

有志者事竟成

Where there's a will, there's a way

University of Alberta

BAYESIAN APPROACH FOR CONTROL LOOP DIAGNOSIS

by

Fei Qi

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

©Fei Qi

Edmonton, Alberta

Fall 2011

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.

To the love of my life, Ying

Abstract

The large number of control loops in a modern industrial plant poses a serious challenge for operators and engineers to monitor these loops to maintain them at optimal conditions continuously. Much research has been done on control loop performance assessment and monitoring of individual components within a control loop. The literature, however, has been sparse in presenting a systematic approach for control loop diagnosis.

This thesis is concerned with establishing a data-driven Bayesian approach for control loop diagnosis. Observations from various monitoring algorithms and *a priori* knowledge of the control loop are synthesized under the Bayesian framework to pinpoint the underlying source of poor control performance. Several challenging practical issues under the proposed framework will also be discussed.

To address the incomplete evidence problem that is often encountered in reality, the missing pattern concept is introduced. The incomplete evidence problems are categorized into single missing pattern ones and multiple missing pattern ones. A novel method based on marginalization over an underlying complete evidence matrix (UCEM) is proposed to include the incomplete evidences into the diagnostic framework, such that information in all the evidence samples can be effectively utilized in the diagnosis.

Data auto-correlation is common in engineering applications. The temporal information hidden in the historical data is extracted by considering evidence and mode dependency in this thesis. Data-driven algorithms for evidence and mode transition probability estimation are developed. An auto-regressive hidden Markov model is built to consider both mode and evidence dependencies. When both the mode and evidence transitions are considered, the temporal information is effectively synthesized under the Bayesian framework.

An approach to estimate the distributions of monitor readings with sparse historical samples is proposed to alleviate the intensive requirement of historical data.

The statistical distribution functions for several monitoring algorithm outputs are analytically derived. A bootstrap based method is proposed to handle the challenging problem of estimating the statistical distribution for valve stiction monitoring. The proposed approach has the potential to estimate evidence distribution with as few as only one evidence sample.

Acknowledgements

This thesis has benefited from many people's help. First and foremost, I would like to express my sincere gratitude to my supervisor, Dr. Biao Huang, for his constant support and vigorous guidance. His insightful advices guided me through every step of this research. His constant encouragement has been a true inspiration. He always spent significant amount of time to give a careful review and detailed corrections of my writings. I am grateful to him for giving me the freedom to explore in the research, and being patient with me during unproductive times. Without all these, it would not have been possible to finish this thesis. I would also like to express my appreciation to my thesis defense committee members, Dr. Ian Craig, Dr. Tongwen Chen, Dr. Vinay Prasad, and Dr. Sirish Shah, for their helpful suggestions on the thesis.

I would like thank Dr. Kwanho Lee, and Mahdi Mohammad for providing sincerely advice and support with the valve stiction monitoring part of this research. I am grateful to Dr. Nina Thornhill, and Dr. Mohieddine Jelali for their generosity of granting the permit to use their data. I would also like to thank Moshood Olanrewaju, and Artin Afacan. They were always willing to help me with the distillation column experiment, even with extended hours. The help from Ruben Gonzalez with the writing of the thesis is also gratefully appreciated.

Evaluating the proposed approaches at Syncrude Canada Ltd. was a rewarding experience, and it led me to a better understanding of industrial process control problems. Without the support from Syncrude Canada Ltd. and many of its employees, it would be impossible to carry on the projects. Special thanks are extended to Edgar Tamayo, Huang Su, Bo Li, Aris Espejo, and Dan Brown, who were always supportive to my research.

I would like to thank my many friends and colleagues at the University of Alberta with whom I have had the pleasure of working over the years. These include Hailei Jiang, Xinguang Shao, Xin Jing, Yu Zhao, Elom Domlan, Yuri Shardt, Liqian

Zhang, Chunyu Liu, Seyi Akande, Fadi Ibrahim, Prabhpreet Grover, Salim Ahmed, Shanshan Liu, Zhijun Mo, and all the members of the CPC group. I will always cherish my friendship with them.

I would like to acknowledge the Department of Chemical and Materials Engineering, University of Alberta, for providing me the opportunity to pursue my doctoral degree. I also gratefully acknowledge the financial support of this research from the Natural Sciences and Engineering Research Council of Canada (NSERC), and the Alberta Ingenuity Fund. Financial support in the form of ConocoPhillips Canada Limited Graduate Scholarship in Hydrocarbon Development, ISA Chemical & Petroleum Industries Division Scholarship, ISA Norman E. Huston Scholarship, Captain Thomas Farrell Greenhalgh Memorial Scholarship, Jacob H. Masliyah Graduate Award in Oil Sands Engineering, and Provost Doctoral Entrance Award are also gratefully acknowledged.

I am eternally grateful to my parents. They always encouraged and supported me to seek out my own goals. Finally, I would like to thank my wife Ying for her unconditional love and support during the ups and downs of my life. In a million ways, both big and small, she has gotten me to this point. Words cannot express my appreciation and love to her.

Contents

1	Introduction	1
1.1	Objective of the thesis	1
1.2	A brief literature overview	2
1.2.1	Control performance assessment	2
1.2.2	Control loop diagnosis	4
1.2.3	Bayesian diagnosis	7
1.3	Outline of the thesis	8
2	Data-driven Bayesian Approach for Control Loop Diagnosis	11
2.1	Introduction	11
2.2	Data-driven Bayesian approach for control loop diagnosis	13
2.2.1	Control loop diagnosis	13
2.2.2	Preliminaries	14
2.2.3	Bayesian control loop diagnosis	16
2.3	Simulation example	20
2.3.1	Process description	20
2.3.2	Monitor selection	21
2.3.3	Diagnostic settings and results	23
2.4	Industry evaluation	25
2.4.1	Process description	25
2.4.2	Data-driven Bayesian diagnosis	26
2.5	Conclusions	28
3	A Bayesian Approach for Control Loop Diagnosis with Incomplete Historical Evidence Data	29
3.1	Introduction	29
3.2	Incomplete evidence problem	30

3.3	Diagnosis with incomplete evidence	32
3.3.1	Single missing pattern problem	33
3.3.2	Multiple missing pattern problem	38
3.4	Simulation example	40
3.4.1	Single missing pattern	40
3.4.2	Approximation of multiple missing pattern solution	47
3.5	Industrial case study and evaluation	48
3.6	Conclusions	49
4	Bayesian Methods for Control Loop Diagnosis in Presence of Temporal Dependent Evidences	51
4.1	Introduction	51
4.2	Temporally dependent evidences	52
4.3	Estimation of evidence transition probability	54
4.4	Reduction of evidence transition space	61
4.5	Simulation example	64
4.5.1	Diagnostic settings	65
4.5.2	Diagnostic results	66
4.6	Pilot scale experiment	70
4.6.1	Process description	70
4.6.2	Diagnostic settings and results	72
4.7	Conclusions	75
5	Dynamic Bayesian Approach for Control Loop Diagnosis with Underlying Mode Dependency	76
5.1	Introduction	76
5.2	Mode dependency	77
5.2.1	Dependent mode	77
5.2.2	Estimation of mode transition probability	80
5.3	Dependent evidence and mode	82
5.4	Simulation example	84
5.4.1	Diagnostic settings	84
5.4.2	Diagnostic results	86
5.5	Pilot scale experiment	88
5.5.1	Process description	88

5.5.2	Diagnostic settings and results	89
5.6	Conclusions	91
6	Control Loop Diagnosis with Sparse Historical Data	92
6.1	Introduction	92
6.2	Analytical estimation of monitor output distribution function	93
6.2.1	Control performance monitor	94
6.2.2	Process model monitor	95
6.2.3	Sensor bias monitor	97
6.3	Bootstrap approach to estimate monitor output distribution function	99
6.3.1	Valve stiction identification	99
6.3.2	The bootstrap method	102
6.3.3	Illustrative example	107
6.3.4	Applications	112
6.4	Simulation example	115
6.4.1	Diagnostic settings and results	118
6.4.2	Weighting of historical data	120
6.5	Experimental example	122
6.5.1	Process description	123
6.5.2	Diagnostic settings and results	125
6.6	Conclusions	127
7	Concluding Remarks and Future Work	130
7.1	Concluding remarks	130
7.2	Recommendations for future work	132
	Bibliography	134
A	Estimation of Likelihood with Incomplete Evidences	141
A.1	Evidence likelihood estimation	141
A.2	Proof of Lemma A.1	144
B	Estimation of Mode Transition Probability	147

List of Figures

2.1	Typical control system structure	14
2.2	Likelihood updating	20
2.3	Distillation column simulation system	21
2.4	Posterior probability assigned to each mode	24
2.5	Structure of diluted oil pre-heater	26
2.6	Posterior probability assigned to each mode	28
3.1	Estimation of expected complete evidence numbers out of the incomplete samples	36
3.2	Comparison of complete evidence numbers	41
3.3	Diagnostic results with different data set	42
3.4	Mean value of posteriors with Monte Carlo simulations	43
3.5	Standard deviations of posteriors with Monte Carlo simulations	43
3.6	Histogram of posteriors assigned to m_9	44
3.7	Comparison of complete evidence numbers	45
3.8	Reconstruction error of two data recovery methods	46
3.9	Diagnostic results with different data set	49
4.1	Bayesian model with independent evidence data samples	52
4.2	Monitor outputs of the illustrative problem	53
4.3	Bayesian model considering dependent evidence	54
4.4	Illustration of evidence transition samples	55
4.5	Evidence transition probability updating	61
4.6	Correlation ratios of monitors	67
4.7	Numbers of occurrences diagnosed for each mode	68
4.8	Average posteriors for validation data from mode m_3 and m_7	70
4.9	Pilot scale tank process	71
4.10	Numbers assigned to each mode	73

4.11	Average posterior probability for each mode	74
5.1	Bayesian model considering dependent mode	78
5.2	Historical composite mode data set for mode transition probability estimation	81
5.3	Dynamic Bayesian model that considers both mode and evidence de- pendency	83
5.4	Number of occurrences diagnosed for each mode	86
5.5	Posterior of mode m_3 and m_7	88
5.6	Numbers of occurrences diagnosed for each mode	90
5.7	Average posteriors assigned to each mode	91
6.1	Operation diagram of sticky valve	100
6.2	Stiction model flow diagram	101
6.3	Bounded stiction parameter search space	102
6.4	Bootstrap method flow diagram	108
6.5	Histogram of simulated \hat{S} and \hat{J}	109
6.6	Auto-correlation coefficient of residuals	110
6.7	Histogram of residual distribution	110
6.8	Comparison of parameter histograms	111
6.9	Histogram of bootstrapped \hat{S}^b and \hat{J}^b for chemical 55	113
6.10	Histogram of bootstrapped \hat{S}^b and \hat{J}^b for chemical 60	113
6.11	Histogram of bootstrapped \hat{S}^b and \hat{J}^b for paper 1	113
6.12	Histogram of bootstrapped \hat{S}^b and \hat{J}^b for paper 9	114
6.13	TE process scheme diagram	117
6.14	TE problem diagnosis with all historical data	119
6.15	TE diagnosis with only one sample from mode IDV 8	119
6.16	Comparison of likelihood distributions	120
6.17	TE diagnosis with only one sample from mode IDV 2	121
6.18	TE diagnosis with only one sample from mode IDV 7	121
6.19	TE diagnosis with only one sample from mode IDV 14	121
6.20	Diagnosis with different number of historical samples for IDV 14	122
6.21	Schematic diagram of the distillation column	124
6.22	Distillation column diagnosis with all historical data	126
6.23	Distillation column diagnosis with only one sample from mode m_1	127

6.24	Distillation column diagnosis with only one sample from mode m_2	. 128
6.25	Distillation column diagnosis with only one sample from mode m_3	. 128
6.26	Distillation column diagnosis with only one sample from mode m_4	. 129

List of Tables

2.1	Operating modes	21
2.2	Summary of monitors	23
2.3	Summary of Bayesian diagnostic parameters	25
2.4	Summary of Bayesian diagnostic parameters	27
3.1	Summary of historical and prior samples	36
3.2	Estimated likelihood	37
3.3	Summary of historical and prior samples	38
3.4	Realizations of evidences with different data set	49
4.1	Likelihood estimation of the illustrative problem	54
4.2	Summary of Bayesian diagnostic parameters	66
4.3	Overall correct diagnosis rate	69
4.4	Correct diagnosis rates for m_3 and m_7	70
4.5	Summary of Bayesian diagnostic parameters	73
4.6	Correct diagnosis rate for each single mode	74
5.1	Summary of Bayesian diagnostic parameters	85
5.2	Overall correct diagnosis rates	87
5.3	Correct diagnosis rate for m_3 and m_7	87
5.4	Summary of Bayesian diagnostic parameters	89
5.5	Overall diagnosis rates	90
5.6	Correct diagnosis rates for each single mode	91
6.1	Comparison of sample standard deviations	111
6.2	Confidence intervals of the identified stiction parameters	112
6.3	List of simulated modes	118
6.4	Summary of Bayesian diagnostic parameters	118
6.5	Detail dimensions of the column and trays	123

6.6	Operating modes for the column	125
6.7	Commissioned monitors for the column	125
6.8	Summary of Bayesian diagnostic parameters	126

Chapter 1

Introduction

1.1 Objective of the thesis

Control loops play important roles in chemical engineering processes. Malfunctioning components in process control loops, including sensors, actuators and other components, are not unusual in industrial environments. Their effects introduce excess variation throughout the process thereby reducing machine operability, increasing costs and disrupting final product quality control. It is reported as many as 60% of industrial controllers have some kind of control performance problems [29]. Some of the common causes of poorly performing control loops are [16, 94]:

- Improper controller tuning;
- Changing process dynamics (transitions, unmeasured disturbances);
- Limited controller output range;
- Large dead time or inaccurate determination of dead time;
- Inappropriate sampling interval;
- Incorrect controlled and manipulated variable pairings;
- Poor hardware (sensor, actuator) maintenance.

The incentives of this research arise from the important task of isolating and diagnosing abnormalities of control loops in complex industry processes. A typical modern process industry operation consists of hundreds or even thousands of control loops, which is overwhelming for the plant personnel to detect as well as to isolate control loops having deteriorated performance. Moreover, even if poor performance

is detected in some control loops, due to the reason that a problem in a single process component may invoke a wide spread of control performance degradation, locating the underlying problem source is not a trivial task. Without an advanced information analysis framework, it is difficult to handle the overwhelming information flood of process data and alarms to determine the source of the underlying problem. Human beings' inability of synthesizing large amount and high dimensional process data is the main reason behind those problems. The purpose of control performance monitoring and diagnosis is to provide an automated procedure that delivers information to plant personnel for determining whether specified performance targets are being met by the controlled process variables and that evaluates the performance of control loops [41], as well as suggests possible problem sources and troubleshooting sequence.

The main objective of this study is to establish a Bayesian approach for control loop diagnosis, synthesizing observations of different monitoring algorithms and *a priori* knowledge of the control loop, to suggest possible faulty sources based on Bayesian probabilistic framework. Some related open problems and issues will be investigated. An equally important objective of this study is to apply the proposed Bayesian diagnostic approach to experimental and industrial processes to verify and demonstrate validity and practicality of this method.

1.2 A brief literature overview

1.2.1 Control performance assessment

Performance assessment is concerned with the analysis of available process data against certain benchmarks. The research was started by the ground-breaking study of Harris (1989) [29] for proposing the Minimum Variance Control (MVC) benchmark. Huang and Shah (1995) [38] developed a filtering and correlation (FCOR) algorithm to estimate the MVC benchmark effectively. A state space framework for MVC benchmark was proposed by McNabb and Qin (2005) [65]. The MVC index was extended to MIMO systems by Harris et al. (1996) [28]. Huang and Shah (1997) [37] tackled MIMO MVC benchmark by introducing the unitary interactor matrix. MVC benchmark provides a readily computable and physically significant bound on control performance.

Although the MVC benchmark provides a simple way to evaluate control performance, *a priori* knowledge of process time delay or interactor matrix is a requirement. Extended horizon performance index based on MVC was introduced by Desborough and Harris (1992) [16]. This method does not need *a priori* knowledge of system delay or interactor matrix, and can also reflect deterministic performance, for instance, settling time. Extended horizon index can also be extended to multivariate case [28].

The theoretical variance lower bound of MVC may not be achievable for most of practical controllers. More realistic performance indices are needed. Ko and Edgar (1998) [48] discussed PID benchmark. An interesting result was presented by Qin (1998) [76], stating that MVC can be achievable for PID controller when process time delay is small or large, but not medium. Huang and Shah (1999) [36] proposed the linear quadratic Gaussian (LQG) regulator benchmark as an alternative to the MVC benchmark, based on process model. Model-based approaches also exist for benchmarking model predictive control (MPC) systems; see Shah et al. (2001) [89], and Gao et al. (2003) [22].

The benchmarks discussed above mainly focus on stochastic performance. However, those benchmarks can also be related with deterministic performances, such as overshoot, decay ratio, settling time, etc. Ko and Edgar (2000) [49] modified the MVC index to include setpoint variations in the inner loop of cascade control. Influence of setpoint changes on the MVC index was discussed by Seppala et al. (2002) [87], where a method that decomposes the control error into the one resulted from setpoint changes and a setpoint detrended signal was proposed. Thornhill et al. (2003) [97] examined the reasons why performance during setpoint change differs from the performance during operation at a constant setpoint. The extension of the MVC index to the varying setpoint case has also been discussed by McNabb and Qin (2005) [65].

In practice, current operation performance is often compared with historical data during a time period when the control performance was benchmarked as “good” from the user’s viewpoint. Such criteria are called baselines, historical data benchmarks, or reference data set benchmarks [25, 22, 84]. Although such historical data benchmarks are pragmatic and practical, sometimes they may be too subjective and rely heavily on how the history data are selected.

Some other methods have also been proposed for control performance assess-

ment. Kendra and Cinar (1997) [47] applied frequency analysis to evaluate control performance. r -statistic was introduced by Venkataramanan et al. (1997) [100], which detects deviations from setpoint, regardless of the output noise. Li et al. (2003) [57] proposed a relative performance monitor, which compares the performance of a control loop to that of a reference model.

A number of commercial control performance assessment software packages are available in the market, such as the Intune software tools by Control Soft, Loop-Scout by Honeywell Hi-Spec Solutions, Performance Surveyor by DuPont, etc. [41]. Various successful industrial applications have also been reported [31, 41].

1.2.2 Control loop diagnosis

The most common reasons for the downgrade of control performance include: mistuned controller parameters, sticky valves, model plant mismatch. Significant work has been done on the diagnosis on those single problems. However, the diagnosis of the overall control loop is still of an open problem. Generally the controller performance is evaluated by the various control performance assessment techniques as discussed in the previous section. The following discussion will focus on overview of other aspects of control loop monitoring.

Valve stiction diagnosis

The undesirable behavior of control valves is the biggest single contributor to poor control loop performance [43]. According to Jelali and Huang (2009) [43], 20-30% of control loop oscillations are induced by valve nonlinearities, including stiction, deadband, hysteresis, etc. Among these problems, stiction is the most common one in the process industry [45]. Oscillation in control loops increases the variability of process variables, which in turn affects product quality, increases energy consumption, and accelerates equipment wear. Detecting valve stiction in a timely manner will bring significant economic benefits, and thus there is a strong incentive for the valve stiction detection research. A comprehensive review and comparison of valve stiction detection methods can be found in Jelali and Huang (2009) [43].

Singhal and Salsbury (2005) [91] proposed a stiction detection methodology by calculating the ratio of the areas before and after the oscillation peaks of PV signal. A method for diagnosing valve stiction was developed based on observations of control loop signal patterns by Yamashita (2006) [104]. The method determines

typical patterns from valve input and valve output/process variable in the control loop, and thus does not allow detection of stiction which shows up in different patterns. Scali and Ghelardoni (2008) [83] improved the work of Yamashita (2006) [104] to allow different possible stiction patterns to be considered. Choudhury et al. (2007) [12] proposed a controller gain change method, which is based on the change in the oscillation frequency due to changes in the controller gain to detect valve stiction. Yu et al. (2008) [105] showed that this method can fail to detect the presence of the sticky valve in interacting multi-input multi-output systems. A strategy based on the magnitude of relative change in oscillation frequency due to changes in controller gain is proposed to overcome the limitations of the existing method.

Despite of the various work regarding stiction detection, valve stiction quantification remains to be a challenging problem. Choudhury et al. (2008) [13] proposed a method to quantify stiction using the ellipse fitting method. The PV vs. OP plot is fitted to an ellipse and the amount of stiction is estimated as the maximum length of the ellipse in the OP direction. Chitralkha et al. (2010) [11] treated the problem of estimating the valve position as an unknown input estimation problem. The valve position is estimated via a Kalman filter type unknown input estimator. Jelali (2008) [42] presented a global optimization based method to quantify valve stiction. A Similar method was also proposed by Srinivasan et al. (2005) [93]. The approach is based on identification of a Hammerstein model consisting of a sticky valve and a linear process. The stiction parameters and the model parameters are estimated simultaneously with a global grid search optimization method. Lee et al. (2009) [43] presented a closed-loop stiction quantification approach using routine operating data. A suitable model structure of valve stiction is chosen prior to conducting valve stiction detection and quantification. Given the stiction model structure, a feasible search domain of stiction model parameters is defined, and a constrained optimization problem is solved for search of stiction model parameters.

The aforementioned stiction qualification methods all assume that the process is linear. Nallasivama et al. (2010) [69] proposed a method to qualify the stiction for closed-loop nonlinear systems. The key idea used in the approach is based on the identification of extra information available in process output PV compared to the controller output, OP. Stiction phenomenon leads to many harmonic components compared to the Fourier transform of the Volterra system, which allows stiction

detection in nonlinear loops.

Model mismatch diagnosis

A large volume of work has been done for open loop model validation. However, the literature has been relatively sparse on studies concerned with on-line model validation using closed-loop data.

Huang (2001) [34] established a method for the analysis of detection algorithms in the frequency domain under closed-loop conditions. The divergence algorithm is extended to the model validation for the general Box-Jenkins model under closed-loop conditions through the frequency domain approach. Based on the two-model divergence method, Jiang et. al. (2009) [44] developed two closed-loop model validation algorithms, which are only sensitive to the plant changes. Of the two algorithms, one is sensitive to changes in both plant and disturbance dynamics, while the other one is only sensitive to the changes in plant dynamics, regardless of changes in disturbance dynamics and additive process faults, such as sensor bias.

Badwe et. al. (2009) [5] proposed a model mismatch detection method based on the analysis of partial correlations between the model residuals and the manipulated variables. The more significant this correlation, the higher is the possibility that there exists model mismatch. Badwe et. al. (2010) further extended their previous work by analyzing the impact of model mismatch on the control performance in [6].

In Selvanathan and Tangirala (2010) [86], a plant model ratio (PMR) is introduced as a measure to quantify the model-plant mismatch in the frequency domain. The PMR provides a mapping between its signatures and changes in process models, and thus the changes in model gain, time constant and time delay can be identified. Although it is claimed that the PMR can be estimated from closed-loop operating data, a significant underlying assumption is that the set-point contains at least a pulse change. This assumption, however, can be restrictive in practice.

Overall control loop diagnosis

Despite the large amount of work on single component diagnosis in control loop, little has been done for the overall control loop diagnosis. The most significant challenge for control loop diagnosis is the the existence of similar symptoms among different problem sources [35]. The monitoring or diagnostic methods discussed previously often focus on one specific problem, and the potential abnormalities in the other

unattended components are ignored [75]. A systematic approach is required to take all possible faults into consideration. A Bayesian frame for control loop diagnosis has been proposed by Huang (2008) [35]. The outputs from different monitoring algorithms are synthesized to provide a probabilistic result for control loop diagnosis.

1.2.3 Bayesian diagnosis

Bayesian approach has proven to be useful for a variety of monitoring and predictive maintenance purposes. Applications of Bayesian approach have been reported in medical science, image processing, target recognition, pattern matching, information retrieval, reliability analysis, and engineering diagnosis [73, 17]. It provides a flexible structure for modeling and evaluating uncertainty. In the presence of noise and disturbances, Bayesian inference provides a well-suited way to solve the process monitoring and diagnosis problem, providing quantifiable measure of uncertainty for decision making. It is one of the most widely applied techniques in probabilistic inferencing [17].

Mehranbod et al. (2003) [66] expanded Bayesian model to detect sensor faults in a dynamic process, whereas most other work utilizing Bayesian model for process fault detection and diagnosis mainly focusing on steady state operations. An intelligent automation system for predictive maintenance of machine tools, based on Bayesian model was proposed by Gilabert and Arnaiz (2006) [27]. Wolbrecht et al. (2000) [103] designed “part models” to represent individual parts in a process. These “part models” were combined to form a Bayesian model of the entire manufacturing process. Similar Bayesian model structure can also be found in Mehranbod and Soroush (2003) [66], where single-sensor models are used as building blocks to develop a Bayesian model for all sensors in the process under consideration. Some commercial software packages are also available for Bayesian diagnosis, such as Netica, and MonteJade [7].

How to build a Bayesian model is of great interest to many researchers. Dey and Stori (2005) [17] used a data-driven method with Dirichlet prior distribution to build Bayesian model to diagnose root causes of process variations. Such kind of data-driven method based on Dirichlet prior distribution for Bayesian model learning has also been addressed by Pernestal (2007) [73]. Sahin et al. (2007) [82] implemented a fault diagnostic technique for airplane engines using the particle swarm optimization algorithm to learn the structure of Bayesian model from a large

data set. The methods discussed above are all data based. However, in the case that there are not enough data, model based methods should be considered. Romessis and Mathioudakis (2006) [80] proposed a method to build Bayesian models from mathematical models for aircraft engine diagnosis, without the need of hard-to-find flight data with faults.

Besides the learning problem, evaluation of posterior probabilities also attracts much attention in Bayesian diagnosis community. The storage space and computation burden of Bayesian approach increase exponentially over the number of nodes in the Bayesian model. A compact storage strategy was introduced by Pernestal (2007) [73]. In Lewis and Ransing (1997) [56], the notion of conditional probability was generalized to enable the belief revision even in the presence of partial evidence. Dealing with continuous variables is very expensive for the inferences computation, so Flores-Loredo et al. (2005) [21] utilized automatic learning algorithms, together with expert advices to determine the Bayesian model of the most common faults in gas turbines.

Another issue widely considered is the evolution of Bayesian model with time. Extension of the Bayesian model in time domain, called Dynamic Bayesian Network (DBN), is used to model changes of Bayesian network over time. Kawahara et al. (2005) [46] built DBN for diagnosis from *a priori* knowledge, and modified it by statistical learning with operation data. A well summarized research regarding DBN was presented by Murphy (2002) [68].

Nevertheless, few results have been reported about implementing Bayesian model in diagnosing industry applications. This is possibly owing to the complexity of industry processes, which makes modeling, storage, and updating of belief of Bayesian model difficult tasks. Also, unknown disturbances are not unusual in industry processes. Omitting these unknown nodes may make the modeling of Bayesian model impossible, since the disturbances may be main contribution sources of certain symptoms (measurements/observations).

1.3 Outline of the thesis

This thesis begins with an introduction to provide an overview of the main areas of focus in this work by outlining the research scope and major objectives.

Chapter 2 establishes a control loop diagnostic strategy through a data-driven Bayesian approach. This approach synthesizes information from different monitoring

algorithms to isolate possible problem sources. Performance of the proposed data-driven Bayesian approach is examined through simulations as well as an industrial application example to verify its ability of information synthesis.

An outstanding problem of the diagnostic procedure in Chapter 2 is its inability to handle incomplete evidence data. The missing pattern concept is introduced in Chapter 3 to incorporate incomplete evidence samples into the data-driven diagnostic framework. The incomplete evidence problems are classified into single missing pattern ones and multiple missing pattern ones. A novel method based on marginalization over underlying complete evidence matrix (UCEM) is proposed to circumvent the incomplete evidence problems. Performance of the proposed incomplete evidence handling approach is examined through simulations and an industrial application example.

Conventional Bayesian methods commonly assume that the evidences are temporally independent. The assumptions regarding evidence independency, however, are restrictive in most engineering applications. In Chapter 4, the important evidence dependency problem is solved by a data-driven Bayesian approach with consideration of evidence transition probability. The sparse data problem induced by high dimensional evidence transition space is circumvented by analyzing the correlation ratio of monitors. The applications in a simulated distillation column and a pilot scale process are presented to demonstrate the data dependency handling ability of the proposed diagnostic approach.

Chapter 5 further considers the mode dependency to extract more information from historical samples. First of all, a hidden Markov model is built to address the temporal mode dependency problem in control loop diagnosis. A data-driven algorithm is developed to estimate the mode transition probability. The new solution to mode dependency is then further synthesized with the solution to evidence dependency to develop a recursive auto-regressive hidden Markov model for the online control loop diagnosis. When both the mode and evidence transitions are considered, the temporal information is effectively synthesized under the Bayesian framework. A simulated distillation column example and a pilot scale experiment example are presented to investigate the ability of the proposed diagnostic approach.

A major concern with the data-driven Bayesian approach is the intensive requirement for historical data. While in industry, the faulty data is often sparse. In extreme cases, a fault may only happen once or none, which the data-driven

approaches cannot handle. Chapter 6 proposes an approach to estimate the distribution of monitor readings in the presence of insufficient historical data. The distributions of monitor readings are estimated with analytical approaches and the bootstrap method. The applications of the proposed approach to the Tennessee Eastman Challenge problem and an experimental distillation column are presented to examine the performance of the proposed likelihood reconstruction methods. This is followed by concluding remarks and suggestions for future work in Chapter 7.

Chapter 2

Data-driven Bayesian Approach for Control Loop Diagnosis

This chapter is concerned with determination of the underlying source of problematic control performance through a data-driven Bayesian approach. This approach synthesizes information from different monitoring algorithms to isolate possible problem sources. Performance of the proposed data-driven Bayesian approach is examined through simulations as well as an industrial application example to verify its ability of information synthesis.

2.1 Introduction

Control loop performance monitoring and diagnosis has been and remains one of the most active research areas in process control community. A number of control performance assessment methods have been developed, including the ones based on minimum variance control (MVC), linear quadratic Gaussian control (LQG), historical data trajectories, and user-specified control, etc [36, 29, 76, 41, 84, 72]. Several surveys on control performance assessment research are available, and a number of successful applications of control performance monitoring algorithms have been reported [29, 76, 31, 41, 39, 90]. Besides performance assessment of control loops, significant progress has also been made in the development of instrument and process monitors, including sensor monitors, actuator monitors, and model validation monitors [77, 3, 13, 67]. However, many problems remain. One of the outstanding problems is that monitoring algorithms are often designed for detection of one specific abnormality. An implicit assumption that other unattended components are in

A version of this chapter has been published in F. Qi, B. Huang, and E.C. Tamayo. A Bayesian approach for control loop diagnosis with missing data. *AIChE Journal*, 56:179-195, 2010.

good shape is made. Clearly this assumption does not always hold. Different problems can produce similar symptoms, thus triggering the same monitor to alarm. On the other hand, one problem source can also affect several monitors simultaneously. Although there exists a large volume of papers addressing control loop monitoring, the literature has been relatively sparse in reporting a systematic way for control loop diagnosis [41, 84, 72, 98]. Continuous improvement in control performance must be accompanied by constantly monitoring the performance of control loops, and diagnosing the source of poor performance such as poor tuning, a sticky valve, a major disturbance upset, or other root causes [41, 98]. It is necessary to develop methods that not only monitor individual components, but are also capable of synthesizing information from different monitors to isolate the underlying source of problematic control performance.

There are several challenging issues in monitor synthesis [35]. First, although problem sources may be different, the symptoms can be similar. For instance, oscillations can either be introduced by a sticky valve or an improperly tuned controller. Second, all processes operate in an uncertain environment to some extent, and there are uncertainties in the links between problem sources and monitor readings due to disturbances. No monitor has a 100% successful detection rate or 0% false alarm rate, and thus a probabilistic framework should be built to describe the uncertainties. Last but not least, how to incorporate *a priori* knowledge in the diagnostic system to improve diagnostic performance is also worth consideration. Most of the existing monitoring methods are data based. However, incorporating *a priori* knowledge such as causal relations between variables is not only helpful, but is necessary for an accurate diagnosis [35].

In view of the challenges listed above, the Bayesian method sheds lights on the problem solution by providing a probabilistic information synthesizing framework. It has been proven useful for a variety of monitoring and predictive diagnosis purposes. Applications of Bayesian methods have been reported in medical science, image processing, target recognition, pattern matching, information retrieval, reliability analysis, and engineering diagnosis [73, 17, 67, 95]. It is one of the most widely applied techniques in probabilistic inferencing [17]. Built upon previous work in Bayesian fault diagnosis [73] and a framework outlined in Huang (2008) [35], this chapter presents a data-driven algorithm for control loop diagnosis based on the Bayesian approach.

The remainder of this chapter is organized as follows: In Section 2.2, a general description of the control loop diagnosis problem is given first, and some presumptions are made. A systematic approach for data-driven control loop diagnosis is also presented. Simulations for a binary distillation column are shown in Section 2.3. The diagnostic approach is applied to an industry process in Section 2.4. Section 2.5 concludes this chapter.

2.2 Data-driven Bayesian approach for control loop diagnosis

2.2.1 Control loop diagnosis

Typically a control loop consists of the following components: controller, actuator, process, and sensor, all subject to disturbances. These components may all suffer from certain abnormalities. For example, a valve acting as an actuator may suffer from a stiction problem; the output of a sensor may be biased. All these problems may cause degradation of control performance, such as large variation of process variables, loop oscillation, etc.

In this work, measurements of manipulated variables (MVs) and controlled variables (CVs), and the nominal operating point are assumed to be available. If validation of the process model or the disturbance model is of interest, then their corresponding nominal models should naturally be available. We further assume that all or some monitors are available for the components of interest in the control loop. There may be, for example, the control performance monitor, valve stiction monitor, process model validation monitor, and sensor monitor. These monitors, however, are all subject to disturbances and thus false alarms, and each monitor can be sensitive to abnormalities of other problem sources. For instance, a valve with stiction problem in a univariate control loop may cause the alarms of several monitors, in addition to the valve stiction monitor itself, as shown in Figure 2.1, where the monitors marked with gray may respond to the valve stiction problem. It is challenging to determine where the problem source is with several simultaneous alarming monitors. Our goal is to determine the underlying source of problematic control performance based on the outputs of all monitors.

In the presence of disturbances, Bayesian inference provides a well suited way to solve the diagnostic problem, quantifying the uncertainty in its conclusion. In

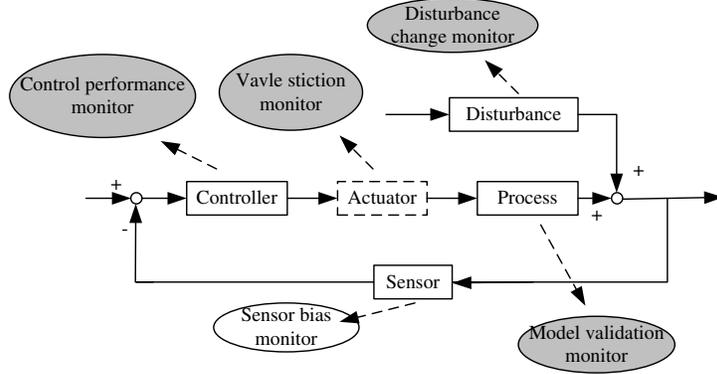


Figure 2.1: Typical control system structure

the work of Pernestål (2007) [73] and Pernestål and Nyberg (2007) [74], a Bayesian approach for diesel engine diagnosis based on complete sensor readings is studied. This section adopts this approach to control loop diagnostic problem based on the readings of control loop monitors.

2.2.2 Preliminaries

To apply the Bayesian method to control loop diagnosis, several notations need to be introduced.

Mode M

Assume that a control loop under diagnosis consists of P components of interest: C_1, C_2, \dots, C_P , among which the problem source may lie in. All these components are subject to possible abnormality or performance deterioration. Each component is said to have a set of discrete operating status. For instance, the sensor might be “biased” or “unbiased”. The control loop diagnostic problem is to determine the operating status of all these components in the case of problematic loop performance, i.e., to locate the underlying problem source of degraded control performance. An assignment of operating status to all the components of interest in the control loop is called a mode, and is denoted as M ; M can take different values and its specific value is denoted by m . For example, $m=(C_1=\textit{well tuned controller}, C_2=\textit{valve with stiction}, \dots)$. A mode representing normal operation is referred to as NF (normal functioning), which means that all components operate normally.

Suppose that component C_i has q_i different status. Then the total number of

possible modes is

$$Q = \prod_{i=1}^P q_i,$$

and the set of all possible modes can be denoted as

$$\mathcal{M} = \{m_1, m_2, \dots, m_Q\}.$$

Evidence E

The monitor readings, called evidence, are input to the diagnostic system, and are denoted as $E = (\pi_1, \pi_2, \dots, \pi_L)$, where π_i is the output of the i th monitor, and L is the total number of monitors.

Often the monitor readings, which are generally continuous, are discretized according to predefined thresholds. In this work, monitor readings are all assumed to be discrete. For example, the control performance monitor may indicate “optimal”, “normal”, or “poor”, depending on the thresholds adopted. The specific value of evidence E is denoted as e ; for example, $e = (\pi_1 = \textit{optimal control performance}, \pi_2 = \textit{no sensor bias}, \dots)$. Suppose that the single monitor output π_i has k_i different discrete values. Then there are

$$K = \prod_{i=1}^L k_i$$

different evidences, and the set of all evidences can be denoted as

$$\mathcal{E} = \{e_1, e_2, \dots, e_K\},$$

where e_i is the i th possible evidence value of E .

Historical data set \mathcal{D}

Historical data are retrieved from the past data record where the mode of control loop, namely, status of all components of interest in the control loop, is available, and the monitor readings are also recorded.

Each sample d^t at time t in the historical data set \mathcal{D} consists of the evidence e^t and the underlying mode m^t . This can be denoted as $d^t = (e^t, m^t)$, and the set of historical data is denoted as

$$\mathcal{D} = \{d^1, d^2, \dots, d^{\hat{N}}\},$$

where \hat{N} is the number of historical data samples. The historical data set can be further divided into subsets under different modes,

$$\mathcal{D} = \{\mathcal{D}_{m_1}, \mathcal{D}_{m_2}, \dots, \mathcal{D}_{m_Q}\},$$

and

$$\mathcal{D}_{m_i} = \{d_{m_i}^1, \dots, d_{m_i}^{N_{m_i}}\}$$

includes all historical samples with the underlying mode being m_i , where N_{m_i} is the number of historical samples corresponding to mode m_i , and $\sum_i N_{m_i} = \hat{N}$.

Different historical data samples may be auto-dependent or independent, contingent on how they are sampled as well as how the disturbances affect the monitors. Each monitor reading is calculated from a segment (window) of recorded process data. If there is no overlap of the windows between two consecutive monitor calculations and there is a sufficient gap between the two windows, then the monitor readings are considered to be independent. In this chapter, all the historical data samples are assumed to be independent, i.e.

$$p(\mathcal{D}) = p(d^1, d^2, \dots, d^{\hat{N}}) = p(d^1)p(d^2) \cdots p(d^{\hat{N}}). \quad (2.1)$$

2.2.3 Bayesian control loop diagnosis

Given current evidence E , the historical data set \mathcal{D} , Bayes' rule can be stated as follows:

$$p(M|E, \mathcal{D}) = \frac{p(E|M, \mathcal{D})p(M|\mathcal{D})}{p(E|\mathcal{D})}, \quad (2.2)$$

where $p(M|E, \mathcal{D})$ is the conditional probability of mode M in the control loop given current evidence E , historical data set \mathcal{D} , which is also known as posterior probability or simply posterior; $p(E|M, \mathcal{D})$ is the probability of having current evidence E , conditioning on mode M with historical data \mathcal{D} , also known as likelihood probability or simply likelihood; $p(M|\mathcal{D})$ is the prior probability of mode M ; and $p(E|\mathcal{D})$ is a scaling factor, and can be calculated as $p(E|\mathcal{D}) = \sum_M p(E|M, \mathcal{D})p(M|\mathcal{D})$. Note that historical data are selectively collected when control loop operates under different modes; therefore they provide no information of prior probability of the abnormality, $p(M|\mathcal{D}) = p(M)$ [73]. As a result, Equation 2.2 is often written as

$$p(M|E, \mathcal{D}) \propto p(E|M, \mathcal{D})p(M). \quad (2.3)$$

Since prior probability is determined by *a priori* information, the main task of building a Bayesian diagnostic system is the estimation of the likelihood probability

$p(E|M, \mathcal{D})$, whose objective is to make the estimated likelihood probability be consistent with historical data \mathcal{D} . Pernestál (2007) [73] presented a data-driven Bayesian algorithm to estimate the likelihood in diesel engine diagnosis. This method is adopted here for control loop diagnosis.

Suppose that the likelihood of evidence $E = e_i$ under mode $M = m_j$ is to be calculated, where

$$e_i \in \mathcal{E} = \{e_1, \dots, e_K\},$$

and

$$m_j \in \mathcal{M} = \{m_1, \dots, m_Q\}.$$

The likelihood $p(e_i|m_j, \mathcal{D})$ can only be estimated from the historical data subset \mathcal{D}_{m_j} where the mode $M = m_j$,

$$p(e_i|m_j, \mathcal{D}) = p(e_i|m_j, \mathcal{D}_{m_j}, \mathcal{D}_{-m_j}) = p(e_i|m_j, \mathcal{D}_{m_j}), \quad (2.4)$$

where \mathcal{D}_{-m_j} is the data set whose underlying mode is not m_j .

The likelihood probability can be computed by marginalization over all possible likelihood parameters,

$$p(e_i|m_j, \mathcal{D}_{m_j}) = \int_{\Omega} p(e_i|\Theta_{m_j}, m_j, \mathcal{D}_{m_j}) f(\Theta_{m_j}|m_j, \mathcal{D}_{m_j}) d\Theta_{m_j}, \quad (2.5)$$

where $\Theta_{m_j} = \{\theta_{1|m_j}, \theta_{2|m_j}, \dots, \theta_{K|m_j}\}$ are the likelihood parameters for all possible evidences of mode m_j , and K is the total number of possible evidences; for example, $\theta_{i|m_j} = p(e_i|m_j)$ is the likelihood of evidence e_i when the underlying mode is m_j ; Ω is the space of all likelihood parameters Θ_{m_j} . In Equation 2.5, $f(\Theta_{m_j}|m_j, \mathcal{D}_{m_j})$ can be calculated according to Bayes' rule:

$$f(\Theta_{m_j}|m_j, \mathcal{D}_{m_j}) = \frac{p(\mathcal{D}_{m_j}|\Theta_{m_j}, m_j) f(\Theta_{m_j}|m_j)}{p(\mathcal{D}_{m_j}|m_j)}. \quad (2.6)$$

In Equation 2.6, Dirichlet distribution is commonly used to model priors of the likelihood parameters with Dirichlet parameters $a_{1|m_j}, \dots, a_{K|m_j}$ [73],

$$f(\Theta_{m_j}|m_j) = \frac{\Gamma(\sum_{i=1}^K a_{i|m_j})}{\prod_{i=1}^K \Gamma(a_{i|m_j})} \prod_{i=1}^K \theta_{i|m_j}^{a_{i|m_j} - 1}, \quad (2.7)$$

where $a_{i|m_j}$ can be interpreted as the number of prior samples for evidence e_i under mode m_j , which will be elaborated shortly; $\Gamma(\cdot)$ is the gamma function,

$$\Gamma(x) = \int_0^{\infty} t^{x-1} e^{-t} dt. \quad (2.8)$$

Here all the independent variables x of the gamma functions are prior numbers of evidences, which take positive integers, so

$$\Gamma(x) = (x - 1)!. \quad (2.9)$$

The likelihood of historical data subset \mathcal{D}_{m_j} can be written as

$$p(\mathcal{D}_{m_j} | \Theta_{m_j}, m_j) = \prod_{t=1}^{N_{m_j}} p(d_{m_j}^t | \Theta_{m_j}, m_j). \quad (2.10)$$

The data sample at time t in the historical data subset \mathcal{D}_{m_j} includes the underlying mode m_j and the evidence e^t ,

$$d_{m_j}^t = (e^t, m_j).$$

Thus when $e^t = e_i$,

$$p(d_{m_j}^t | \Theta_{m_j}, m_j) = \theta_{i|m_j}. \quad (2.11)$$

Combining Equation 2.10 and Equation 2.11, we have

$$p(\mathcal{D}_{m_j} | \Theta_{m_j}, m_j) = \prod_{i=1}^K \theta_{i|m_j}^{n_{i|m_j}}, \quad (2.12)$$

where $n_{i|m_j}$ is the number of historical samples where the evidence $E = e_i$, and the underlying mode $M = m_j$.

Substituting Equation 2.12 and Equation 2.6 in Equation 2.5, the following result can be obtained for the likelihood [73]:

$$p(E = e_i | M = m_j, \mathcal{D}) = \frac{n_{i|m_j} + a_{i|m_j}}{N_{m_j} + A_{m_j}}, \quad (2.13)$$

where $n_{i|m_j}$ is the number of historical samples with the evidence $E = e_i$, and mode $M = m_j$; $a_{i|m_j}$ is the number of prior samples that is assigned to evidence e_i under mode m_j ; $N_{m_j} = \sum_i n_{i|m_j}$, and $A_{m_j} = \sum_i a_{i|m_j}$. To simplify notations, the subscript m_j will be omitted when it is clear from the context.

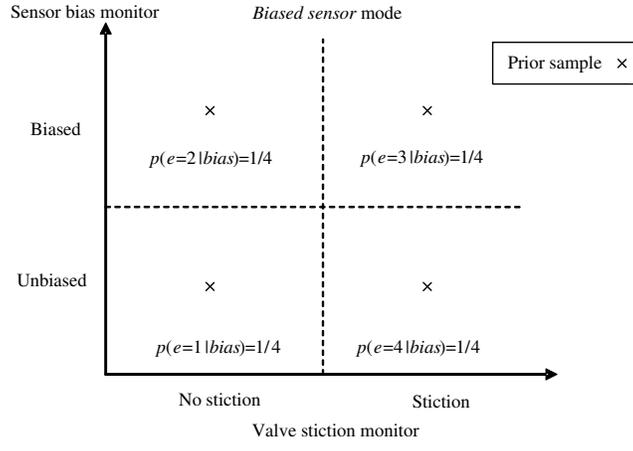
This is a concise yet intuitive result. The likelihood probability is determined by both prior samples and historical samples. As the number of historical data increases, the likelihood probability will converge to the relative frequency determined by the historical data samples, and the influence of priors will decrease. The number of prior samples can be interpreted as prior belief of the likelihood distribution, where a uniform distribution indicates that prior sample numbers are equal across

all evidences under a given underlying mode. It is important to set nonzero prior sample numbers; otherwise the diagnostic system may yield unexpected results [73]. For example, an extreme situation occurs when there is only one sample in the historical data set. Without any prior samples defined, the likelihood for the evidence corresponding to the historical data sample will be assigned with one, whilst likelihood of the other evidences will be zero. This result can be rather misleading during the diagnosis. One may consider that the numbers of the prior samples represent the belief of the prior likelihood. The larger the prior sample numbers are, the stronger belief in the prior likelihood. In general, the numbers of prior samples of all possible evidences are set to be equal as a non-informative prior if there is no prior information available.

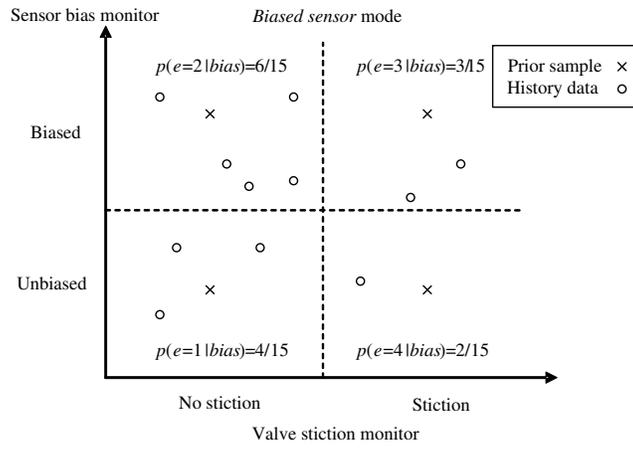
Consider a univariate control loop under diagnosis with two possible problematic components: a valve subject to the possible stiction problem, and a sensor subject to the possible bias problem. Each possible problematic component has a corresponding monitor. The reading of each monitor is discretized into two bins with predefined thresholds; therefore the overall evidence space is discretized into four bins, as shown in Figure 2.2(a). Consider that the underlying system mode is $m=(no\ valve\ stiction, sensor\ bias)$. Each discrete evidence bin is assigned with one prior sample under the assumption of uniformly distributed prior samples. See Figure 2.2(a). Hence, $a_{j|m} = 1$, $A_m = 4$, and the likelihoods of all the evidences equal $1/4$. With the historical data collected under the same underlying mode m , the likelihood probabilities can be updated according to Equation 2.13, as presented in Figure 2.2(b).

With the estimated likelihood probabilities for current evidence E under different modes m_i , $P(E|m_i, \mathcal{D})$, and the user defined prior probabilities $p(m_i)$, posterior probabilities of each mode $m_i \in \mathcal{M}$ can be calculated according to Equation 2.3. Among these modes, the one with largest posterior probability is typically picked up as the underlying mode based on the maximum *a posteriori* (MAP) principle, and the abnormality associated with this mode is then diagnosed as the problem source.

The above procedure illustrates a data-driven approach for control loop diagnosis. Results from different monitors can be synthesized within the Bayesian framework to generate posterior probability for diagnosis.



(a) Likelihood with only prior samples



(b) Updated likelihood with historical data

Figure 2.2: Likelihood updating

2.3 Simulation example

2.3.1 Process description

To investigate diagnostic performance of the proposed Bayesian approach, we apply the diagnostic scheme to a simulated binary distillation column [102]. The column has five inputs, four of which are manipulated variables (MVs) operated by a model predictive controller (MPC). Of the ten outputs, three are controlled quality variables (CVs). They are: top product (distillate) quality measured as final boiling point (FBP top), bottom product (pressure compensated) temperature (PCT bottom), and column pressure. The process is subject to several different problems. All the possible modes, and the corresponding problematic components, are listed in Table 2.1.

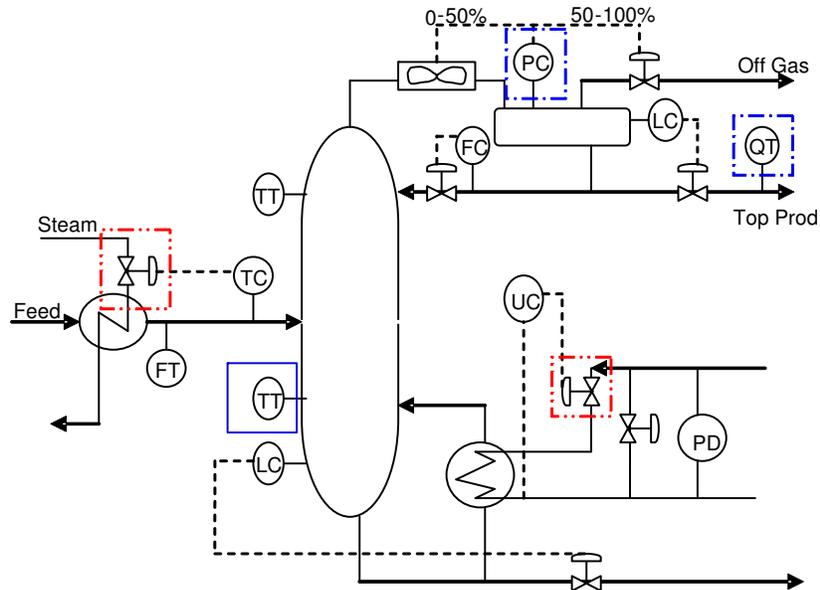


Figure 2.3: Distillation column simulation system

Table 2.1: Operating modes

Mode	Problematic components
NF	None
m_1	Poorly tuned MPC controller
m_2	Feed temperature valve stiction
m_3	Duty valve stiction
m_4	FBP top & PCT bottom model mismatch
m_5	PCT bottom model mismatch
m_6	PCT bottom disturbance dynamic change
m_7	Pressure disturbance dynamic change
m_8	FBP top sensor bias
m_9	Pressure sensor bias
UC	Other unknown errors or combinations of errors

2.3.2 Monitor selection

To evaluate the information synthesizing ability of the Bayesian diagnostic approach, monitors are chosen rather arbitrarily, some of which have high false-alarm/misdetection rate.

Control performance monitor

The minimum variance control benchmark is adopted to evaluate control performances for both univariate and multivariate cases. The FCOR algorithm [36] is employed to compute control performance indices based on both univariate CVs and multivariate CVs.

Valve stiction monitor

For illustrative purposes, we consider the following simplified scenario: if a control loop has oscillation, then the oscillation is caused either by valve stiction or by external oscillatory disturbance. The latter has sinusoid form while the former does not.

If the CV and the MV of a control loop oscillate sinusoidally, by plotting CV versus MV, an ellipse will be obtained. It has been observed that an ellipse will be distorted if the oscillation is caused by valve stiction. The method adopted here is based on the evaluation of how well the shape of the CV versus MV plot can be fitted by an ellipse. An empirical threshold of distance between each data point and the ellipse is used to determine the goodness-of-fit, and thereafter the valve stiction.

Process model validation monitor

The local approach based on the output error (OE) method [3] is employed to validate the nominal process model. This method applies to MISO systems. A MIMO system can be separated into several MISO subsystems. Models of each MISO part can be monitored with the local approach.

Disturbance model monitor

According to the assumption made before, the nominal model for the output disturbance, namely G_l , is available when the disturbance model validation is of concern. Multiplying the residual of the process model with inverse of the disturbance model yields the input to the disturbance model $\tilde{e}(t)$,

$$\tilde{e}(t) = G_l^{-1}[y(t) - \hat{y}(t)], \quad (2.14)$$

where $y(t)$ is the process output, and $\hat{y}(t)$ is the simulated output from nominal process model. If there is no mismatch in the disturbance model, the generated sequence should be white noise. Thus the disturbance model validation problem

can be transformed into a whiteness test problem. The index $\tilde{e}^T(t)R_{\tilde{e}}^{-1}\tilde{e}(t)$, which should follow χ^2 distribution, is used as the output of the disturbance dynamics monitor, where $R_{\tilde{e}}$ is variance of $\tilde{e}(t)$.

Sensor bias monitor

An analytical redundancy method which eliminates the unknown states is applied to detect sensor bias [77].

2.3.3 Diagnostic settings and results

Since the three quality CVs are of the main interest, the selected monitors mainly target these CVs, as shown in Table 2.2.

Table 2.2: Summary of monitors

Monitor	Description
π_1	Overall control performance monitor
π_2, π_3, π_4	Univariate control performance monitors for the three quality variables
π_5, π_6	Valve stiction monitors for the two possible problematic valves
π_7, π_8, π_9	Process model validation monitors for the three quality variables
$\pi_{10}, \pi_{11}, \pi_{12}$	Disturbance change monitors for the three quality variables
$\pi_{13}, \pi_{14}, \pi_{15}$	Sensor bias detection monitors for the three quality variables

The parameter settings of the Bayesian diagnostic system are summarized in Table 2.3. Note that *UC* represents unknown problems as well as combinations of two or more problems occurring simultaneously, so data from *PCT bottom sensor bias* mode, which represents unknown problems, and data from *simultaneous poorly tuned controller and pressure sensor bias* mode, which represents combination of two or more problems, are collected for the validation of *UC* mode.

Diagnostic results in Figure 2.4 are obtained from evaluation (cross-validation) data which are generated independently of historical samples. In Figure 2.4, the title of each plot denotes the true underlying mode, and the numbers on the horizontal axis stand for the diagnosed eleven possible modes numbered according to

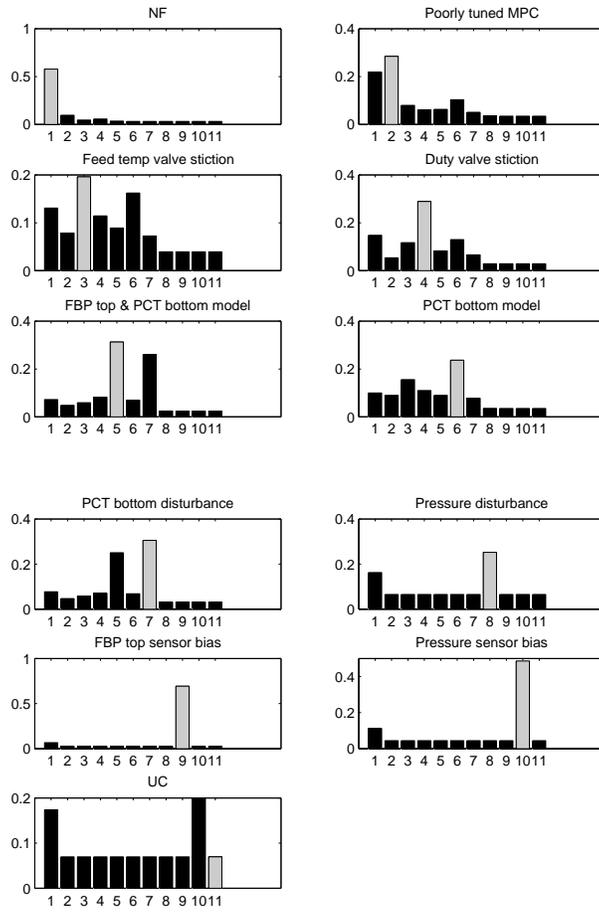


Figure 2.4: Posterior probability assigned to each mode

Table 2.3: Summary of Bayesian diagnostic parameters

Discretizaion	$k_i = 3(\text{“low”}, \text{“medium”}, \text{“high”}),$ $K = 3^{15} = 14348907$
Historical data	300 samples for each mode, except <i>UC</i>
Prior samples	Uniformly distributed with prior sample, $a_j = 1, A = 14348907$
Prior probabilities	$p(NF) = 0.1, p(m_{other}) = 0.09$
Evaluation data	300 samples for each mode, from training modes and <i>UC</i>

the sequence shown in the first column in Figure 2.1. In each plot, the posterior probability corresponding to the true underlying mode is highlighted with gray bars, while the others are in dark bars. The diagnostic conclusion is determined by selecting the mode with the largest posterior probability. If the largest probability happens to be the gray one, then the problem source is correctly identified. From Figure 2.4, we can see that all the true underlying modes are assigned with the largest posterior probabilities, except *UC*. Even in the presence of low-performance monitors, the Bayesian approach can synthesize information from these monitors to provide good diagnostic results. Performance of the diagnostic system for the *UC* mode, however, is poor as expected, owing to the lack of historical data for that mode.

2.4 Industry evaluation

The data-driven Bayesian diagnostic approach has been tested on an industry diluted oil pre-heater process.

2.4.1 Process description

The scheme diagram of the process is presented in 2.5. The function of this process is to heat the diluted oil from the upstream process to a desired temperature with a furnace. The diluted oil is fed into the furnace through eight passes, and the oil is heated within the coils in the furnace. The eight oil passes are mixed at the outlet, and then fed into downstream process. Flow control of the oil feed is provided for each of the eight heater passes with eight PID controllers, FIC1 to FIC8 respectively. The set-points of the flow PID controllers are set by a multivariate MPC controller to

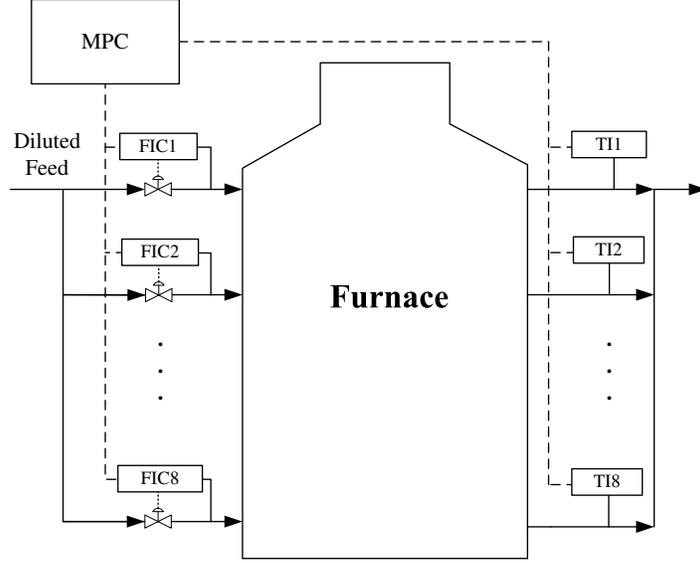


Figure 2.5: Structure of diluted oil pre-heater

control the temperatures of the eight passes at the outlet TI1 to TI8, i.e. coil outlet temperatures ($COTs$). The $COTs$ are controlled such that differences between the eight $COTs$ and their average COT_{ave} can be minimized, and the COT_{ave} is always within the limit range. This MPC application is known as the *pass balance*.

One of the flow control loops, FIC4 is subject to the problem of a sticky valve and frequent problematic PID control performance. Thus, the control system has three problematic modes: (1) valve stiction problem, (2) control tuning problem, and (3) simultaneous valve stiction and controller tuning problem. They share almost the same symptoms, such as process oscillation and large process variance. The interest is to isolate the problem source. The historical data we obtained contain the valve stiction mode and the simultaneous valve stiction and problematic controller tuning mode. The proposed Bayesian approach is used to synthesize monitor outputs to distinguish different problems with similar phenomena, so as to enhance the stiction detection and control performance monitoring.

2.4.2 Data-driven Bayesian diagnosis

Three monitors are chosen for the Bayesian diagnostic system, including the univariate control performance monitor π_1 , model validation monitor π_2 , and valve stiction monitor π_3 . The univariate MVC benchmark is employed to monitor the control performance of the difference between TI4 and COT_{ave} . The model validation mon-

itor π_2 uses the local output error approach to monitor the model change between the input FI4 and the output TI4. The ellipse fitting method with oscillation detection is utilized for the FIC4 valve stiction monitor. Each complete evidence consists of three monitor readings, i.e. $E = (\pi_1, \pi_2, \pi_3)$. Each single monitor reading is discretized into two values, “*abnormal*” and “*normal*”. Thus totally there are $2^3 = 8$ discrete evidence bins.

The historical data contain two problematic modes together with the normal operation mode. The two problematic modes are denoted as *Sticky* (sticky FIC4 valve problem only), and *SP* (simultaneous sticky FIC4 valve and problematic PID control). The normal operation mode is denoted as *NF* (normal functioning). It should be noted that the historical data are collected when there is no setpoint change or other major upsets, such that the process operation status is consistent with the defined mode. The sampling interval of process data is set to one minute. Each window of data consists of approximately 8-hour process data for a calculation of one monitor reading or one “historical sample”. The collected data of the three modes are divided into two parts. One part is for estimation of the likelihood, and the other is for cross validation of the Bayesian diagnostic system. Table 2.4 summarizes parameters for the Bayesian diagnosis.

Table 2.4: Summary of Bayesian diagnostic parameters

Evidence	$E = (\pi_1, \pi_2, \pi_3)$
Discretization	$k_i = 2, K = 2^3 = 8$
Historical data	41 samples for <i>NF</i> mode, 23 samples for <i>SP</i> mode, and 8 samples for <i>sticky</i> mode
Prior samples	Uniformly distributed with prior sample $a_j = 1, A = 8$
Prior probabilities	$p(NF) = p(SP) = p(sticky) = 1/3$
Evaluation data	10 samples for <i>NF</i> and <i>SP</i> mode, and 2 samples for <i>sticky</i> mode

With the data-driven Bayesian approach, the diagnostic results shown in Figure 2.6 are obtained for the cross validation data. In each plot, the posterior probability corresponding to the true underlying mode is shown with gray bars, while others are in dark bars. Thus, if the grey bar is highest then the correct diagnosis is obtained. From Figure 2.6, we can see that all the true underlying modes are assigned with

the largest probabilities, indicating correct diagnosis of the three modes.

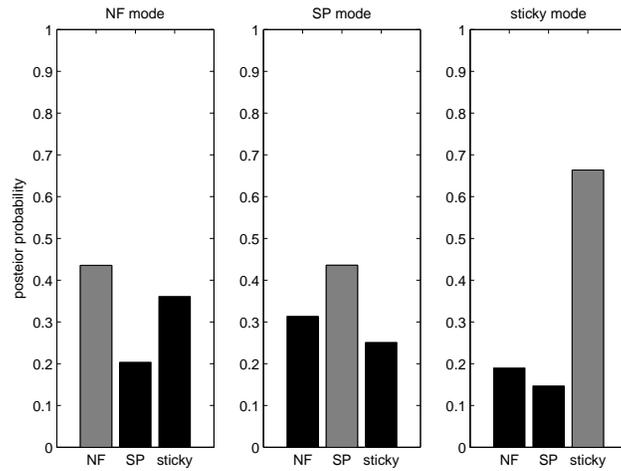


Figure 2.6: Posterior probability assigned to each mode

2.5 Conclusions

In this Chapter, a data-driven Bayesian approach is introduced for control loop diagnosis. The Bayesian methods are employed to synthesize control loop monitors and to isolate the underlying problem sources. Some Bayesian diagnosis concepts are adapted to fit the control loop diagnosis problem. The proposed method is verified by a simulated binary distillation column and an industrial process, where the features of the Bayesian approach to the synthesis of a variety of monitors are demonstrated.

Chapter 3

A Bayesian Approach for Control Loop Diagnosis with Incomplete Historical Evidence Data

A main issue encountered in the application of the data-driven Bayesian approach is the problem of missing monitor readings. By introducing the concept of missing pattern, incomplete problems are classified into single and multiple missing patterns. A novel method based on marginalization over underlying complete evidence matrix (UCEM) is proposed to circumvent the incomplete evidence problems. Performance of the proposed incomplete evidence handling approach is examined through simulations as well as an industrial application example.

3.1 Introduction

Although a data-driven Bayesian procedure for control loop diagnosis has been discussed in Chapter 2, practical problems remain. An outstanding problem of the procedure in the previous chapter is its inability to handle incomplete evidences. In the process industry environment, missing data is not an uncommon problem. Due to instrument reliability, heavy control network traffic, or historian storage problem, some key process variable measurements may not be available in the historian. Depending on the monitor calculation algorithms, the missing process variable will often lead to incomplete evidences where some of the monitors readings are missing. However, in the Bayesian diagnostic method discussed in Chapter 2, only com-

A version of this chapter has been published in F. Qi, B. Huang, and E.C. Tamayo. A Bayesian approach for control loop diagnosis with missing data. *AIChE Journal*, 56:179-195, 2010.

plete historical evidence samples can be used. If any monitor reading is missing, then all the other monitor readings sampled in the same evidence, have to be discarded, which will reduce the number of available historical samples and compromise the performance of the diagnosis. This can be very problematic in many practical applications where certain modes only appear infrequently. This problem will be addressed in this Chapter.

Missing data has been a popular topic in various research areas [40]. A large number of missing data reconstruction approaches have been developed, for instance, maximum likelihood (ML) estimation [15], EM algorithms [59, 85], multi-imputation [81, 92], generalized estimating equations [106], selection models [58, 101], etc. However, most of the available methods are computationally intensive, and cannot be readily fit into the Bayesian diagnostic framework. In this chapter, we develop a missing monitor handling strategy based on the marginalization of the incomplete evidence likelihood. The proposed approach requires much less computational power than the conventional methods. It is shown that the result obtained is the limit of the ML method, and thus better incomplete evidence reconstruction performance is expected from the proposed method.

The remainder of this chapter is organized as follows: The concepts of a missing pattern and UCEM are introduced in Section 3.2 to classify incomplete evidence problems into single and multiple missing pattern ones. The solution for evidence likelihood estimation in the presence of incomplete evidences is derived in Section 3.3. The proposed approach is applied to a simulation example and an industrial process in Sections 3.4 and 3.5. Section 3.6 concludes this chapter.

3.2 Incomplete evidence problem

In the following discussions, the subscript for denoting the mode M will be omitted for simplicity without causing confusion.

In the presence of missing monitor readings, the historical data set can be segregated into two parts, the complete evidence samples and the incomplete ones,

$$\mathcal{D} = \{\mathcal{D}_c, \mathcal{D}_{ic}\},$$

where \mathcal{D}_c is the data set with complete evidences, and the \mathcal{D}_{ic} is the remaining data set with incomplete evidences. The two data sets are therefore named complete data set and incomplete data set respectively.

Let a complete evidence consist of L monitor readings,

$$E = (\pi_1, \pi_2, \dots, \pi_L),$$

and each single monitor reading π_i has k_i different discrete values. In reality, the missing data problem may occur in any of these monitors. A concept named “missing pattern” needs to be introduced to describe and classify the monitor missing problems.

A missing pattern is determined by the locations of the missing monitor, i.e. how the missing monitor readings occur in an evidence. If two incomplete evidences have missing data from the same monitors, they belong to the same missing pattern, regardless of the value of the available monitor readings. For example, two evidence readings $(\times, \times, 0)$ and $(\times, \times, 1)$ belong to the same missing pattern, where \times denotes the missing value. Otherwise, they are said to fall into different missing patterns. For instance, the two evidence readings $(\times, 1, 0)$ and $(1, \times, 1)$ belong to two different missing patterns; it should be noted that two evidence readings $(\times, \times, 0)$ and $(1, \times, 1)$ are also from two distinct missing patterns. By enumerating the numbers of missing patterns in the historical data set, the incomplete evidence problems can be classified into single missing pattern ones and multiple missing pattern ones.

In a multiple missing pattern problem, the incomplete historical data set is divided into groups of different single missing patterns. For each single missing pattern, all the missing data are from the same monitors and missing data occur across these problematic monitors simultaneously.

Without loss of generality, assume that the first q monitor readings have missing data for a given missing pattern. Thus an incomplete evidence in this missing pattern can be represented as

$$E = (\times, \dots, \times, \pi_{q+1}, \dots, \pi_L).$$

Each of the available monitor readings π_{q+1}, \dots, π_L can take one of its k_i discrete values; thus there are in total

$$S = \prod_{i=q+1}^L k_i$$

different combinations of incomplete evidences. Each missing reading of the first q monitors could have been anyone of its possible output values; thus each incomplete evidence may have come from one of the $R = \prod_{i=1}^q k_i$ underlying complete evidences.

A unique underlying complete evidence matrix (UCEM) can be formed for each missing pattern. This matrix is constructed by enumerating all possible incomplete evidences in a column in front of the matrix, and then listing all possible underlying complete evidences corresponding to each incomplete evidence in the same row. Therefore the size of a UCEM is $S \times R$. Such a matrix looks like, for instance, for one of single missing patterns

$$\begin{matrix} (\times, \times, 0) \\ (\times, \times, 1) \end{matrix} \begin{bmatrix} (0, 0, 0) & (0, 1, 0) & (1, 0, 0) & (1, 1, 0) \\ (0, 0, 1) & (0, 1, 1) & (1, 0, 1) & (1, 1, 1) \end{bmatrix}. \quad (3.1)$$

Each row of UCEM contains all possible underlying complete evidences corresponding to an incomplete evidence. All elements of this matrix are unique, so there should be no coincidence between any two different rows in a UCEM. Therefore, all underlying complete evidences can be located in the UCEM uniquely.

In the following derivation, first consider the k th single missing pattern corresponding to a UCEM (denoted as k th UCEM); the underlying complete evidence with location (i, j) in the k th UCEM is denoted as $\epsilon_{i,j/k}$; its likelihood parameter is denoted as $\theta_{i,j/k}$; the number of historical samples with this underlying complete evidence is denoted as $\eta_{i,j/k}$. The corresponding incomplete evidence is denoted as $\epsilon_{i/k}$; its likelihood is denoted as $\lambda_{i/k}$; the historical sample number of this incomplete evidence is denoted as $\eta_{i/k}$. For instance, in the UCEM in Equation 3.1, which is assumed to be the first UCEM, the evidence $(0,0,0)$ is denoted as $\epsilon_{1,1/1}$; the number of historical samples with this evidence is denoted as $\eta_{1,1/1}$. The corresponding incomplete evidence, $(\times, \times, 0)$, is denoted as $\epsilon_{1/1}$; the historical sample number of $(\times, \times, 0)$ is $\eta_{1/1}$, and its likelihood is $\lambda_{1/1}$.

3.3 Diagnosis with incomplete evidence

Recall the posterior calculation Equation 2.3

$$p(M|E, \mathcal{D}) \propto p(E|M, \mathcal{D})p(M). \quad (3.2)$$

As discussed in Chapter 2, prior probability $p(M)$ is determined by *a priori* information. Thus our interest remains in how to derive the likelihood of historical data $p(\mathcal{D}|\Theta, M)$ in the presence of incomplete evidences.

When there are only complete historical samples, $p(\mathcal{D}|\Theta, M)$ can be calculated with method outlined in Chapter 2. Missing data occur from the problematic monitor readings, and the rest of the monitor outputs still provide partial information of

evidence. Thus, the incomplete samples need to be taken into consideration when evaluating the likelihood. The likelihood probability of an incomplete data sample $d_{ic}^t = (\epsilon_{i/k}, M)$ equals the likelihood of the incomplete evidence $\epsilon_{i/k}$,

$$p(d_{ic}^t|\Theta, M) = p((\epsilon_{i/k}, M)|\theta, M) = p(\epsilon_{i/k}|\Theta, M) = \lambda_{i/k}, \quad (3.3)$$

which is the summation of the probabilities of all the possible underlying complete evidences, i.e., marginalization of the likelihood over all possible complete evidences in the i th row of the k th UCEM,

$$p((\epsilon_{i/k}, M)|\theta, M) = \sum_{j=1}^R p(\epsilon_{i,j/k}|\Theta, M) = \sum_{j=1}^R \theta_{i,j/k}. \quad (3.4)$$

Take the UCEM in Equation 3.1 as an example. The likelihood of incomplete evidence $(\times, \times, 0)$ equals the summation of likelihood from $(0, 0, 0)$ to $(1, 1, 0)$ in the first row,

$$p((\epsilon_{1/1}, M)|\Theta, M) = \lambda_{1/1} = \sum_{j=1}^4 \theta_{1,j/1}. \quad (3.5)$$

3.3.1 Single missing pattern problem

For a single missing pattern problem, where only one UCEM exists, the likelihood probability over the historical data set, including both complete and incomplete samples, is

$$\begin{aligned} p(\mathcal{D}|\Theta, M) &= p(\mathcal{D}_c|\Theta, M)p(\mathcal{D}_{ic}|\Theta, M) \\ &= \prod_{i=1}^S \prod_{j=1}^R \theta_{i,j}^{\eta_{i,j}} \cdot \prod_{i=1}^S \lambda_i^{\eta_i} = \prod_{i=1}^S \left[\prod_{j=1}^R \theta_{i,j}^{\eta_{i,j}} \cdot \left(\sum_{k=1}^R \theta_{i,k} \right)^{\eta_i} \right]. \end{aligned} \quad (3.6)$$

Note that since only one UCEM exists for the single missing pattern problem, the subscript denoting the number of UCEM is omitted here without causing confusion.

Again, let the prior distribution of likelihood parameters be Dirichlet distributed with parameters $a_{i,j}$ for the complete evidences [73],

$$f(\Theta|M) = \frac{\Gamma(\sum_{i=1}^S \sum_{j=1}^R a_{i,j})}{\prod_{i=1}^S \prod_{j=1}^R \Gamma(a_{i,j})} \prod_{i=1}^S \prod_{j=1}^R \theta_{i,j}^{a_{i,j}-1}. \quad (3.7)$$

Then,

$$\begin{aligned}
f(\Theta|M, \mathcal{D}) &= \frac{p(\mathcal{D}|\Theta, M)f(\Theta|M)}{p(\mathcal{D}|M)} \\
&= \frac{1}{p(\mathcal{D}|M)} \cdot \prod_{i=1}^S \left[\prod_{j=1}^R \theta_{i,j}^{\eta_{i,j}} \cdot \left(\sum_{k=1}^R \theta_{i,k} \right)^{\eta_i} \right] \\
&\quad \cdot \frac{\Gamma(\sum_{i=1}^S \sum_{j=1}^R a_{i,j})}{\prod_{i=1}^S \prod_{j=1}^R \Gamma(a_{i,j})} \prod_{i=1}^S \prod_{j=1}^R \theta_{i,j}^{a_{i,j}-1}. \tag{3.8}
\end{aligned}$$

Let

$$c = \frac{\Gamma(\sum_{i=1}^S \sum_{j=1}^R a_{i,j})}{\prod_{i=1}^S \prod_{j=1}^R \Gamma(a_{i,j})},$$

and then Equation 3.8 can be written as

$$\begin{aligned}
f(\Theta|M, \mathcal{D}) &= \frac{c}{p(\mathcal{D}|M)} \cdot \prod_{i=1}^S \left[\prod_{j=1}^R \theta_{i,j}^{\eta_{i,j}} \cdot \left(\sum_{k=1}^R \theta_{i,k} \right)^{\eta_i} \right] \cdot \prod_{i=1}^S \prod_{j=1}^R \theta_{i,j}^{a_{i,j}-1} \\
&= \frac{c}{p(\mathcal{D}|M)} \cdot \prod_{i=1}^S \left[\prod_{j=1}^R \theta_{i,j}^{\eta_{i,j}+a_{i,j}-1} \cdot \left(\sum_{k=1}^R \theta_{i,k} \right)^{\eta_i} \right] \\
&= \frac{c}{\int_{\Omega} p(\mathcal{D}|\Theta, M)f(\Theta|M)d\Theta} \cdot \prod_{i=1}^S \left[\prod_{j=1}^R \theta_{i,j}^{\eta_{i,j}+a_{i,j}-1} \cdot \left(\sum_{k=1}^R \theta_{i,k} \right)^{\eta_i} \right]. \tag{3.9}
\end{aligned}$$

With derivations presented in Appendix A, the likelihood of evidence $\epsilon_{s,r}$ can be determined as

$$\begin{aligned}
p(\epsilon_{s,r}|M, \mathcal{D}) &= \frac{\eta_{s,r} + a_{s,r}}{N + A} \cdot \frac{\sum_{j=1}^R (\eta_{s,j} + a_{s,j}) + \eta_s}{\sum_{j=1}^R (\eta_{s,j} + a_{s,j})} \\
&= \frac{\eta_{s,r} + a_{s,r}}{N + A} \cdot \left(1 + \frac{\eta_s}{\sum_{j=1}^R (\eta_{s,j} + a_{s,j})} \right), \tag{3.10}
\end{aligned}$$

where $N = \sum_i \eta_i + \sum_i \sum_j \eta_{i,j}$ is the total number of historical data samples for mode M , including both complete and incomplete samples; $A = \sum_i \sum_j a_{i,j}$ is the total number of prior samples, which is, however, only applicable to the complete evidences.

Let us use an example to illustrate the likelihood calculation. Suppose that there are two monitors π_1 and π_2 , both with possible output 0 and 1, and π_2 is missing in part of the historical data samples. Using the complete samples only, the likelihood of evidence (0,0), for example, is

$$p((0,0)|M, \mathcal{D}) = \frac{\eta_{(0,0)} + a_{(0,0)}}{N_c + A}, \tag{3.11}$$

where $\eta_{(0,0)}$ is the number of complete data samples with evidence $(0,0)$, and $a_{(0,0)}$ is the prior sample number for evidence $(0,0)$; N_c is the number of complete historical data samples. When the incomplete data samples are also used, according to Equation 3.10, the likelihood becomes

$$\begin{aligned} & p((0,0)|m, \mathcal{D}) \\ &= \frac{\eta_{(0,0)} + a_{(0,0)}}{N + A} \cdot \left(1 + \frac{\eta_{(0,\times)}}{\sum_{\pi_2} \eta_{(0,\pi_2)} + a_{(0,\pi_2)}} \right), \end{aligned} \quad (3.12)$$

where $\eta_{(0,\times)}$ is the number of samples with incomplete evidence pattern $(0, \times)$.

The estimated likelihood in Equation 3.10 has an intuitive explanation. It can be rewritten as

$$p(\epsilon_{s,r}|M, \mathcal{D}) = \frac{\eta_{s,r} + a_{s,r} + \eta'_{s,r}}{N + A}, \quad (3.13)$$

where

$$\eta'_{s,r} = \eta_s \cdot \frac{\eta_{s,r} + a_{s,r}}{\sum_{j=1}^R (\eta_{s,j} + a_{s,j})} \quad (3.14)$$

can be interpreted as the expected number of samples with evidence $\epsilon_{s,r}$ in the incomplete data set. Interestingly, this number equals the expected value of a variable subject to Dirichlet distribution with parameters $(\eta_{s,1} + a_{s,1}, \eta_{s,2} + a_{s,2}, \dots, \eta_{s,R} + a_{s,R})$. As a result, the expected number of the underlying complete evidences in the incomplete data set can be estimated from the distribution of the complete evidences. See Figure 3.1 for illustration, where the sensor bias monitor reading is assumed to be missing in some historical samples. The incomplete data samples are located on the boundary between “biased” and “unbiased” zone; namely the underlying missing values could be sensor bias or unbiased. With totally two samples in the evidence bin labeled with $(no\ stiction, unbiased)$, and three samples in the evidence bin labeled with $(no\ stiction, biased)$, the expected numbers of the underlying complete evidences in the incomplete samples can be calculated by Equation 3.14, as shown in Figure 3.1. Namely, among the four incomplete evidences, 1.6 of them are expected to be in the bin $(no\ stiction, unbiased)$ and 2.4 of them are expected to be in the bin $(no\ stiction, biased)$.

The estimated sample numbers from the incomplete data set, together with the numbers of complete data samples and prior samples, are used to calculate the final likelihood according to Equation 3.13.

In summary, the following routine can be developed to estimate the likelihood from historical data with incomplete evidences.

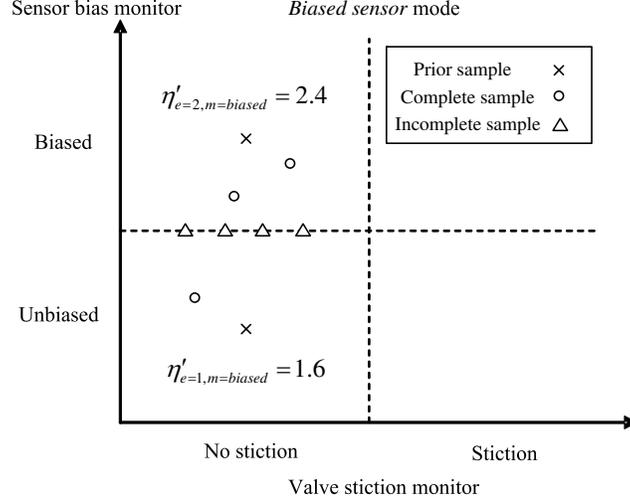


Figure 3.1: Estimation of expected complete evidence numbers out of the incomplete samples

- (1) *Construct a UCEM for the incomplete data samples;*
- (2) *Calculate the expected numbers of all possible underlying evidences $\eta'_{s,r}$ according to Equation 3.14;*
- (3) *Calculate the likelihood according to Equation 3.13.*

By considering the incomplete samples, how will the likelihood be changed? Consider a data set with incomplete samples. Table 3.1 contains the details of the data set. The prior samples are uniformly distributed with one for each complete evidence.

Table 3.1: Summary of historical and prior samples

Evidence	(0,0)	(0,1)	(1,0)	(1,1)	(0,x)	(1,x)
Historical samples	6	5	4	8	20	4
Prior samples	1	1	1	1	N/A	N/A

Compare the two different incomplete evidence handling strategies, i.e. omitting all the incomplete evidences, and considering the incomplete evidences according to the proposed approach. The results are displayed in Table 3.2.

With the difference being significant, it is demonstrated that the information from the incomplete evidences is useful for the estimation of likelihood. Simply

Table 3.2: Estimated likelihood

Evidence	(0,0)	(0,1)	(1,0)	(1,1)
Likelihood(discarded)	0.2593	0.2222	0.1583	0.3333
Likelihood(marginalization)	0.3484	0.2989	0.1261	0.2269

omitting all the incomplete evidences will lose information for the likelihood estimation.

An interesting question at this point is when the proposed approach does not change the likelihood estimation, i.e., generating the same likelihood as that when only the complete evidences are considered. In view of Equation 3.10, if the *missing ratio*

$$\rho_s^{miss} = \frac{\eta_s}{\sum_{j=1}^R (\eta_{s,j} + a_{s,j})} \quad (3.15)$$

is identical for all rows in UCEM, namely for different s , the likelihood of evidence $\epsilon_{s,r}$ is

$$\begin{aligned} & p(\epsilon_{s,r} | M, \mathcal{D}) \\ &= \frac{1}{N + A} \cdot \left[(\eta_{s,r} + a_{s,r}) + \eta_s \cdot \frac{\eta_{s,r} + a_{s,r}}{\sum_{j=1}^R (\eta_{s,j} + a_{s,j})} \right] \\ &= \frac{\eta_{s,r} + a_{s,r}}{N + A} \cdot [1 + \rho_s^{miss}] \\ &= \frac{\eta_{s,r} + a_{s,r}}{\sum_{s,j} (\eta_{s,j} + a_{s,j}) + \rho_s^{miss} \sum_{s,j} (\eta_{s,j} + a_{s,j})} \cdot [1 + \rho_s^{miss}] \\ &= \frac{\eta_{s,r} + a_{s,r}}{\sum_{s,j} (\eta_{s,j} + a_{s,j})} = \frac{\eta_{s,r} + a_{s,r}}{N_c + A}, \end{aligned} \quad (3.16)$$

which is the same as the result obtained with only consideration of the complete evidences. Thus, it can be concluded that if the missing ratios ρ_s^{miss} are the same in all the incomplete evidences, consideration of incomplete evidences will not introduce extra information, and hence the same likelihood will be obtained. For example, consider a set of historical evidences summarized in Table 3.3, with one prior sample being assigned to each complete evidence.

According to the definition of missing ratio in Equation 3.15,

$$\rho_{(0,\times)}^{miss} = \frac{7}{(2+1) + (3+1)} = 1,$$

and

$$\rho_{(1,\times)}^{miss} = \frac{11}{(4+1) + (5+1)} = 1.$$

Table 3.3: Summary of historical and prior samples

Evidence	(0,0)	(0,1)	(1,0)	(1,1)	(0,×)	(1,×)
Historical sample	2	3	4	5	7	11
Prior sample	1	1	1	1	N/A	N/A

In this example, introduction of the incomplete evidences will not change the estimation of the likelihood. Other than this special case, the proposed method will in general generate different likelihood estimation from that obtained by merely using the complete evidences.

3.3.2 Multiple missing pattern problem

The solution for the multiple missing pattern is more complex. In this section, we will sketch the derivation and present the general solution.

When there are more than one missing pattern in the historical data set, several UCEMs exist. Suppose that a historical data set, where each complete evidence contains two monitor readings, $E = (\pi_1, \pi_2)$, has two missing patterns (\times, π_2) and (π_1, \times) . Let both π_1 and π_2 have two possible values 0 and 1. Two UCEMs are constructed for the historical data set:

$$\begin{matrix} (\times, 0) \\ (\times, 1) \end{matrix} \left[\begin{matrix} (0, 0) & (1, 0) \\ (0, 1) & (1, 1) \end{matrix} \right], \quad (3.17)$$

and

$$\begin{matrix} (0, \times) \\ (1, \times) \end{matrix} \left[\begin{matrix} (0, 0) & (0, 1) \\ (1, 0) & (1, 1) \end{matrix} \right]. \quad (3.18)$$

Both UCEMs need to be taken into consideration when calculating the likelihood of historical data, including both complete and incomplete samples. Since all complete evidences can be located in either one of the two UCEMs, we only need to use one to denote the likelihood of underlying complete evidences. The likelihood parameter for complete evidence (0,0), for example, is represented as $\theta_{1,1/1}$ according to the UCEM in Equation 3.17.

In general, assume there are P missing patterns in the historical data set, the

likelihood of the data set is calculated as

$$\begin{aligned}
p(\mathcal{D}|\Theta, M) &= p(\mathcal{D}_c|\Theta, M)p(\mathcal{D}_{ic}|\Theta, M) \\
&= \prod_{i=1}^{S_1} \prod_{j=1}^{R_1} \theta_{i,j/1}^{\eta_{i,j/1}} \cdot \prod_{k=1}^P \prod_{i=1}^{S_k} \lambda_{i/k}^{\eta_{i/k}} = \prod_{i=1}^{S_1} \prod_{j=1}^{R_1} \theta_{i,j/1}^{\eta_{i,j/1}} \cdot \prod_{k=1}^P \prod_{i=1}^{S_k} \left(\sum_{j=1}^{R_k} \theta_{i,j/k} \right)^{\eta_{i/k}},
\end{aligned} \tag{3.19}$$

where the size of k th UCEM is $S_k \times R_k$. Considering the fact that all the complete evidences can be located in a single UCEM uniquely, each complete evidence $\epsilon_{i,j/k}$ must have one and only one match in the first UCEM. As such, the first UCEM can be used as the target, to project all the marginalization of the incomplete evidences onto it.

Following the similar derivation procedure as developed for single missing pattern case, the likelihood of complete evidence $\epsilon_{s,r/1}$ can be derived as

$$\begin{aligned}
p(\epsilon_{s,r/1}|M, \mathcal{D}) &= \frac{1}{N + A} \\
&\cdot \frac{\sum_{\mathcal{T}^1 \dots \mathcal{T}^P} \prod_{k=1}^P \prod_{i=1}^{S_k} \left[\mathbf{C}_{\eta_{i/k}}^{\mathcal{T}^k} \Gamma(\sum_{j=1}^P \mathcal{Z}_{s,r} + 1) \prod_{(u,v) \neq (s,r)} \Gamma(\mathcal{Z}_{u,v}) \right]}{\sum_{\mathcal{T}^1 \dots \mathcal{T}^P} \prod_{k=1}^P \prod_{i=1}^{S_k} \left[\mathbf{C}_{\eta_{i/k}}^{\mathcal{T}^k} \prod_{u=1}^{S_1} \prod_{v=1}^{R_1} \Gamma(\mathcal{Z}_{u,v}) \right]}
\end{aligned} \tag{3.20}$$

where \mathcal{T}^i is all the combination of \mathcal{T}_j for the i th missing pattern; $\mathcal{Z}_{s,r} = \sum_{j=1}^P \tilde{t}_{s,r/j} + \eta_{s,r/1} + a_{s,r/1}$; $\tilde{t}_{s,r/j}$ is the possible number of complete evidence $\epsilon_{s,r/1}$ in the j th missing pattern; $\eta_{s,r/1}$ and $a_{s,r/1}$ are the historical counts and prior counts of the complete evidence $\epsilon_{s,r/1}$ respectively; and $N = \sum_{i,j} \eta_{i,j/1} + \sum_{i,k} \eta_{i/k}$, $A = \sum_{i,j} a_{i,j/1}$.

In view of the general solution for multiple missing patterns presented in Equation 3.20, we can see that the computation load increases exponentially with the growth of missing patterns. Thus simplification or approximation of the likelihood calculation is necessary for a computationally affordable solution.

It is however worthy to point out that the single missing pattern solution does have its practical merit whether the data has single missing pattern or multiple missing patterns. In general, multiple missing patterns do exist. The evidence of a typical process consists of a large number of monitor readings but not all these variables suffer from the missing data problem. In other words, there are some monitors that have the missing data problem. If data missing does not occur simultaneously across all the problematic variables, a multiple missing pattern occurs, but we can still apply the single missing pattern method to get an approximate solution. In this case, a single missing pattern which includes all monitors that have missing

data problems is constructed. All multiple missing patterns can be fitted into this single missing pattern by discarding some monitor readings. For instance, consider that an evidence of a process contains five monitors and two of them have missing data problem, say monitor 1 and 2. If in an instant, one monitor reading has missing data, say monitor 1, but the other one (monitor 2) has a reading at this instant (i.e. multiple missing pattern occurs). The traditional method would ignore readings of all five monitor readings at this instant. To apply the proposed single missing-pattern method, however, one only needs to omit the reading of monitor 2, but the readings from monitors 3-5 can still be used, rather than omitting all five monitor readings. An improvement is therefore expected. Comparing to the computation load of applying full missing pattern method, this single missing pattern approximation can be a valuable alternative. We will demonstrate its application to a multiple missing pattern problem in the next section.

3.4 Simulation example

To investigate diagnostic performance of the proposed Bayesian approach for a process with advanced process control scheme, we apply the algorithm to the simulated binary distillation column described in Section 2.3. The process and diagnostic settings are the same as Section 2.3, and hence will not be repeated here.

3.4.1 Single missing pattern

Assume that duty valve stiction monitor reading π_6 and process model monitor reading π_9 have missing data in some historical samples collected under m_9 , *pressure sensor bias* mode. The incomplete evidence can be denoted as

$$e = (\pi_1, \dots, \pi_5, \times, \pi_7, \pi_8, \times, \pi_{10}, \dots, \pi_{15}).$$

In the simulation, let that π_6 and π_9 tend to be missing when the discrete output of the pressure sensor bias monitor reading is “high”. π_6 and π_9 have 0.9 probability to be missing under such a condition. When the discrete output of the pressure sensor bias monitor indicates “low” or “medium”, the probability that the readings of monitors π_6 and π_9 are missing is 0.1. By analyzing the original historical data set for m_9 mode, it can be observed that π_9 in most of the samples indicates “high”. While most of such data samples are incomplete, the distribution of the complete

evidences for m_9 is distorted as shown in Figure 3.2, which will deteriorate the diagnostic performance.

Diagnosis with the proposed approach

We use the proposed approach for single missing pattern to tackle this problem. A UCEM is constructed, and the expected number of each possible underlying complete evidence is calculated according to Equation 3.14. These estimated samples are added to the samples with complete evidences as the new set of historical data. We may regard this procedure as recovery of the incomplete data with the proposed approach.

Consider a figure with the horizontal axis indicating different realizations of complete evidences (different combinations of each monitor readings to form evidences that appeared in data records) and the vertical axis indicating the numbers of occurrences in the historical data set of the corresponding evidences. The original data set (no missing monitor problem), the set with incomplete evidences, and the recovered data set are shown in Figure 3.2. It should be noted that the missing evidences which do not appear in the set with incomplete set are now recovered and displayed in Figure 3.2 together with the original ones.

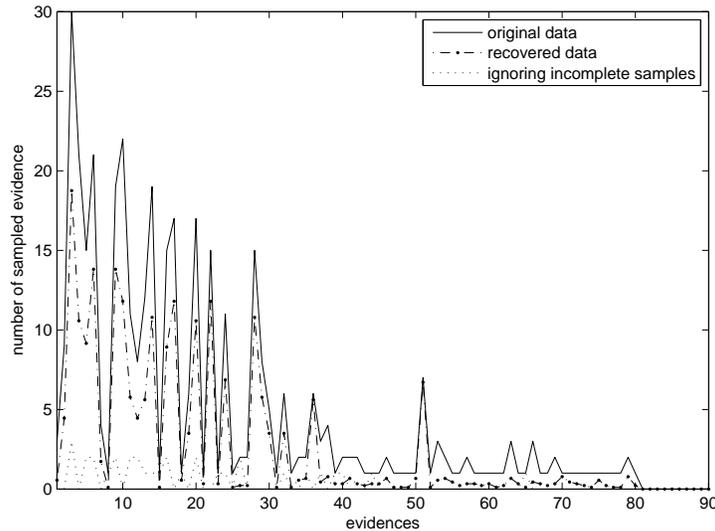


Figure 3.2: Comparison of complete evidence numbers

Clearly, the evidence distribution is seriously distorted if only complete samples

in the set with incomplete evidences are considered. Although the proposed approach can not recover all underlying missing evidences, i.e. to make the dotted line in Figure 3.2 coincide with the solid one, the dotted line (recovered data) can follow the overall trend of the original data well, which implies that likelihood distribution is well recovered. Further, compare the diagnostic results of m_9 mode using the original data set, and the diagnostic result from the set with incomplete evidences using two different strategies, i.e. (1) simply ignoring all the incomplete samples and (2) using the proposed approach. The comparison results are displayed in Figure 3.3. By ignoring all the incomplete samples, the posterior probability assigned to mode m_9 , which is the true underlying mode, is only 0.1606. This probability is not the largest one assigned, which means that the diagnostic system generates an erroneous result. By using the proposed approach, the posterior assigned to m_9 mode is 0.3736, which is the largest probability assigned to all the potential modes. Thus the diagnostic system generates correct result with the proposed approach.

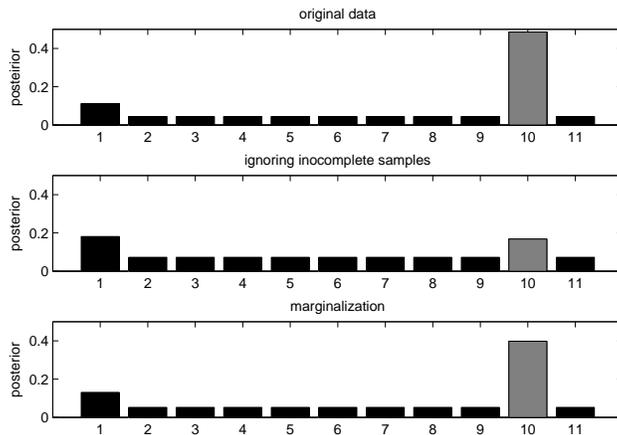


Figure 3.3: Diagnostic results with different data set

To further examine the performance of the proposed approach, Monte-Carlo simulations are performed. Totally 1000 runs are performed for the above simulations. The mean values of the posterior probabilities using original data and incomplete data set with two incomplete evidence handling strategies as discussed above are displayed in Figure 3.4. In comparison of Figures 3.3 and 3.4, the posteriors in two figures are very close, indicating small deviation of each simulation from mean values. The standard deviations of the Monte-Carlo simulations are summarized in

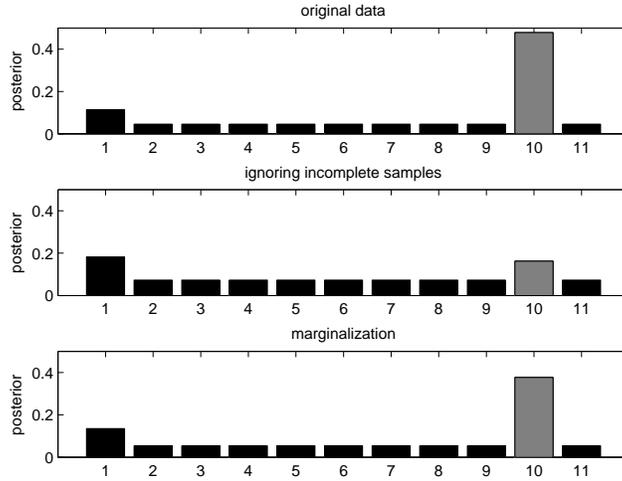


Figure 3.4: Mean value of posteriors with Monte Carlo simulations

Figure 3.5. The small standard deviations of the Monte-Carlo simulations indicate once again consistent results of each simulation.

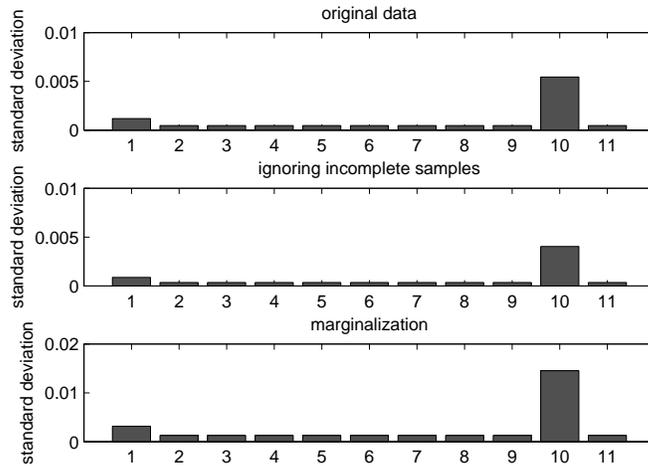


Figure 3.5: Standard deviations of posteriors with Monte Carlo simulations

In the histogram of the posterior probabilities assigned to mode m_9 shown in Figure 3.6, one can see that the posteriors obtained from original data set are always the highest. Although the posteriors calculated from incomplete data set with the proposed approach are smaller than those by complete data, they are all higher than the results when the incomplete samples are simply discarded. In summary, it can

be concluded that the proposed approach has improved diagnostic performance in the presence of incomplete evidences.

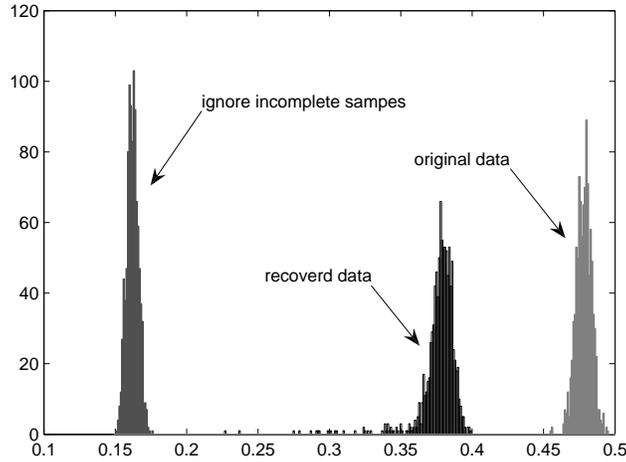


Figure 3.6: Histogram of posteriors assigned to m_9

Gibbs sampling method

The Gibbs sampling method is selected as an example to compare the performance of the proposed incomplete evidence handling strategy with those of traditional ones. Gibbs sampling is a Monte-Carlo Markov chain (MCMC) method introduced by Geman and Geman (1984) [24]. It is capable of estimating the parameters in the case of missing data, and thereby reconstruct the missing data according to the estimated parameters. Readers can refer to Korb and Nicholson (2004) [50] for details of the algorithm. A software named BUGS (Bayesian inference Using Gibbs Sampling) [1] is utilized in this work to perform the Gibbs sampling.

By plotting the figure with the horizontal axis indicating different realizations of complete evidences and the vertical axis indicating the numbers of occurrences in the historical data set of the corresponding evidences, the original data set (complete data), the recovered data with the proposed approach and the recovered data with Gibbs sampling are shown in Figure 3.7. It can be observed that the evidence distribution recovered with Gibbs sampling by a large number of iterations (4000 iterations in this example) converges to that recovered with the proposed approach.

If the numerical iterations are not sufficient, however, the result obtained by

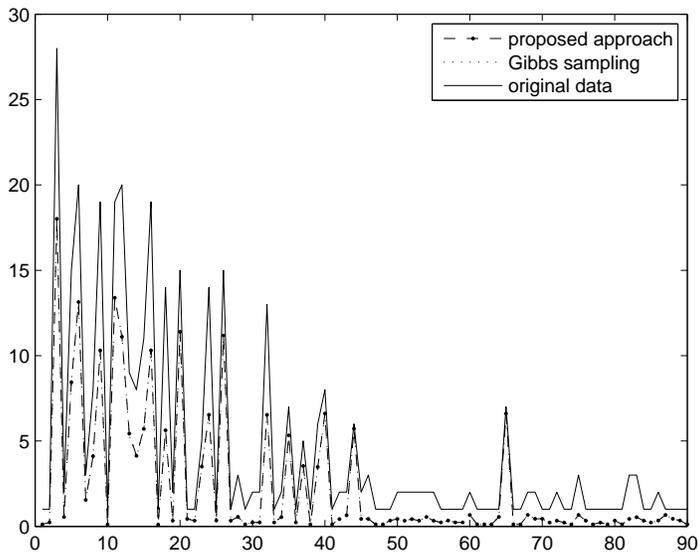


Figure 3.7: Comparison of complete evidence numbers

Gibbs sampling method will deviate more from that obtained by the proposed method. Figure 3.8 shows the evidence reconstruction error in terms of differences between the number of occurrence of the evidences in the original data set and the number of occurrence of the evidences after reconstructions by the two methods as discussed. The closer the difference from zero, the better the reconstruction performance is. It is observed that with insufficient iterations (10 iterations in this example), Gibbs sampling method yields a larger reconstruction error than the proposed method. In other words, the performance of Gibbs sampling method is by all means no better than that of the proposed approach.

The above results are expected. In the historical data set for a single mode, evidences are subject to categorical distribution with parameters p_1, p_2, \dots, p_Q , where Q is the total number of all possible evidences. In the Gibbs sampling method, it is these parameters that are estimated first, and then the missing data are reconstructed.

Recall that the elements between the rows of a UCEM are independent of each other, which follow the categorical distribution with parameters $[\frac{p_{i,1}}{\sum_j p_{i,j}}, \dots, \frac{p_{i,R}}{\sum_j p_{i,j}}]$. That is, for an evidence $\epsilon_{s,r}$, it follows categorical distribution $[\theta_{s,1}, \dots, \theta_{s,R}]$, where $\theta_{s,i} = p_{s,i} / \sum_j p_{s,j}$.

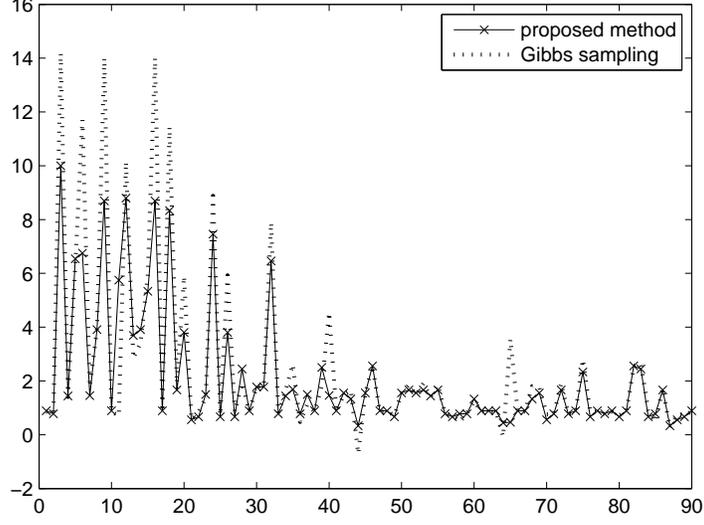


Figure 3.8: Reconstruction error of two data recovery methods

The procedures of the Gibbs sampling in BUGS can be described as following: Suppose that there are several parameters $\alpha_1, \dots, \alpha_N$ to estimate. With hierarchy distribution given, sample through all the parameters,

$$\begin{aligned} \alpha_1 &\sim p(\alpha_1 | \tilde{\alpha}_1, \mathcal{D}) \\ &\vdots \\ \alpha_N &\sim p(\alpha_N | \tilde{\alpha}_N, \mathcal{D}), \end{aligned}$$

where $\tilde{\alpha}_i = \{\alpha_1, \dots, \alpha_N\} \setminus \alpha_i$ is the parameter set excluding α_i , and \mathcal{D} is all the available data.

Consider the parameters for the distribution of evidences in the s th row in the UCEM, $\theta = (\theta_{s,1}, \dots, \theta_{s,R})$. Since $\theta_{s,i}$ are not independent of each other, θ has to be estimated as a whole:

$$P(\theta | \mathcal{D}) = \frac{P(\mathcal{D} | \theta) P(\theta)}{\int_{\Omega} P(\mathcal{D} | \theta) P(\theta) d\theta}, \quad (3.21)$$

where $P(\mathcal{D} | \theta) = \prod_{i=1}^R \theta_{(s,i)}^{n_{s,i}}$; $P(\theta)$ is assigned with Dirichlet distribution with parameters $[a_{s,1}, \dots, a_{s,L}]$, namely

$$f(\theta) = \frac{\Gamma(\sum_{j=1}^R a_{s,j})}{\prod_{j=1}^R \Gamma(a_{s,j})} \prod_{j=1}^R \theta_{s,j}^{a_{s,j}-1}.$$

According to the previous derivations in the evidence likelihood computation, the distribution of θ is

$$P(\theta|\mathcal{D}) = \frac{\Gamma(\sum_{i=1}^R \eta_{s,i} + a_{s,i})}{\prod_{i=1}^R \Gamma(\eta_{s,i} + a_{s,i})} \cdot \prod_{i=1}^R \theta_{s,i}^{\eta_{s,i} + a_{s,i} - 1}, \quad (3.22)$$

which is actually a Dirichlet distribution with parameter $(\eta_{s,1} + a_{s,1}, \dots, \eta_{s,R} + a_{s,R})$. Based on the samples from this distribution in each iteration step, the missing data are then reconstructed. Hence, more iteration steps mean more samples from the distribution of the missing data, which implies that the fitted distribution is closer to the real distribution.

Note that the expected value of a Dirichlet distribution is

$$E[\theta_{s,r}] = \frac{\eta_{s,r} + a_{s,r}}{\sum_{i=1}^R \eta_{s,i} + a_{s,i}},$$

so the expected number of evidence $\epsilon_{s,r}$ in the incomplete evidence samples is

$$\eta_s \cdot \frac{\eta_{s,r} + a_{s,r}}{\sum_{i=1}^R \eta_{s,i} + a_{s,i}}.$$

With this reconstructed incomplete evidences, the likelihood can be computed as

$$\begin{aligned} & \frac{1}{N + A} \cdot \left(\eta_s \frac{\eta_{s,r} + a_{s,r}}{\sum_{i=1}^R \eta_{s,i} + a_{s,i}} + \eta_{s,r} + a_{s,r} \right) \\ &= \frac{\eta_{s,r} + a_{s,r}}{N + A} \left(1 + \frac{\eta_s}{\sum_{i=1}^R \eta_{s,i} + a_{s,i}} \right), \end{aligned} \quad (3.23)$$

which is the likelihood equation we derived for the single missing pattern problem using marginalization. Therefore, we can conclude that Gibbs sampling method simply produces sufficient samples from a numerous iterations of simulations, and estimates parameters of the distribution from the sampled data, which is expectation in our example. The solution developed in this chapter is the limit that the Gibbs sampling method tries to achieve through iterative numerical simulations. Our method is an analytical solution with straightforward computation and hence the advantage is obvious.

3.4.2 Approximation of multiple missing pattern solution

As discussed in previous sections, the solution of single missing pattern problem provides a viable alternative in the presence of multiple missing patterns. An example is provided here to further demonstrate the idea.

In the data missing case studied so far, two monitor readings π_4 and π_9 are assumed to be unavailable simultaneously if missing occurs, and hence there is only one missing pattern considered. This problem can be modified into a multiple missing pattern problem by removing the assumption that the two monitors must be unavailable simultaneously if missing occurs. Thus there can be three different missing patterns in the evidence data set: missing π_4 only, missing π_9 only, and missing π_4 and π_9 simultaneously. Let π_6 or π_9 tend to be missing when the discrete output of the pressure sensor bias monitor reading is “high”; π_6 or π_9 has 90% chance to miss under such a condition. When the discrete output of the pressure sensor bias monitor indicates “low” or “medium”, the chance that the readings of monitors π_6 or π_9 are missing is 10%.

By applying the single missing-pattern method, we only need to deal with the single missing pattern with missing π_4 and π_9 simultaneously, i.e., omitting some available monitor reading π_4 or π_9 whenever the other one is missing. As such, all the incomplete evidences are categorized into a single missing pattern $(\pi_1, \dots, \pi_5, \times, \pi_7, \pi_8, \times, \pi_{10}, \dots, \pi_{15})$. Not surprisingly, the diagnostic results by this approximation are the same as those presented in Figures 3.2 and 3.3, and it has been shown they are considerably better than simply ignoring an entire evidence when one of the monitor readings is not available.

3.5 Industrial case study and evaluation

The diluted oil pre-heater process presented in Section 2.4 is used to investigate the incomplete evidence handling ability of the proposed method. The diagnostic settings are kept the same.

Consider a scenario that some historical data sampled under *SP* mode have the missing control performance monitor reading π_1 . The missing ratio, or the probability of missing monitor changes according to the output of π_1 . When π_1 is “*normal*”, the probability that π_1 is missing is 0.3; when π_1 is “*abnormal*”, the probability is 0.9. Thus, π_1 is more likely to be unavailable when the control performance is poor. Column 2 of Table 3.4 shows the true occurrences of complete evidences in the historical data set for mode *SP*, and column 3 shows actual appearance of the evidences. It can be observed that the distribution of evidences is distorted with missing monitor, and so does the likelihood of evidences. If only the samples with complete evidence are used to calculate the likelihood, the diagnostic results of *SP*

mode is shown in the middle panel of Figure 3.9, where the highest probability is not assigned to the true underlying mode.

Table 3.4: Realizations of evidences with different data set

Evidences	Numbers of occurrences		
	true data	sampled data	recovered data
(101)	18	2	12.2
(001)	2	1	7.8
(100)	2	0	0.667
(000)	1	1	2.333

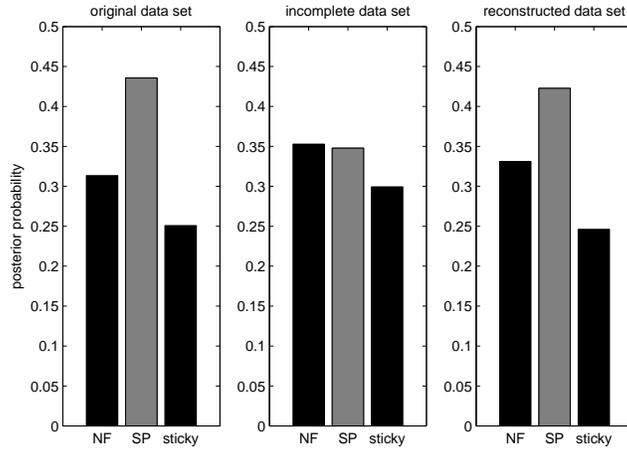


Figure 3.9: Diagnostic results with different data set

With the proposed approach, the recovered historical samples are presented in the last column of Figure 3.4. It can be seen that evidence distribution of the recovered data cannot be exactly recovered to the true one, but is able to be closer to the true evidence distribution. Thus in the diagnostic results obtained with the recovered data, shown in the right panel of Figure 3.9, the highest probability is indeed assigned to the true underlying mode, and the probabilities assigned to the three modes are close to the results obtained with the true data set.

3.6 Conclusions

In this chapter, a novel data-driven Bayesian approach for control loop diagnosis with incomplete evidences is presented. Monitor missing problems, including single

missing pattern ones and multiple missing pattern ones are handled via marginalization of evidences over the UCEM. The proposed method is verified by a simulated binary distillation column and an industrial process, where the features of the Bayesian approach to the incomplete evidences are demonstrated, and are compared to the performance of Gibbs sampling method.

Chapter 4

Bayesian Methods for Control Loop Diagnosis in Presence of Temporal Dependent Evidences

Conventional Bayesian methods commonly assume that the evidences are temporally independent. This condition, however, does not hold for most engineering problems. With the evidence transition information being considered, the temporal information can be synthesized within the Bayesian framework to improve diagnostic performance. In this chapter, the important evidence dependency problem is solved by a data-driven Bayesian approach with consideration of evidence transition probability. The sparse data problem induced by the high-dimensional evidence transition space is circumvented by analyzing the correlation ratio of the evidence. The applications in a simulated distillation column and a pilot scale process are presented to demonstrate the data dependency handling ability of the proposed diagnostic approach.

4.1 Introduction

In Chapter 2, the Bayesian diagnostic method proposed by Pernestål (2007) [73] is applied to the control loop diagnostic problem. The approach is further developed with consideration of incomplete evidences in Chapter 3. The algorithms have been tested through simulation as well as an industrial example, where the information synthesizing ability of the proposed approaches is demonstrated. However, the developed algorithms, along with majority of other existing data-driven Bayesian

A version of this chapter has been published in F. Qi, and B. Huang. Bayesian methods for control loop diagnosis in the presence of temporal dependent evidences. regular paper, *Automatica*, 2011, doi:10.1016/j.automatica.2011.02.015

methods, have not yet considered the temporal dependency problem.

Note that in the approaches described in the previous two chapters, an assumption is made in that the current evidence only depends on the underlying mode, and is independent of the previous monitor readings. Based on this assumption, the data-driven Bayesian diagnostic approach is developed. The corresponding graphic model is shown in Figure 4.1. The assumptions regarding evidence independency,

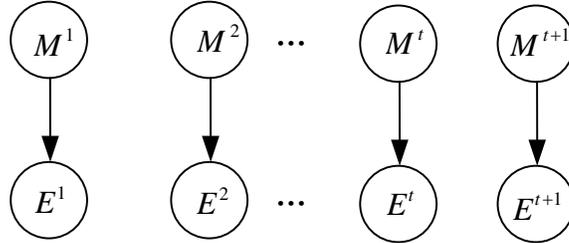


Figure 4.1: Bayesian model with independent evidence data samples

however, are restrictive in engineering applications; some important temporal information that is helpful for the the diagnosis is neglected. In this chapter, a new algorithm is developed with consideration of evidence temporal dependency, so as to solve the important evidence dependency problem with Bayesian methods.

The remainder of this chapter is organized as follows. The rationale to consider evidence dependency is detailed in Section 4.2. The estimation algorithm for the evidence transition probability is developed in Section 4.3. Section 4.4 proposes a dimension reduction solution for the high order evidence transition space. Sections 4.5 and 4.6 present applications of the proposed diagnostic approach to a simulated example and a pilot scale process. Section 4.7 concludes this chapter.

4.2 Temporally dependent evidences

An evidence is a statistic estimated from a section (window) of process data. The independency among evidences relies on how the evidence data are sampled, and how the disturbance affects the monitor outputs. If the evidence samples are collected with sufficiently large intervals, or if the disturbance has no or weak correlation among the evidence samples, the evidences can be considered as independent. Generally the first requirement regarding the sampling interval can be easily satisfied by leaving sufficient gap between consecutive monitor readings. If disturbance has long-term autocorrelation and the gap between consecutive monitor readings is not large

enough, then the temporal independency assumption of monitor readings cannot apply. A simple practical example of long-term auto-correlation of the disturbance is the ambient temperature change. Consider that each evidence, which includes one monitor reading, is calculated based on one-hour process data and there is no overlap in the use of data in consecutive monitor reading calculations. Due to the cyclic change of temperature within 24 hours, the monitor readings may follow a predictable pattern. Apparently it is more justifiable to consider the dependency between those evidence samples than simply ignoring it in this example if the monitor is indeed affected by the temperature.

Besides the practical issues, another limitation with the conventional Bayesian approach ignoring evidence dependency is its inability to capture all time domain information. An illustrative problem is presented in the following. Suppose that the system under diagnosis has two modes m_1 and m_2 . One monitor π , with two discrete outcomes, 0 and 1, is available. A set of 100 samples of the monitor outputs is shown in Figure 4.2. The title in each plot indicates the underlying operating mode under which the evidence data are collected. The likelihood probability of evidence being 0

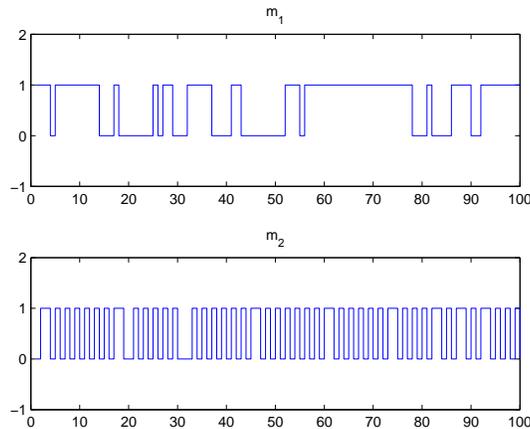


Figure 4.2: Monitor outputs of the illustrative problem

or 1 is summarized in Table 4.1. The likelihoods of the evidence being 0 or 1 under the two modes are almost identical. This may invoke confusion in the diagnosis, which will lead to high false diagnosis rate. By looking at the data plot in Figure 4.2, one can argue that distinguishing the two modes should not be such a difficult task. Although the evidences under m_1 and m_2 share similar likelihood, the frequencies of the evidence change apparently differ far from each other. The limitation with

Table 4.1: Likelihood estimation of the illustrative problem

	$e = 0$	$e = 1$
m_1	0.46	0.54
m_2	0.48	0.52

the conventional Bayesian method without considering evidence dependency is that the temporal information has not been completely explored, leading to less efficient diagnostic performance. In summary, it is desirable to take the evidence dependency into consideration when building the diagnostic model.

Assume that current evidence depends on both current underlying mode and previous evidence, and the evidence dependency follows a Markov process. The corresponding graphic model is shown in Figure 4.3. With the consideration of

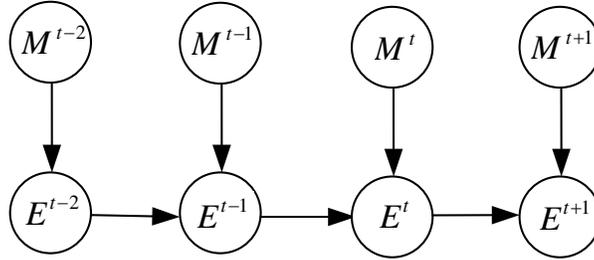


Figure 4.3: Bayesian model considering dependent evidence

evidence dependency, the mode posterior probability is calculated as

$$p(M^t|E^{t-1}, E^t, \mathcal{D}) \propto p(E^t|M, E^{t-1}, \mathcal{D})p(M^t). \quad (4.1)$$

In view of Equation 4.1, the main task of building a Bayesian diagnostic system boils down to the estimation of the evidence transition probability with historical evidence data \mathcal{D} , $p(E^t|M^t, E^{t-1}, \mathcal{D})$.

4.3 Estimation of evidence transition probability

The intention of the estimation of evidence transition probability is to let the estimated probabilities be consistent with the historical evidence data set \mathcal{D} in which the evidence dependency exists. Our goal is to calculate the likelihood probability of an evidence E^t given current underlying mode M^t and previous evidence E^{t-1} to

reflect the dependency with the Markov property, so every evidence transition sample, which is defined for evidence transition probability estimation purpose, should include these three elements,

$$d_E^{t-1} = \{M^t, E^{t-1}, E^t\}. \quad (4.2)$$

The new evidence transition data set \mathcal{D}_E , which is assembled from historical evidence data set \mathcal{D} to estimate the evidence transition probability, is defined as

$$\begin{aligned} \mathcal{D}_E &= \{d_E^1, \dots, d_E^{\hat{N}-1}\} \\ &= \{(M^2, E^1, E^2), \dots, (M^{\hat{N}}, E^{\hat{N}-1}, E^{\hat{N}})\}, \end{aligned} \quad (4.3)$$

which may also be called transition data set for simplicity. Figure 4.4 depicts how the collected historical evidence data are divided to form transition samples. In Figure 4.4, the part highlighted with shadows or gray and enclosed by the dash-lined or solid-lined frame is a transition sample described by Equation 4.2.

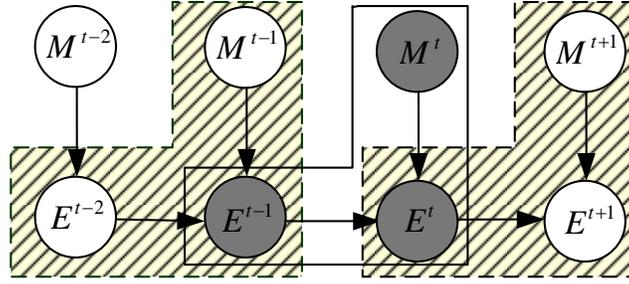


Figure 4.4: Illustration of evidence transition samples

Suppose that the evidence transition probability from $E^{t-1} = e_s$ to $E^t = e_r$ under mode $M^t = m_c$ is to be estimated from the transition data set,

$$p(E^t | E^{t-1}, M^t, \mathcal{D}_E) = p(e_r | e_s, m_c, \mathcal{D}_E) \quad (4.4)$$

where

$$e_s, e_r \in \mathcal{E} = \{e_1, \dots, e_K\}, \quad (4.5)$$

and

$$m_c \in \mathcal{M} = \{m_1, \dots, m_Q\}. \quad (4.6)$$

The evidence transition probability $p(e_r | e_s, m_c, \mathcal{D}_E)$ can only be estimated from the subset $\mathcal{D}_{E|m_c}$ where the mode $M^t = m_c$,

$$\begin{aligned} p(e_r | e_s, m_c, \mathcal{D}_E) &= p(e_r | e_s, m_c, \mathcal{D}_{E|m_c}, \mathcal{D}_{E|\neg m_c}) \\ &= p(e_r | e_s, m_c, \mathcal{D}_{E|m_c}), \end{aligned} \quad (4.7)$$

where $D_{E|m_c}$ is the historical transition data set when the underlying mode M^t is not m_c . To simplify notations, the subscript m_c will be omitted when it is clear from the context.

The evidence transition probability can be computed by marginalization over all possible evidence transition probability parameters,

$$\begin{aligned} & p(e_r|e_s, m_c, \mathcal{D}_E) \\ &= \int_{\Psi_1, \dots, \Psi_K} p(e_r|\Phi_1, \dots, \Phi_K, e_s, m_c, \mathcal{D}_E) \cdot f(\Phi_1, \dots, \Phi_K|e_s, m_c, \mathcal{D}_E) d\Phi_1 \dots \Phi_K. \end{aligned} \quad (4.8)$$

where $\Phi_i = \{\phi_{i,1}, \phi_{i,2}, \dots, \phi_{i,K}\}$ represents the probability parameter set for all possible evidence transitions from evidence e_i under mode m_c , and K is the total number of possible evidence values. For example, $\phi_{s,r} = p(e_r|e_s, m_c)$ is the probability of evidence transition from $E^{t-1} = e_s$ to $E^t = e_r$ with underlying mode $M^t = m_c$. According to the definition of $\phi_{s,r}$, we have

$$\sum_{i=1}^K \phi_{s,i} = \sum_{i=1}^K p(e_i|e_s, m_c) = 1. \quad (4.9)$$

Ψ_i is the space of all the probability parameter sets Φ_i subject to Equation 4.9.

Among all parameter sets Φ_1, \dots, Φ_K , the value of $p(e_r|\Phi_1, \dots, \Phi_K, e_s, m_c, \mathcal{D}_E)$ only depends on the parameter set Φ_s ; thus

$$p(e_r|\Phi_1, \dots, \Phi_K, e_s, m_c, \mathcal{D}_E) = p(e_r|\Phi_s, e_s, m_c, \mathcal{D}_E). \quad (4.10)$$

Substituting Equation 4.10 in Equation 4.8 yields

$$\begin{aligned} & p(e_r|e_s, m_c, \mathcal{D}_E) \\ &= \int_{\Psi_1, \dots, \Psi_K} p(e_r|\Phi_s, e_s, m_c, \mathcal{D}_E) \cdot f(\Phi_1, \dots, \Phi_K|e_s, m_c, \mathcal{D}_E) d\Phi_1 \dots \Phi_K \\ &= \int_{\Psi_1, \dots, \Psi_K} \phi_{s,r} \cdot f(\Phi_1, \dots, \Phi_K|e_s, m_c, \mathcal{D}_E) d\Phi_1 \dots \Phi_K. \end{aligned} \quad (4.11)$$

The calculation of the second term, $f(\Phi_1, \dots, \Phi_K|e_s, m_c, \mathcal{D}_E)$, is tackled from a Bayesian perspective. It can be regarded as the posterior probability of parameter sets $\{\Phi_1, \dots, \Phi_K\}$,

$$\begin{aligned} & f(\Phi_1, \dots, \Phi_K|e_s, m_c, \mathcal{D}_E) \\ &= \frac{p(\mathcal{D}_E|e_s, m_c, \Phi_1, \dots, \Phi_K) f(\Phi_1, \dots, \Phi_K|e_s, m_c)}{p(\mathcal{D}_E|e_s, m_c)}, \end{aligned} \quad (4.12)$$

where

$$\begin{aligned}
& p(\mathcal{D}_E|e_s, m_c) \\
&= \int_{\Psi_1, \dots, \Psi_K} p(\mathcal{D}_E|e_s, m_c, \Phi_1, \dots, \Phi_K) \cdot f(\Phi_1, \dots, \Phi_K|e_s, m_c) d\Phi_1 \cdots \Phi_K \quad (4.13)
\end{aligned}$$

is the scaling factor.

In Equation 4.12, the first term in the numerator, $p(\mathcal{D}_E|e_s, m_c, \Phi_1, \dots, \Phi_K)$, is the likelihood of transition data set given parameter sets $\{\Phi_1, \dots, \Phi_K\}$. It should be noted that likelihood of transition data \mathcal{D}_E is solely determined by the mode and parameter sets $\{\Phi_1, \dots, \Phi_K\}$, and thereby is independent of current evidence e_s , i.e.,

$$\begin{aligned}
p(\mathcal{D}_E|e_s, m_c, \Phi_1, \dots, \Phi_K) &= p(\mathcal{D}_E|m_c, \Phi_1, \dots, \Phi_K) \\
&= \prod_{i=1}^K \prod_{j=1}^K \phi_{i,j}^{\tilde{n}_{i,j}}, \quad (4.14)
\end{aligned}$$

where $\tilde{n}_{i,j}$ is the number of evidence transitions from e_i to e_j in the transition data set.

Similar to the likelihood of historical transition data set, the prior probability of transition parameter set Φ_i is solely determined by the underlying mode m_c , and thus is independent of current evidence e_s ,

$$f(\Phi_1, \dots, \Phi_K|e_s, m_c) = f(\Phi_1, \dots, \Phi_K|m_c).$$

With the common assumption that the priors for different parameter sets Φ_i and Φ_j , where $i \neq j$, are independent [73],

$$f(\Phi_1, \dots, \Phi_K|m_c) = f(\Phi_1|m_c) \cdots f(\Phi_K|m_c). \quad (4.15)$$

Dirichlet distribution is used to model priors of the likelihood parameters with Dirichlet parameters $b_{i,1}, \dots, b_{i,K}$,

$$f(\Phi_i|m_c) = \frac{\Gamma(\sum_{j=1}^K b_{i,j})}{\prod_{j=1}^K \Gamma(b_{i,j})} \prod_{j=1}^K \phi_{i,j}^{b_{i,j}-1}, \quad (4.16)$$

so

$$f(\Phi_1, \dots, \Phi_K|e_s, m_c) = \prod_{i=1}^K \frac{\Gamma(\sum_{j=1}^K b_{i,j})}{\prod_{j=1}^K \Gamma(b_{i,j})} \prod_{j=1}^K \phi_{i,j}^{b_{i,j}-1} \quad (4.17)$$

where $b_{i,j}$ can be interpreted as the number of prior samples for evidence transition from e_i to e_j . $\Gamma(\cdot)$ is the gamma function,

$$\Gamma(x) = \int_0^\infty t^{x-1} e^{-t} dt. \quad (4.18)$$

Since in this chapter all the independent variables x of gamma functions are counts of evidence transitions, which are positive integers, so

$$\Gamma(x) = (x-1)!. \quad (4.19)$$

Substituting Equation 4.17 and Equation 4.14 in Equation 4.12, we have

$$\begin{aligned} & f(\Phi_1, \dots, \Phi_K | e_s, m_c, \mathcal{D}_E) \\ &= \frac{p(\mathcal{D}_E | e_s, m_c, \Phi_1, \dots, \Phi_K) f(\Phi_1, \dots, \Phi_K | e_s, m_c)}{p(\mathcal{D}_E | e_s, m_c)} \\ &= \frac{1}{p(\mathcal{D}_E | e_s, m_c)} \cdot \prod_{i=1}^K \frac{\Gamma(\sum_{j=1}^K b_{i,j})}{\prod_{j=1}^K \Gamma(b_{i,j})} \prod_{j=1}^K \phi_{i,j}^{b_{i,j}-1} \cdot \prod_{i=1}^K \prod_{j=1}^K \phi_{i,j}^{\tilde{n}_{i,j}} \end{aligned} \quad (4.20)$$

Let

$$\mu = \prod_{i=1}^K \frac{\Gamma(\sum_{j=1}^K b_{i,j})}{\prod_{j=1}^K \Gamma(b_{i,j})}, \quad (4.21)$$

and then Equation 4.20 can be written as

$$\begin{aligned} & f(\Phi_1, \dots, \Phi_K | e_s, m_c, \mathcal{D}_E) \\ &= \frac{\mu}{p(\mathcal{D}_E | e_s, m_c)} \cdot \prod_{i=1}^K \prod_{j=1}^K \phi_{i,j}^{b_{i,j}-1} \cdot \prod_{i=1}^K \prod_{j=1}^K \phi_{i,j}^{\tilde{n}_{i,j}} \\ &= \frac{\mu}{p(D | e_s, m_c)} \prod_{i=1}^K \prod_{j=1}^K \phi_{i,j}^{\tilde{n}_{i,j} + b_{i,j} - 1}. \end{aligned} \quad (4.22)$$

Therefore the transition probability from evidence e_s to e_r can be derived as

$$\begin{aligned} & p(e_r | e_s, m_c, \mathcal{D}_E) \\ &= \int_{\Psi_1, \dots, \Psi_K} \phi_{s,r} \cdot f(\Phi_1, \dots, \Phi_K | e_s, m_c, \mathcal{D}_E) d\Phi_1 \dots \Phi_K \\ &= \int_{\Psi_1, \dots, \Psi_K} \phi_{s,r} \cdot \frac{\mu}{p(\mathcal{D}_E | e_s, m_c)} \cdot \prod_{i=1}^K \prod_{j=1}^K \phi_{i,j}^{\tilde{n}_{i,j} + b_{i,j} - 1} d\Phi_1 \dots \Phi_K \\ &= \frac{\mu}{p(\mathcal{D}_E | e_s, m_c)} \int_{\Psi_1} \prod_{j=1}^K \phi_{1,j}^{\tilde{n}_{1,j} + b_{1,j} - 1} d\Phi_1 \dots \int_{\Psi_s} \phi_{s,r}^{\tilde{n}_{s,r} + b_{s,r}} \prod_{j \neq r} \phi_{s,j}^{\tilde{n}_{s,j} + b_{s,j} - 1} d\Phi_s \\ & \quad \dots \int_{\Psi_K} \prod_{j=1}^K \phi_{K,j}^{\tilde{n}_{K,j} + b_{K,j} - 1} d\Phi_K \end{aligned} \quad (4.23)$$

In the above equation, $p(\mathcal{D}_E|e_s, m_c)$ is the scaling factor as defined in Equation 4.12. According to Equation 4.13,

$$\begin{aligned}
& p(\mathcal{D}_E|e_s, m_c) \\
&= \int_{\Psi_1, \dots, \Psi_K} p(\mathcal{D}_E|e_s, m_c, \Phi_1, \dots, \Phi_K) \cdot f(\Phi_1, \dots, \Phi_K|e_s, m_c) d\Phi_1 \dots \Phi_K \\
&= \int_{\Psi_1, \dots, \Psi_K} \prod_{i=1}^K \frac{\Gamma(\sum_{j=1}^K b_{i,j})}{\prod_{j=1}^K \Gamma(b_{i,j})} \cdot \prod_{i=1}^K \prod_{j=1}^K \phi_{i,j}^{\tilde{n}_{i,j} + b_{i,j} - 1} d\Phi_1 \dots \Phi_K \\
&= \mu \cdot \int_{\Psi_1} \prod_{j=1}^K \phi_{1,j}^{\tilde{n}_{1,j} + b_{1,j} - 1} d\Phi_1 \dots \int_{\Psi_K} \prod_{j=1}^K \phi_{K,j}^{\tilde{n}_{K,j} + b_{K,j} - 1} d\Phi_K \\
&= \mu \cdot \prod_{i=1}^K \frac{\prod_{j=1}^K \Gamma(\tilde{n}_{i,j} + b_{i,j})}{\Gamma(\tilde{N}_i + B_i)}, \tag{4.24}
\end{aligned}$$

where $\tilde{N}_i = \sum_j \tilde{n}_{i,j}$ is the total number of transition samples, from previous evidence e_i under mode m_c , and $B_i = \sum_j b_{i,j}$ is the corresponding total number of transition samples under mode m_c .

Similarly, we can derive

$$\begin{aligned}
& \int_{\Psi_1} \prod_{j=1}^K \phi_{1,j}^{\tilde{n}_{1,j} + b_{1,j} - 1} d\Phi_1 \dots \int_{\Psi_s} \phi_{s,r}^{\tilde{n}_{s,t} + b_{s,r}} \cdot \prod_{j \neq r} \phi_{s,j}^{\tilde{n}_{s,j} + b_{s,j} - 1} d\Phi_s \dots \int_{\Psi_K} \prod_{j=1}^K \phi_{K,j}^{\tilde{n}_{K,j} + b_{K,j} - 1} d\Phi_K \\
&= \frac{\Gamma(\tilde{n}_{s,r} + b_{s,r} + 1)}{\Gamma(\tilde{N}_s + B_s + 1)} \cdot \frac{\prod_{i,j \neq s,r} \Gamma(\tilde{n}_{i,j} + b_{i,j})}{\prod_{i \neq s} \Gamma(\tilde{N}_i + B_i)} \tag{4.25}
\end{aligned}$$

Thus Equation 4.23 can be simplified as

$$\begin{aligned}
& p(e_r|e_s, m_c, \mathcal{D}_E) \\
&= \mu \cdot \frac{\Gamma(\tilde{n}_{s,r} + b_{s,r} + 1)}{\Gamma(\tilde{N}_s + B_s + 1)} \cdot \frac{\prod_{i,j \neq s,r} \Gamma(\tilde{n}_{i,j} + b_{i,j})}{\prod_{i \neq s} \Gamma(\tilde{N}_i + B_i)} \cdot \frac{\prod_{i=1}^K \Gamma(\tilde{N}_i + B_i)}{\mu \cdot \prod_{i=1}^K \prod_{j=1}^K \Gamma(\tilde{n}_{i,j} + b_{i,j})} \\
&= \frac{\tilde{n}_{s,r} + b_{s,r}}{\tilde{N}_s + B_s} \tag{4.26}
\end{aligned}$$

Rewrite the evidence likelihood equation without consideration of evidence dependency,

$$p(e_i|m_c, \mathcal{D}) = \frac{n_i + a_i}{N + A}. \tag{4.27}$$

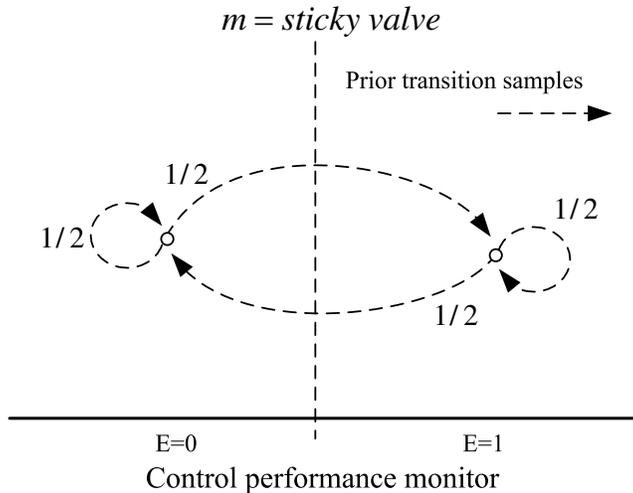
By comparing Equation 4.27 and Equation 4.26, we can see that the evidence transition probability is also determined by both prior samples and historical samples, similar to the single evidence likelihood calculation when the evidences are considered as independent. The difference lies in how the numbers of prior and historical samples are counted. In Equation 4.27 the prior and historical samples refer

to a count of the single evidence samples corresponding to a target mode, while in Equation 4.26 the prior and historical samples refer to the count of the transition samples under the target mode. However, the evidence transition space is much larger than the single evidence space. This may lead to insufficient historical data problem, which will be elaborated shortly.

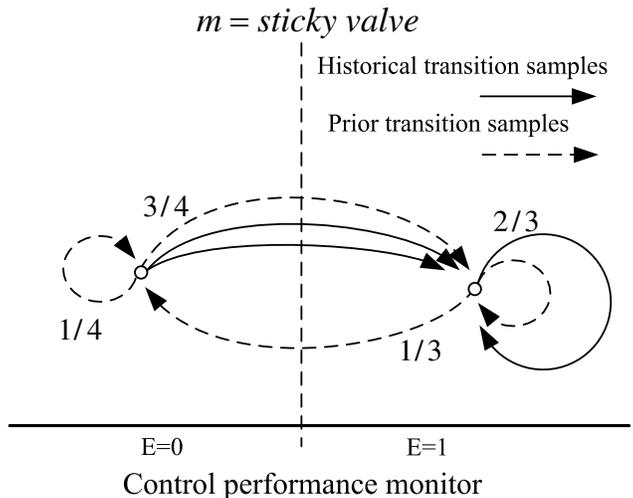
As the number of historical transition samples increases, the transition probability will converge to the relative frequency determined by the historical samples, and the influence of the priors will decrease. The number of prior samples can be interpreted as prior belief of the evidence transition probability distribution, where the uniform distribution requires that prior sample numbers are equal across all possible transitions under any given underlying mode. It is important to set nonzero prior sample numbers; otherwise unexpected results may occur in the case of limited historical samples [73].

As an example, consider a univariate control loop under diagnosis with one problematic component: a valve subject to the stiction problem. A control performance monitor is commissioned on the loop. The reading of the monitor is discretized into two bins according to a predefined threshold. One is corresponding to $E = 0$ and the other is for $E = 1$. This is illustrated in Figure 5(a), where the bin for $E = 0$ is to the left of the dashed vertical line and to the right is the bin for $E = 1$. Therefore there are four possible evidence transitions, namely $E^{t-1} = 0 \rightarrow E^t = 0$, $E^{t-1} = 0 \rightarrow E^t = 1$, $E^{t-1} = 1 \rightarrow E^t = 0$ and $E^{t-1} = 1 \rightarrow E^t = 1$, which are depicted as the directed arcs in Figure 4.5(a). Let the underlying system mode be $m = \text{sticky valve}$. Each evidence transition is assigned with one prior sample under the assumption of uniformly distributed prior transition samples. See Figure 4.5(a). Hence, $b_{i,j} = 1$, $B_i = 2$, and all the evidence transition probabilities equal $1/2$. With the historical data collected under the same underlying mode m , the evidence transition probabilities can be updated. In Figure 4.5(b), the solid directed arcs denote the historical transition samples. For instance, there are two solid arcs from the bin $E = 0$ to $E = 1$, indicating that two historical samples have been collected, $\tilde{n}_{0,1} = 2$. Meanwhile no solid arc is seen from $E = 0$ to itself, and thus there is no historical sample corresponding to that transition, $\tilde{n}_{0,0} = 0$, and $\tilde{N}_0 = \sum_i \tilde{n}_{0,i} = 2$. According to Equation 4.26, the transition probability from $E^{t-1} = 0$ to $E^t = 1$ is calculated,

$$p(E^t = 1 | E^{t-1} = 0, m, \mathcal{D}_E) = \frac{\tilde{n}_{0,1} + b_{0,1}}{\tilde{N}_0 + B_0} = \frac{3}{4}, \quad (4.28)$$



(a) Evidence transition probability with only prior transition samples



(b) Updated evidence transition probability with historical transition samples

Figure 4.5: Evidence transition probability updating

as seen in Figure 4.5(b).

4.4 Reduction of evidence transition space

Define the evidence transition space as the set that contains all possible evidence transitions. The dimension of the evidence transition space is the total number of monitor readings that an evidence transition sample can include. A practical problem encountered in the evidence transition probability estimation is the large

combinatorial number of evidence transitions. Consider a diagnostic system with ten monitors,

$$E = \{\pi_1, \pi_2, \dots, \pi_{10}\},$$

and each single monitor output is discretized into two bins. The total number of possible evidence values equals

$$K = 2^{10} = 1024; \tag{4.29}$$

the total number of possible evidence transitions equals

$$K' = 2^{10} \cdot 2^{10} = 1048576, \tag{4.30}$$

which is the square of the total number of single evidence values. Although the number of evidence transition combinations that have actually occurred in the data set, may be lower, the number of historical transition samples still needs to increase dramatically in general to generate an accurate estimation of the evidence transition probability. Otherwise insufficient data will lead to degraded diagnostic performance. This issue will be illustrated shortly.

Obtaining such a large number of historical samples is challenging, especially for an industrial process, where the available historical data corresponding to faulty modes may be sparse. To circumvent this problem, we propose an approximate solution based on correlation ratio analysis of the evidence, to reduce the dimension of the evidence transition space, and thereby to relieve the requirement of large amount of historical evidence data.

The temporal dependency (auto-dependency) of the evidences is mainly introduced by the temporally dependent disturbances. Consider, for example, that the disturbance is a sequence of filtered white noise. If the time gap between two consecutive monitor readings is much larger than the time constant of the filter, those monitor readings may be treated as auto-independent; otherwise the auto-dependency introduced by the disturbance has to be taken into consideration. By analyzing dependence of the evidence data, we can divide the monitor readings into two groups: the auto-dependent ones, and the auto-independent ones.

Correlation ratio, which measures the functional dependence of two random variables, is selected to model the dependence of the monitors. In contrast to the correlation coefficient, the correlation ratio is capable of detecting almost any functional dependency, not only the linear dependency [19].

For illustration, consider a diagnostic system with one monitor, $E = \{\pi\}$. Suppose that totally N evidence samples are collected, and are represented as

$$\{(\pi^1), (\pi^2), \dots, (\pi^N)\},$$

where π^t is the t th collected sample of π . Assume that the monitor has two discrete-valued outcomes, 0 and 1. To estimate the correlation ratio of π^t on π^{t-1} , shift the data sequence of π by one sample and then create a set of paired data as follows:

$$\{(\pi^1, \pi^2), (\pi^2, \pi^3), \dots, (\pi^{N-1}, \pi^N)\}.$$

The collected samples of π^t are classified into two categories according to first value of each pairs:

$$\begin{aligned} \text{The 1st value of the pair} = 0 : & \quad \pi^{0,1} \quad \pi^{0,2} \quad \dots \quad \pi^{0,n_0} \\ \text{The 1st value of the pair} = 1 : & \quad \pi^{1,1} \quad \pi^{1,2} \quad \dots \quad \pi^{1,n_1}, \end{aligned}$$

where $\pi^{i,j}$ the second element of the j th pair whose first value is i . Following [63], the correlation ratio of π^t on the π^{t-1} can be calculated as

$$\eta(\pi^t | \pi^{t-1}) = \sqrt{\frac{\sum_{i=0}^1 n_i (\bar{\pi}^i - \bar{\pi})^2}{\sum_{i,j}^{1,n_i} (\pi^{i,j} - \bar{\pi})^2}}, \quad (4.31)$$

where $i = 0, 1$, $\bar{\pi}^i$ is the average of the monitor readings in the i th category, and $\bar{\pi}$ is the average of all the collected monitor readings. The calculated correlation ratio quantifies the dependency of π^t on π^{t-1} , namely, the effect of the value of π^{t-1} on the value of π^t . This value is an indication of how previous π value affects current π value. $\eta(\pi^t | \pi^{t-1}) = 0$ means that π is auto-independent, while $\eta(\pi^t | \pi^{t-1}) = 1$ indicates that current output of π is completely determined by its previous value. Generally if $\eta(\pi^t | \pi^{t-1})$ is sufficiently small, π can be treated as auto-independent. Similarly, we can calculate the cross correlation ratio between different monitor outputs, say, π_1 and π_2 , by creating a new set of paired data,

$$\{(\pi_1^1, \pi_2^1), (\pi_1^2, \pi_2^2), \dots, (\pi_1^N, \pi_2^N)\}.$$

By calculating the auto-correlation ratio of each monitor, the temporal dependency of each monitor may be determined. Without loss of generality, assume that the first q monitors generate auto-independent outputs, while remaining monitors do not. An evidence can now be divided into two portions,

$$E = \{\pi_1, \pi_2, \dots, \pi_q : \pi_{q+1}, \dots, \pi_L\} = \{E_c, E_{ic}\} \quad (4.32)$$

where $E_c = \{\pi_1, \pi_2, \dots, \pi_q\}$ is the set of auto-dependent monitors, and E_{ic} is the set of auto-independent monitors. Using the newly defined notations, the evidence transition probability can be written as

$$p(E^t|E^{t-1}, M^t) = p(E_c^t, E_{ic}^t|E_c^{t-1}, E_{ic}^{t-1}, M^t). \quad (4.33)$$

Further assume that E_c and E_{ic} are mutually independent (may be tested approximately by cross correlation ratio),

$$p(E_c^t, E_{ic}^t|E_c^{t-1}, E_{ic}^{t-1}, M^t) = p(E_c^t|E_c^{t-1}, E_{ic}^{t-1}, M^t)p(E_{ic}^t|E_c^{t-1}, E_{ic}^{t-1}, M^t). \quad (4.34)$$

Recall that E_{ic}^t only depends on the underlying mode M^t , and E_c^t depends on both the underlying mode M^t and the previous evidence, E_c^{t-1} , so Equation 4.34 becomes

$$\begin{aligned} & p(E_c^t, E_{ic}^t|E_c^{t-1}, E_{ic}^{t-1}, M^t) \\ &= p(E_c^t|E_c^{t-1}, E_{ic}^{t-1}, M^t)p(E_{ic}^t|E_c^{t-1}, E_{ic}^{t-1}, M^t) \\ &= p(E_c^t|E_c^{t-1}, M^t)p(E_{ic}^t|M^t), \end{aligned} \quad (4.35)$$

which means that the probability of the evidence transition equals the product of transition probability of temporally dependent part of the evidence, i.e., partial evidence transition, and the likelihood of temporally independent monitor outputs, i.e., partial single evidence. Hence the likelihoods of the partial evidence transition and partial single evidence can be estimated separately. Only a subspace of the evidence transition space is needed to estimate the partial evidence transition probability, and the dimension of the evidence transition space is therefore reduced. Let monitor π_i have k_i discrete outputs. The number of complete evidence transition is

$$K' = \prod_{i=1}^L k_i \cdot \prod_{i=1}^L k_i = \prod_{i=1}^L k_i^2. \quad (4.36)$$

In Equation 4.35, the number of partial evidence transitions is $\prod_{i=1}^q k_i^2 < K'$. Consequently fewer historical data samples are required to calculate the evidence transition probability. Better diagnostic results are expected with the same amount of historical data samples.

4.5 Simulation example

The simulated binary distillation column [102] that is used for simulations in the previous chapters, is selected to evaluate the Bayesian diagnostic approaches with and without considering the evidence dependency.

4.5.1 Diagnostic settings

In addition to the simulation settings presented in Chapter 2, random binary sequence of limited frequency bandwidth is introduced into the FBP top and PCT bottom measurement as two-level temporal dependent disturbances for the sake of simulation. The change of disturbance in FBP top follows the transition probability matrices shown in Equation 4.38, where the transition probabilities of FBP disturbance under mode m_3 , $P_{m_3}^{FBP}$, are different from those under the other modes $P_{-m_3}^{FBP}$; the change of disturbance in PCT bottom follows the transition matrices shown in Equation 4.37, where the transition probabilities of PCT disturbance under mode m_7 , $P_{m_7}^{PCT}$ are different from those under the other modes $P_{-m_7}^{PCT}$.

$$P_{m_3}^{FBP} = \begin{matrix} & \begin{matrix} 0 & 1 \end{matrix} \\ \begin{matrix} 0 \\ 1 \end{matrix} & \begin{pmatrix} 0.9 & 0.1 \\ 0.1 & 0.9 \end{pmatrix} \end{matrix}, P_{-m_3}^{FBP} = \begin{matrix} & \begin{matrix} 0 & 1 \end{matrix} \\ \begin{matrix} 0 \\ 1 \end{matrix} & \begin{pmatrix} 0.1 & 0.9 \\ 0.9 & 0.1 \end{pmatrix} \end{matrix}; \quad (4.37)$$

$$P_{m_7}^{PCT} = \begin{matrix} & \begin{matrix} 0 & 1 \end{matrix} \\ \begin{matrix} 0 \\ 1 \end{matrix} & \begin{pmatrix} 0.9 & 0.1 \\ 0.1 & 0.9 \end{pmatrix} \end{matrix}, P_{-m_7}^{PCT} = \begin{matrix} & \begin{matrix} 0 & 1 \end{matrix} \\ \begin{matrix} 0 \\ 1 \end{matrix} & \begin{pmatrix} 0.1 & 0.9 \\ 0.9 & 0.1 \end{pmatrix} \end{matrix}. \quad (4.38)$$

5000 evidence samples are simulated. It should be noted that while the evidence transitions are subject to the rules stated above, the underlying modes are completely randomly generated and independent of each other. Also note that in contrast to the data collection presented in previous chapters, where the historical data are “selectively” collected, namely, neglecting the time domain information, the evidence samples collected here are all in temporal order, so that important time domain information such as evidence transition probability can be persevered and thereafter be retrieved.

Among the collected evidence samples, 3000 of them are used to estimate the evidence likelihood, evidence transition probability, and the other 2000 independently generated evidence samples are used for cross-validation of the diagnostic system. In the validation data set, some data are from the unconsidered (UC) mode, for which no historical data are available. Thus there are totally 11 possible modes to diagnose.

The parameter settings of the Bayesian diagnostic system are summarized in Table 4.2.

Table 4.2: Summary of Bayesian diagnostic parameters

Discretization	$k_i = 3(\text{“low”}, \text{“medium”}, \text{“high”}),$ $K = 3^{15} = 14348907$
Historical evidence data	A mixture of totally 3000 samples from the 10 modes
Prior samples	Uniformly distributed with prior sample, for single evidence space, and evidence transition space
Prior probabilities	Uniformly distributed, $p(m) = 1/11$
Evaluation data	Mixture of 2000 samples from the 10 known modes, and UC mode

4.5.2 Diagnostic results

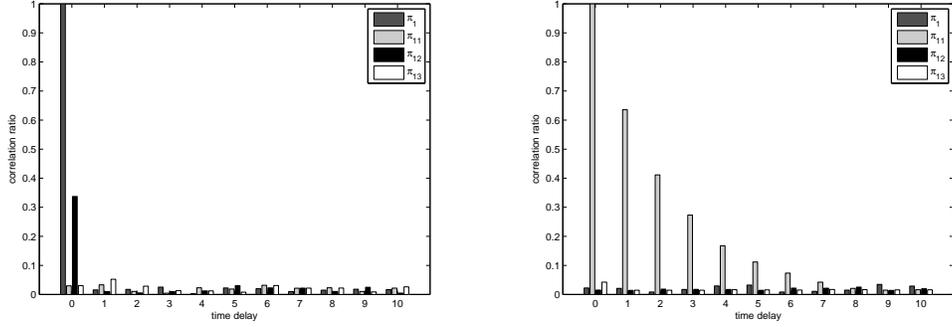
The single evidence space of this system is already a high-dimensional one, where a single evidence includes 15 monitors. The number of possible evidence transitions is the square of that of the single evidences, i.e.,

$$K' = K^2 = 3^{30} = 2.0589 \times 10^{14},$$

which indicates that an evidence transition can take any combination among the 2.0589×10^{14} possible ones. For such a high-dimensional space, 3000 samples are distributed sparsely, and are far from enough to give a meaningful estimation of the evidence transition probability.

To deal with this sparse evidence transition data problem, the correlation ratio of the evidence vector is investigated to reveal the temporal dependency of the monitor outputs. The auto-correlation ratio and cross-correlation ratio of each discrete monitor output are calculated and compared. The correlation ratios of monitor π_1 and π_{11} against some of other monitors are displayed in Figure 4.6 as examples. According to the correlation ratio analysis, it is observed that the monitor π_{11} and π_{13} both have strong temporal auto-dependency, and the other monitors are all auto-independent as well as cross-independent of monitor π_{11} and π_{13} .

This results in reduction of the evidence transition space. According to the



(a) Correlation ratio of π_1 against other monitors (b) Correlation ratio of π_{11} against other monitors

Figure 4.6: Correlation ratios of monitors

discussions in Section 4.4 and the correlation ratio analysis above,

$$\begin{aligned}
& p(E^t | E^{t-1}, M^t) \\
&= p(\{\pi_1^t, \dots, \pi_{11}^t, \pi_{12}^t, \pi_{13}^t, \dots\} | \{\pi_1^{t-1}, \dots, \pi_{11}^{t-1}, \pi_{12}^{t-1}, \pi_{13}^{t-1}, \dots\}, M^t) \\
&= p(\{\pi_1^t, \dots, \pi_{12}^t, \dots\} | M^t) p(\{\pi_{11}^t, \pi_{13}^t\} | \{\pi_{11}^{t-1}, \pi_{13}^{t-1}\}, M^t) \\
&= p(E_{ic}^t | M^t) p(E_c^t | E_c^{t-1}, M^t), \tag{4.39}
\end{aligned}$$

The original high-dimensional evidence transition space can be separated into a low dimensional partial single evidence space and a low dimensional partial evidence transition space. Recall the simulation settings. The disturbances, which affect monitor π_{11} and π_{13} , are both assumed to be a binary signal with predefined high and low levels, which indicates that there exist four possible values of E_c^t . Thus the number of possible partial evidence transitions is $K' = 4^2 = 16$, and the number of possible partial single evidences is

$$K = 3^{13} = 1.59 \times 10^6. \tag{4.40}$$

These numbers are remarkably reduced in comparison with the the original evidence transition space, which has 2.0589×10^{14} different evidence transitions. Although the number of lower-dimensional partial single evidences is still a large one, the real effective dimension is actually even lower, since some combinations of partial single evidence never appear in the data set. In this simulation, for example, including both historical and validation data, only 595 different single partial evidence values have actually occurred. Thus 3000 samples are sufficient to generate a reasonable estimation of the partial single evidence likelihood distribution.

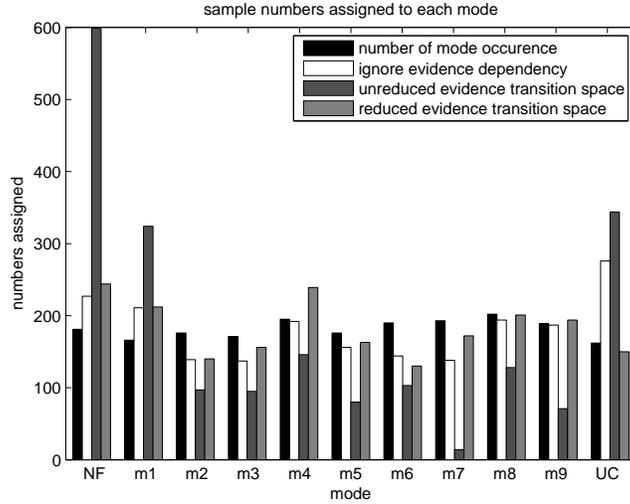


Figure 4.7: Numbers of occurrences diagnosed for each mode

Diagnostic results shown in Figure 4.7 are obtained from the 2000 evaluation (cross-validation) evidence samples which are generated independently of the historical samples. In Figure 4.7, the dark bars are the numbers of occurrences of each mode in the validation data set. The other bars are the diagnostic results shown in the form of numbers of samples assigned to each mode by three different diagnostic strategies, namely: 1) Bayesian approach ignoring evidence dependency; 2) considering evidence dependency but without reducing the evidence transition space; and 3) both considering evidence dependency and reducing the evidence transition space. The closer the bars to the dark ones, the better the diagnostic performance is.

Figure 4.7 clearly illustrates the poor performance of the approach that considers the evidence dependency, but does not reduce the evidence transition space. The number assigned by this strategy, is unreasonably high for mode *NF*. This is due to the insufficiency of historical data. The number of possible values of the unreduced evidence transitions, as discussed above, is 2.0589×10^{14} . Although the actual dimension of the space is smaller for each mode, where not all possible evidence transitions appear in the historical and the validation data set, the number of historical data samples in this case study is still far from enough to yield a reasonable estimation of the evidence transition probabilities. According to further simulations, approximately half million of historical data samples are needed to es-

timate the transition probabilities to obtain a good diagnostic performance without using evidence transition space reduction.

The proposed approach, which considers the evidence dependency and also takes the strategy to reduce the evidence transition space dimension, yields better diagnostic results than other two methods. The numbers of each mode diagnosed by the proposed approach are close to the real distribution of the modes in the validation data set. The better performance of the proposed method is also demonstrated by the overall correct diagnosis rate, as presented in Table 4.3.

Table 4.3: Overall correct diagnosis rate

Ignoring evidence dependency	72.86%
Considering evidence dependency with original evidence transitions space	18.69%
Considering evidence dependency with reduced evidence transitions space	80.56%

Further, consider the diagnostic performance of mode m_3 and m_7 , for which the disturbances are different from the other modes according to Equation 4.37 and 4.38. Figure 4.8 displays the average posteriors assigned to the ten possible modes by the three approaches when the true underlying mode is m_3 or m_7 . In each figure, the title of each plot denotes the diagnostic approach used. In each plot, the posterior probability corresponding to the true underlying mode is denoted with gray bars, while the others are in dark bars. The diagnostic result is determined by picking up the mode with the largest posterior probability. Clearly for both modes, the approach considering evidence dependency without compressing evidence transition space has poor performance owing to insufficient historical data. It is interesting to compare the results obtained between the Bayesian method ignoring evidence dependency and the proposed approach. Both approaches, assign highest probability to the true underlying modes. Yet the probability for the true underlying mode calculated by the proposed approach is higher, indicating better diagnostic performance. This is validated by comparing the correct diagnosis rate of the two modes, as shown in Table 4.4. The proposed approach yields the highest correct diagnosis rate for both m_3 and m_7 among the three diagnostic strategies, indicating more effective handling of evidence dependency.

Table 4.4: Correct diagnosis rates for m_3 and m_7

	m_3	m_7
Ignoring evidence dependency	80.12%	69.27%
Considering evidence dependency with original evidence transitions space	9.36%	6.22%
Considering evidence dependency with reduced evidence transitions space	91.23%	89.12%

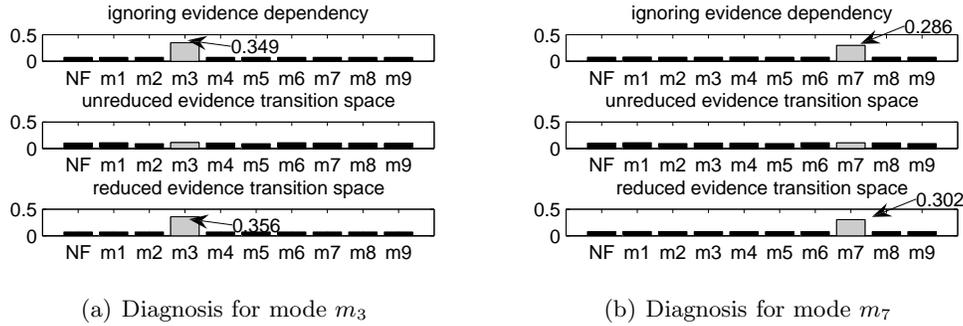


Figure 4.8: Average posteriors for validation data from mode m_3 and m_7

4.6 Pilot scale experiment

A pilot scale experiment is conducted to investigate performance of the proposed Bayesian approach with consideration of the evidence dependency, and to compare it with the Bayesian approach when the evidence dependency is ignored.

4.6.1 Process description

The experiment setup is a water tank with one inlet flow and two outlet flows. The schematic diagram of the process is shown in Figure 4.9. The inlet flow is driven by a pump. Of the two outlet flow valves, one is adjusted by a PID controller to provide level control for the tank, and the other one is a manual bypass valve. The bypass valve is closed when the system is in its normal operation condition.

Three operating modes are defined, including the *normal functioning* (NF) mode, and two problematic modes *leakage* and *bias*: the tank leakage problem defined as *leakage* mode, implemented by opening the bypass valve manually, and the sensor bias problem defined as *bias* mode, implemented by adding a constant

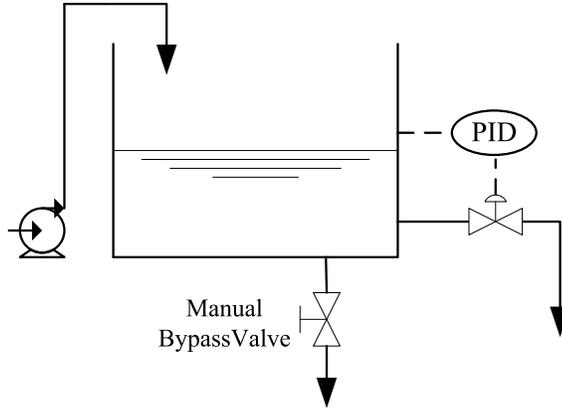


Figure 4.9: Pilot scale tank process

bias to the sensor output. The two problems share similar symptoms in terms of shifting the steady state operation point of the process. For instance, when there is a leakage in the tank, the valve adjusted by the PID controller will close to maintain the water level; when there is a negative sensor bias, the valve will also close. Thus it is not obvious how to distinguish the two faulty modes without any advanced diagnostic method. To make things worse, the external disturbance introduced by changing the pump input will shift the operation point. Thus the operation point may also change during the normal operation.

Consider that disturbance is introduced through flow changes fluctuating between two rates. A random binary sequence of a limited frequency band is introduced into the inlet pump input to simulate temporal dependent disturbances. By defining the high value as 1, and the low value as 0, the disturbance is designed to follow the transition probability matrices presented in Equation 4.41,

$$\begin{aligned}
 P_{NF}^{dis} &= \begin{matrix} & 0 & 1 \\ 0 & \begin{pmatrix} 0.9 & 0.1 \\ 0.2 & 0.8 \end{pmatrix} \\ 1 & \end{matrix}, P_{leakage}^{dis} = \begin{matrix} & 0 & 1 \\ 0 & \begin{pmatrix} 0.1 & 0.9 \\ 0.8 & 0.2 \end{pmatrix} \\ 1 & \end{matrix}, \\
 P_{bias}^{dis} &= \begin{matrix} & 0 & 1 \\ 0 & \begin{pmatrix} 0.5 & 0.5 \\ 0.5 & 0.5 \end{pmatrix} \\ 1 & \end{matrix}, \tag{4.41}
 \end{aligned}$$

where P_M^{dis} represents the transition probability matrix of the introduced disturbance under mode M .

Two process monitors, process model validation monitor and sensor bias monitor, are implemented. Since we are only concerned with the study of the information retrieving and synthesizing abilities of Bayesian approaches with different diagnostic strategies, some of the selected monitor algorithms do not necessarily have high performance.

The output of the process model validation monitor π_1 is given by the squared sum of the nominal model output residuals, scaled by the magnitude of the process output. Let the simulated output of the nominal model be \hat{y}_t at each sampling instance t , and the real output be y_t . The output of the model validation monitor π_1 is calculated as

$$\pi_1 = \frac{\sum_{t=1}^N (y_t - \hat{y}_t)^2}{\bar{y}}, \quad (4.42)$$

where $\bar{y} = \frac{1}{N} \sum_{t=1}^N y_i$ is the mean value of the process output over one monitor reading period, and N is the length of data segment during this period.

The sensor bias monitor output π_2 is obtained by examining shift of the operation point. For illustration, consider the scenario when a negative sensor bias occurs. The steady state in terms of the sensor output will not change, since it is controlled by the PID. The steady state output of the controller, i.e., the valve position, however, will decrease. The valve position will reverse in the presence of the positive sensor bias. Thus we can detect the sensor bias by monitoring the deviation of the mean value of the controller output from the nominal operation point. The output of the sensor bias monitor π_2 is calculated as

$$\pi_2 = \left| u_0 - \frac{1}{N} \sum_{t=1}^N u_t \right|, \quad (4.43)$$

where u_0 is the nominal operation point of the controller output, u_t is the controller output at sampling instant t , and N is the length of process data segment for a calculation of single monitor reading. Note that this monitor only applies to steady state data.

4.6.2 Diagnostic settings and results

Process data are collected for the three predefined modes. The sampling interval is set to be 1 second. Process data sampled over every 100 seconds are used for calculation of one monitor reading. Totally 600 monitor readings are calculated from 16.5 hours of process experiments. The collected evidence data of the three

modes are divided into two parts for estimation of the likelihood, and for cross-validation respectively. Table 4.5 summarizes the Bayesian diagnostic parameter settings.

Table 4.5: Summary of Bayesian diagnostic parameters

Discretization	$k_i = 2, K = 2^2 = 4$
Historical data	Totally 360 evidence samples for the three modes
Prior samples	Uniformly distributed with prior sample, for single evidence space, and evidence transition space
Prior probabilities	$p(NF) = p(m_{other}) = 1/3$
Evaluation data	240 independently generated cross-validation evidence samples

Note that there are only $4^2 = 16$ combinations of evidence transitions, so the reduction of the evidence transition space is not necessary. With the data-driven Bayesian approaches of two different strategies, namely, considering and ignoring the evidence dependency, the diagnostic results in Figure 4.10 are obtained based on the cross-validation data. In the plot, the white bars are the numbers of the underlying modes occurred in the validation data set; the light gray and black bars are the numbers of the diagnosed mode by two diagnostic approaches respectively.

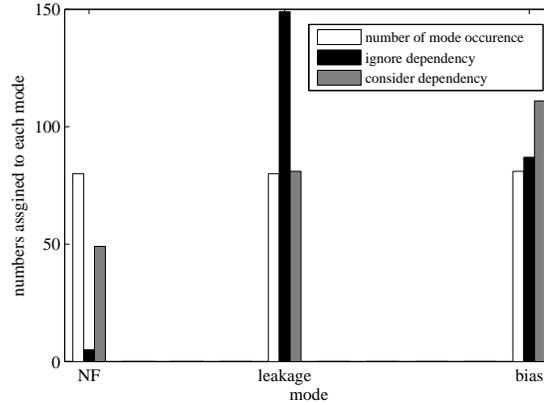


Figure 4.10: Numbers assigned to each mode

Owing to the dependent external disturbance, the Bayesian approach ignoring evidence dependency significantly overestimates the number of leakage mode, and underestimates the number of NF mode. Therefore, its overall correct diagnosis rate is only 51.45%, and is much lower in comparison to the diagnosis rate of the proposed

method, which is 73.86%. Not only can better overall performance be obtained by the proposed approach, the diagnostic performance of each single mode, as will be investigated, is also more favorable.

Figure 4.11 summarizes diagnostic results in the form of average posterior probabilities. The title of each plot denotes the true underlying mode, and the posterior probability corresponding to the true underlying mode is denoted by light gray. The left panel summarizes the diagnostic results calculated by the approach ignoring evidence dependency; the right panel summarizes the diagnostic results obtained by the approach with consideration of evidence dependency. It is observed that for the three modes, the posterior probabilities assigned to the true underlying modes by the proposed approach are all higher than that assigned by the method ignoring dependency. Thus we can conclude that the proposed approach has better performance for diagnosis of all modes. This conclusion is confirmed by computing the correct diagnosis rate for each single mode, as presented in Table 4.6.

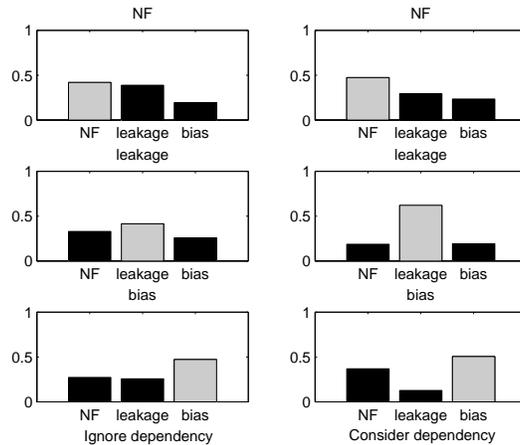


Figure 4.11: Average posterior probability for each mode

Table 4.6: Correct diagnosis rate for each single mode

	<i>NF</i>	<i>leakage</i>	<i>bias</i>
Ignore evidence dependency	6.25%	73.75%	70%
Consider evidence dependency	55%	78.75%	92.5%

4.7 Conclusions

In this chapter, a data-driven Bayesian approach with consideration of temporally dependent evidences is proposed for control loop diagnosis. Temporal dependency problem is solved by introducing transition evidence. The evidence transition probabilities needed for the solution are estimated from the historical evidence data. The large evidence transition space problem is alleviated by analyzing the correlation ratio of monitors. The proposed method is applied to a simulated binary distillation column and a pilot scale experiment setup, where the performance of the proposed approach is shown to be significantly improved compared to the method ignoring evidence temporal dependency.

Chapter 5

Dynamic Bayesian Approach for Control Loop Diagnosis with Underlying Mode Dependency

In this chapter, first of all, a hidden Markov model is built to address the temporal mode dependency problem in control loop diagnosis. A data-driven algorithm is developed to estimate the mode transition probability. The new solution to mode dependency is then further synthesized with the solution to evidence dependency to develop a recursive auto-regressive hidden Markov model for the online control loop diagnosis. When both the mode and evidence transition information is considered, the temporal information is effectively synthesized under the Bayesian framework. A simulated distillation column example and a pilot scale experiment example are investigated to demonstrate the ability of the proposed diagnostic approach.

5.1 Introduction

In Chapter 4, the temporal information in terms of evidence transition is synthesized within the Bayesian framework to improve diagnostic performance. However, the mode dependency has not been considered. In engineering practice, the assumption that the current system mode is independent of previous modes is not general enough. For example, the fact that new equipment has more of a tendency to operate normally in the future than aged equipment has not been considered. On the other hand, if a system is in a faulty mode, without any repair action being taken, it is more likely that the system will remain in the same faulty status. Furthermore, the

A version of this chapter has been published in F. Qi, and B. Huang. Dynamic Bayesian approach for control loop diagnosis with underlying mode dependency. *Industrial Engineering Chemistry Research*, 49:8613-8623, 2010.

system mode may also change due to regular operating condition shifts. All these mode changes may follow some patterns, and thus, it will be beneficial to consider the mode dependency when performing diagnosis. The temporal dependency of the underlying modes will be considered in this chapter. This chapter is concluded by integrating the two solutions for the comprehensive temporal-dependency problem.

The remainder of this chapter is organized as follows. A recursive solution based on the deliberation of mode dependency and the associated data-driven algorithm for mode transition probability estimation are detailed Section 5.2. After that, the evidence and mode dependency are incorporated into an auto-regressive hidden Markov model framework in Section 5.3. The data dependency handling ability of the proposed approach is demonstrated by examining the diagnostic performance in simulation and experiment problems in Sections 5.4 and 5.5. The final section concludes this chapter with a discussion of the results achieved.

5.2 Mode dependency

Note that in this section, we consider only the mode dependency problem. The most general case when both the mode and evidence have temporal dependency will be addressed in the next section.

5.2.1 Dependent mode

The dynamic Bayesian model [26] is used to represent the mode dependency. Consider that the mode transition follows a Markov process, in which a mode is dependent only on its immediate previous neighbor in the temporal domain. By adding directed lines that represent dependencies between the consecutive modes depicted in Figure 4.1, a graphic model with temporal-dependent modes as shown in Figure 5.1 is constructed. The model is also known as the hidden Markov model [68].

To estimate the probability of the current system mode, previous evidence also

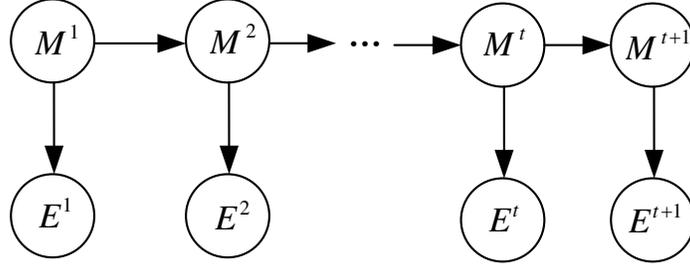


Figure 5.1: Bayesian model considering dependent mode

needs to be taken into consideration due to the dependency of the underlying modes,

$$\begin{aligned}
& p(M^t | E^1, \dots, E^t) \\
&= \frac{p(E^1, \dots, E^t | M^t) p(M^t)}{p(E^1, \dots, E^t)} \\
&= \frac{\sum_{M^{t-1}} p(E^1, \dots, E^t, M^{t-1} | M^t) p(M^t)}{p(E^1, \dots, E^t)} \\
&= \frac{\sum_{M^{t-1}} p(E^t | M^t, M^{t-1}, E^1, \dots, E^{t-1}) p(E^1, \dots, E^{t-1}, M^{t-1} | M^t) p(M^t)}{p(E^1, \dots, E^t)}.
\end{aligned} \tag{5.1}$$

According to the Markov property, given current mode M^t , current evidence E^t is conditionally independent of the previous modes M^{t-i} , where $i \geq 1$, and previous evidences E^{t-j} , where $j \geq 1$, so

$$\begin{aligned}
& p(M^t | E^1, \dots, E^t) \\
&= \frac{\sum_{M^{t-1}} p(E^t | M^t) p(E^1, \dots, E^{t-1}, M^{t-1} | M^t) p(M^t)}{p(E^1, \dots, E^t)}.
\end{aligned} \tag{5.2}$$

In the above equation, the term $p(E^1, \dots, E^{t-1}, M^{t-1} | M^t)$ can be calculated based on Bayes' rule:

$$\begin{aligned}
& p(E^1, \dots, E^{t-1}, M^{t-1} | M^t) \\
&= \frac{p(M^t | E^1, \dots, E^{t-1}, M^{t-1}) p(E^1, \dots, E^{t-1}, M^{t-1})}{p(M^t)}.
\end{aligned} \tag{5.3}$$

Insert Equation 5.3 into 5.2,

$$\begin{aligned}
& p(M^t | E^1, \dots, E^t) \\
&= \frac{\sum_{M^{t-1}} p(E^t | M^t) p(M^t | E^1, \dots, E^{t-1}, M^{t-1}) p(E^1, \dots, E^{t-1}, M^{t-1})}{p(E^1, \dots, E^t)}.
\end{aligned} \tag{5.4}$$

It is also known that M^t should be conditionally independent of E^1, \dots, E^{t-1} given M^{t-1} , so

$$\begin{aligned}
& p(M^t|E^1, \dots, E^t) \\
&= \frac{\sum_{M^{t-1}} p(E^t|M^t)p(M^t|M^{t-1})p(E^1, \dots, E^{t-1}, M^{t-1})}{p(E^1, \dots, E^t)} \\
&= \frac{\sum_{M^{t-1}} p(E^t|M^t)p(M^t|M^{t-1})p(M^{t-1}|E^1, \dots, E^{t-1})p(E^1, \dots, E^{t-1})}{p(E^1, \dots, E^t)} \\
&= p(E^t|M^t)p(E^1, \dots, E^{t-1}) \frac{\sum_{M^{t-1}} p(M^{t-1}|E^1, \dots, E^{t-1})p(M^t|M^{t-1})}{p(E^1, \dots, E^t)},
\end{aligned} \tag{5.5}$$

which means

$$p(M^t|E^1, \dots, E^t) \propto p(E^t|M^t) \sum_{M^{t-1}} p(M^{t-1}|E^1, \dots, E^{t-1})p(M^t|M^{t-1}). \tag{5.6}$$

Equation 5.6 provides a recursive solution for calculating the probability of the current system mode in the presence of mode dependency. To apply this equation, in addition to the single evidence likelihood, $p(E^t|M^t)$, the mode transition probability, $p(M^t|M^{t-1})$, also needs to be estimated, which will be discussed shortly.

In addition to the ability to handle the time domain information in terms of mode dependency, another advantage of the proposed approach lies in the fact that this approach is less likely to be affected by inaccurate prior probability. In view of Equation 5.6, there is no need to calculate prior probability. All the information required to compute the mode posterior probability includes the following: (1) the mode transition probability, $p(M^t|M^{t-1})$, and (2) the single evidence likelihood probability, $p(E^t|M^t)$; both (1) and (2) can be estimated from the historical evidence data set, and (3) the mode posterior probability as calculated diagnostic results from the previous diagnosis, $p(M^{t-1}|E^1, \dots, E^{t-1})$.

Although the prior probability is still necessary to calculate the first mode probability $p(M^1|E^1)$ when no ‘‘previous mode’’ is available,

$$p(M^1|E^1) \propto p(E^1|M^1)p(M^1), \tag{5.7}$$

the impact of the prior probability will diminish as more evidences surge, according to Equation 5.6. Rewrite the posterior probability equation for single evidence diagnosis according to Bayes’ rule,

$$p(M^t|E^t) \propto p(E^t|M^t)p(M^t). \tag{5.8}$$

A comparison of Equation 5.6 and Equation 5.8 shows that the term $p(E^t|M^t)$ appears in both equations, and they both update the information by considering newly emerging evidence; the remaining terms $\sum_{M^{t-1}} p(M^{t-1}|E^1, \dots, E^{t-1})p(M^t|M^{t-1})$ and $p(M^t)$ respectively in the two equations are determined before the arrival of the new evidence E^t . Thus, we may equivalently treat

$\sum_{M^{t-1}} p(M^{t-1}|E^1, \dots, E^{t-1})p(M^t|M^{t-1})$ as the ‘‘prior’’ in the recursion. The difference is that the term $\sum_{M^{t-1}} p(M^{t-1}|E^1, \dots, E^{t-1})p(M^t|M^{t-1})$ is constantly updated with new evidences and mode transition probability. Thus, it is not surprising to see that the impact of prior probability in the first recursion will eventually diminish as more evidence data samples are collected. Therefore, when sufficient evidence data samples are available, we can claim that the term

$\sum_{M^{t-1}} p(M^{t-1}|E^1, \dots, E^{t-1})p(M^t|M^{t-1})$ is completely determined by the data, and thus is free from the negative impact from inaccurate prior probability.

5.2.2 Estimation of mode transition probability

The intention of estimating the mode transition probability is to make the estimated probabilities to be consistent with the historical evidence data set \mathcal{D} in which the mode dependency exists. Our goal is to calculate the likelihood probability of a mode M^t given previous mode M^{t-1} to reflect the Markov property, so every composite mode sample, which is defined for mode transition probability estimation purpose, should include two elements:

$$d_M^{t-1} = \{M^{t-1}, M^t\}. \quad (5.9)$$

It should be noted that since our focus is only on the mode transition, not the evidence transition, the composite mode sample d_M^t includes only the transitions of the underlying modes, i.e., two consecutive modes. Accordingly, the new composite mode data set \mathcal{D}_M , which is assembled from the historical evidence data set \mathcal{D} to estimate of the mode transition probability, is defined as

$$\begin{aligned} \mathcal{D}_M &= \{d_M^1, \dots, d_M^{\tilde{N}-1}\} \\ &= \{(M^1, M^2), \dots, (M^{\tilde{N}-1}, M^{\tilde{N}})\}. \end{aligned} \quad (5.10)$$

Figure 5.2 depicts how the collected historical evidence data are organized to form composite mode samples. In Figure 5.2, the nodes highlighted with shadows or enclosed by the solid-lined frame constitute a single composite mode sample, as described by Equation 5.9.

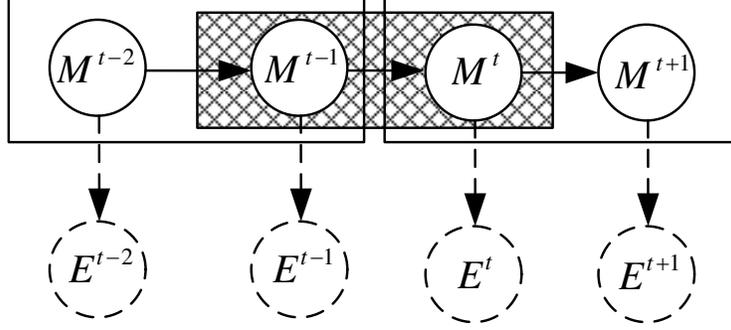


Figure 5.2: Historical composite mode data set for mode transition probability estimation

Following an approach similar to that outlined in Chapter 2, the mode transition probability can be derived as

$$p(m_v|m_u, \mathcal{D}_M) = \frac{\hat{n}_{u,v} + c_{u,v}}{\hat{N}_u + C_u}, \quad (5.11)$$

where $\hat{n}_{u,v}$ is the number of mode transitions from m_u to m_v in the historical composite mode data set; $\hat{N}_u = \sum_j \hat{n}_{u,j}$ is the total number of mode transitions, from mode m_u to any other mode; $c_{u,v}$ is the number of prior samples for the mode transition from m_u to m_v , and $C_u = \sum_j c_{u,j}$ is the total number of prior mode transitions from m_u to any other mode. See Appendix B for the derivation of Equation 5.11.

Rewrite the single evidence likelihood calculation equation

$$p(e_i|m_j, \mathcal{D}) = \frac{n_i + a_i}{N + A}, \quad (5.12)$$

and the evidence transition likelihood calculation equation,

$$p(e_r|e_s, m_c, \mathcal{D}_E) = \frac{\tilde{n}_{s,r} + b_{s,r}}{\tilde{N}_s + B_s} \quad (5.13)$$

By comparing Equations 5.12 and 5.13 with Equation 5.11, we can see that the mode transition probability is also determined by both prior samples and historical data samples, similar to the single evidence likelihood calculation and evidence transition probability calculation introduced in Equations 5.12 and 5.13. The difference lies in how the numbers of the prior and historical samples are counted. In Equations 5.12 and 5.13 the prior and historical samples refer to a count of the evidence samples or evidence transitions that correspond to a target mode, while in Equation 5.11 the prior and historical samples refer to the count of the composite mode samples assembled from the historical evidence samples.

As the number of historical composite modes increases, the transition probability will converge to the relative frequency determined by the historical composite mode samples, and the influence of the priors will decrease. The number of prior samples can be interpreted as prior belief of the mode transition probability distribution, where the uniform distribution indicates that prior sample numbers are equal across all possible transitions. It is important to set nonzero prior sample numbers; otherwise, unexpected results may occur in the case of limited historical samples [73].

In Chapter 4, an evidence space reduction solution is proposed to handle the intensive historical data needed for the evidence transition probability estimation. This space reduction, however, is unnecessary for the mode transition estimation. The problem with the evidence transition probability estimation arises due to the large combinatorial number of evidence transitions. The limited historical evidence transition data are divided into the small subspaces defined by different evidence transitions and the underlying modes. For example, consider a diagnostic system with ten monitors and ten modes, and each single monitor output is discretized into two discrete values. The total number of possible evidence transitions equals

$$K = 2^{10} \cdot 2^{10} = 1048576. \quad (5.14)$$

With the underlying modes being further considered, which have ten different statuses, there is a total of approximately 10^7 possible subspaces for each single evidence transition sample to fall into. However, the total number of possible mode transitions equals

$$K = 10 \cdot 10 = 100, \quad (5.15)$$

which does not increase exponentially with the number of the modes. Mode transition space reduction is not required.

5.3 Dependent evidence and mode

Up to now, we have discussed two different types of temporal dependency: evidence dependency, addressed in Chapter 4, and mode dependency, addressed in this chapter. In reality, both kinds of dependencies may exist. Thus, it is necessary to consider the evidence dependency and the mode dependency simultaneously.

With both mode and evidence dependencies being considered, a dynamic Bayesian model [26] is established to depict the temporal dependency between data samples.

By adding directed lines, which represent dependencies, between consecutive modes and evidences as shown in Figure 4.1, the proposed model is constructed in Figure 5.3. The model structure is also known as the auto-regressive hidden Markov model [68].

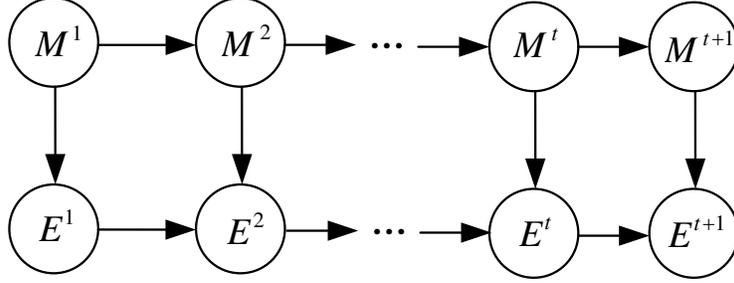


Figure 5.3: Dynamic Bayesian model that considers both mode and evidence dependency

Following similar derivations as in the previous section, a recursive solution can be developed. To estimate the probability of the current system mode, the previous evidences also need to be taken into consideration due to the evidence dependency in the time domain,

$$\begin{aligned}
& p(M^t | E^1, \dots, E^t) \\
&= \frac{p(E^1, \dots, E^t | M^t) p(M^t)}{p(E^1, \dots, E^t)} \\
&= \frac{\sum_{M^{t-1}} p(E^1, \dots, E^t, M^{t-1} | M^t) p(M^t)}{p(E^1, \dots, E^t)} \\
&= \frac{\sum_{M^{t-1}} p(E^t | M^t, M^{t-1}, E^1, \dots, E^{t-1}) p(E^1, \dots, E^{t-1}, M^{t-1} | M^t) p(M^t)}{p(E^1, \dots, E^t)}.
\end{aligned} \tag{5.16}$$

Note that the current evidence is determined by both the current mode M^t and previous evidence E^{t-1} ,

$$\begin{aligned}
& p(M^t | E^1, \dots, E^t) \\
&= \frac{\sum_{M^{t-1}} p(E^t | M^t, E^{t-1}) p(E^1, \dots, E^{t-1}, M^{t-1} | M^t) p(M^t)}{p(E^1, \dots, E^t)}.
\end{aligned} \tag{5.17}$$

By following procedure similar to that in the previous section, a recursive solution

can be expressed as follows,

$$p(M^t|E^1, \dots, E^t) \propto p(E^t|M^t, E^{t-1}) \sum_{M^{t-1}} p(M^{t-1}|E^1, \dots, E^{t-1})p(M^t|M^{t-1}). \quad (5.18)$$

With previous mode posterior $p(M^{t-1}|E^1, \dots, E^{t-1})$, two transition probabilities are required to calculate the current mode posterior, namely, the evidence transition probability $p(E^t|M^t, E^{t-1})$ and the mode transition probability $p(M^t|M^{t-1})$. The estimation algorithms of the first probability have been detailed in Chapter 4, and the second has been developed in the previous section of this chapter.

With data dependency being considered, the ability to retrieve the time domain information hidden in both the evidence and mode transitions, along with insensitivity to inaccurate prior probability, provides a significant advantage to the proposed approach in comparison to the Bayesian diagnostic solution based on information solely from single evidence without considering the data dependency.

5.4 Simulation example

The simulated binary distillation column employed in previous chapters is selected to evaluate the four Bayesian diagnostic approaches. Both the mode and evidence dependencies are introduced into the experiment.

5.4.1 Diagnostic settings

When simulation parameters are set, it is assumed that a hardware abnormality, or an external disturbance change, has a much lower probability of disappearing once the abnormality occurs, which also means that it is more unlikely to shift from other modes to the NF or m_1 mode, which are neither hardware modes nor disturbance modes. Other transitions have the same probability. Further assume that the hardware of the system is prone to problems, and the NF and m_1 modes thus have low probabilities to persist. In summary, the mode transition probability matrix is shown in Equation 5.19:

$$P_m = \begin{matrix} & \begin{matrix} NF & m_1 & m_2 & \dots & m_9 \end{matrix} \\ \begin{matrix} NF \\ m_1 \\ \vdots \\ m_9 \end{matrix} & \begin{pmatrix} 0.02 & 0.02 & 0.12 & \dots & 0.12 \\ 0.02 & 0.02 & 0.12 & \dots & 0.12 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0.02 & 0.02 & 0.12 & \dots & 0.12 \end{pmatrix} \end{matrix}. \quad (5.19)$$

In addition to the original measurement noise, random binary bias with two different levels, which are defined as 0 and 1, respectively, is added into the FBP top and PCT bottom sensor signal to introduce the evidence dependency into the simulation. The change in the bias in the FBP top follows the transition probability matrices shown in Equation 5.21, where the transition probabilities of FBP bias under mode m_3 , $P_{m_3}^{FBP}$, are different from those under the other modes $P_{\neg m_3}^{FBP}$; the change in bias in the PCT bottom follows the transition matrices shown in Equation 5.20, where the transition probabilities of PCT bias under mode m_7 , $P_{m_7}^{PCT}$, are different from those under the other modes $P_{\neg m_7}^{PCT}$.

$$P_{m_3}^{FBP} = \begin{matrix} & 0 & 1 \\ 0 & \begin{pmatrix} 0.9 & 0.1 \\ 0.1 & 0.9 \end{pmatrix} \\ 1 & \end{matrix}, P_{\neg m_3}^{FBP} = \begin{matrix} & 0 & 1 \\ 0 & \begin{pmatrix} 0.1 & 0.9 \\ 0.9 & 0.1 \end{pmatrix} \\ 1 & \end{matrix}; \quad (5.20)$$

$$P_{m_7}^{PCT} = \begin{matrix} & 0 & 1 \\ 0 & \begin{pmatrix} 0.9 & 0.1 \\ 0.1 & 0.9 \end{pmatrix} \\ 1 & \end{matrix}, P_{\neg m_7}^{PCT} = \begin{matrix} & 0 & 1 \\ 0 & \begin{pmatrix} 0.1 & 0.9 \\ 0.9 & 0.1 \end{pmatrix} \\ 1 & \end{matrix}. \quad (5.21)$$

A total of 5000 consecutive evidence data samples are collected. Among the collected samples, the first 3000 are used to estimate the single evidence likelihood, evidence transition probability, and mode transition probability. The remaining 2000 samples are used for cross-validation.

The parameter settings of the Bayesian diagnostic system are summarized in Table 5.1. The evidence transition space is a high dimensional one. According to the procedure introduced in Chapter 4, the evidence transition space is compressed to reduce the intensive requirement of historical evidence data.

Table 5.1: Summary of Bayesian diagnostic parameters

Discretizaion	$k_i = 3(\text{“low”}, \text{“medium”}, \text{“high”}),$ $K = 3^{15} = 14348907$
Historical evidence data	A mixture of a total of 3000 from all the 10 modes
Prior samples	Uniformly distributed with prior samples, for single evidence space, evidence transition space, and mode transition space
Prior probabilities	$p(NF) = 0.2, p(m_1) = 0.1, p(m_{other}) = 0.0875$
Evaluation data	Mixture of 2000 samples from all possible modes

5.4.2 Diagnostic results

The diagnostic results in Figure 5.4 are obtained from the 2000 evaluation (cross-validation) evidence data samples. In Figure 5.4, the bars denote the number of actual occurrences of each mode in the validation data set, as well as the number of modes diagnosed by four different diagnostic approaches, namely, the Bayesian approach that ignores both dependencies, that considers evidence dependency only, that considers mode dependency only, and that considers both mode and evidence dependencies.

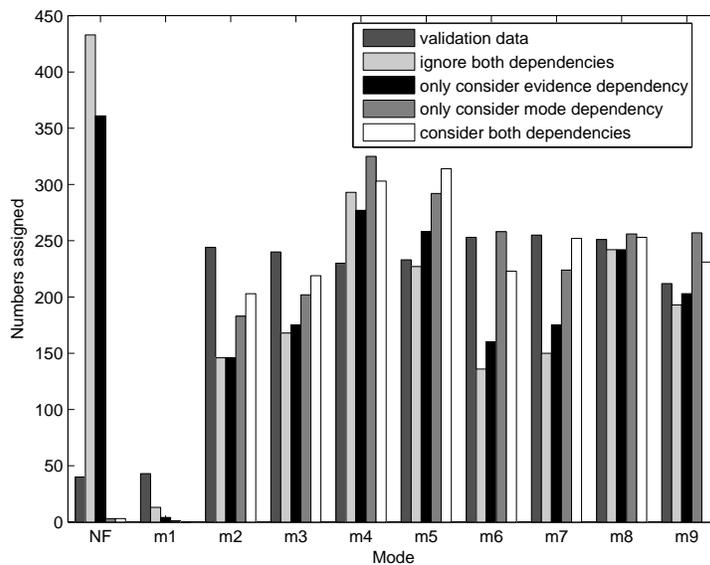


Figure 5.4: Number of occurrences diagnosed for each mode

In view of the numbers of the modes assigned by different approaches, the two Bayesian methods that ignore the mode dependency significantly overestimate the number of NF mode occurrences due to the methods' heavy dependency on prior probability: $p(NF) = 0.2$, $p(m_1) = 0.1$, and $p(m_{other}) = 0.0875$. With a high prior probability assigned to mode NF , the diagnostic system that ignores the mode dependency tends to yield higher posterior probability for the NF mode, and therefore overestimates the occurrences of the NF mode. The numbers of modes diagnosed by the remaining two approaches are close to the real numbers. The overall diagnostic performance is the best when both mode and evidence dependencies are taken into consideration, as summarized in Table 5.2.

Table 5.2: Overall correct diagnosis rates

		Evidence dependency	
		Ignore	Consider
Mode dependency	Ignore	69.92%	73.36%
	Consider	76.66%	82.21%

Further, consider the diagnostic performance for mode m_3 and m_7 , for which the disturbances are different from the other eight modes. Figure 5.5 displays the average posteriors assigned to the ten possible modes by the four approaches when the true underlying mode is m_3 and m_7 . In each subplot, the title denotes the diagnostic approach used, and the posterior probability assigned to the true underlying mode is highlighted with gray. The diagnostic conclusion is determined by picking up the mode with the largest posterior probability. The posterior probabilities assigned to the true underlying mode by the two approaches that consider the evidence dependency are higher than the other two that simply ignore the evidence dependency. Of the two approaches that ignore the evidence dependency, the posteriors assigned to the true underlying mode, generated by the method that considers the mode dependency, are higher, suggesting better diagnostic performance; furthermore, the probability assigned to the true underlying mode calculated by the proposed approach that considers both the mode and evidence dependencies is the highest, indicating the best diagnostic performance. The above discussion is further validated by the comparison of the correct diagnosis rates of the two modes in Table 5.3.

Table 5.3: Correct diagnosis rate for m_3 and m_7

	m_3	m_7
Ignoring both dependencies	70%	58.82%
Considering evidence dependency	84.17%	80%
Considering mode dependency	72.92%	68.63%
Considering evidence and mode dependencies	91.25%	95.29%

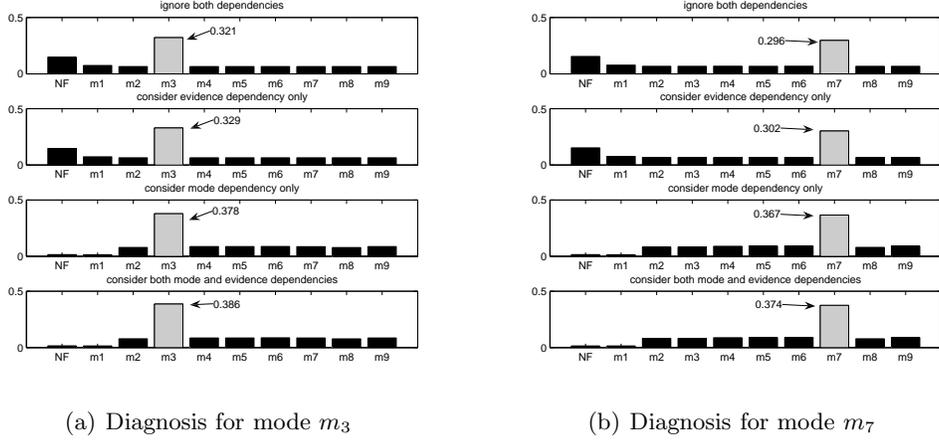


Figure 5.5: Posterior of mode m_3 and m_7

5.5 Pilot scale experiment

A pilot scale experiment is conducted to investigate the performance of the proposed Bayesian approach that considers both mode and evidence dependencies and to compare this strategy with the other Bayesian diagnostic strategies.

5.5.1 Process description

The experiment setup is a water tank with one inlet flow and two outlet flows. This equipment was used to test our previous diagnostic algorithms in Chapter 4, and is used here to test the new algorithm and compare it with the previous ones. Both the mode and evidence dependencies are introduced into the experiment.

The transition of the system modes follows the mode transition probability matrix in Equation 5.22. It is assumed that the system status tends to remain in the current condition, whether it is normal or faulty. Thus, the diagonal elements in the mode transition probability matrix are much larger than the others, as shown in Equation 5.22.

$$P_m = \begin{matrix} & \begin{matrix} NF & leakage & bias \end{matrix} \\ \begin{matrix} NF \\ leakage \\ bias \end{matrix} & \begin{pmatrix} 0.95 & 0.025 & 0.025 \\ 0.025 & 0.95 & 0.025 \\ 0.025 & 0.025 & 0.95 \end{pmatrix} \end{matrix}. \quad (5.22)$$

For the evidence dependency, consider that the disturbance is introduced through flow changes that fluctuate between two predefined rates. The random binary se-

quence of a limited frequency band is introduced into the inlet pump to simulate temporally dependent disturbances. By defining the high value as 1 and the low value as 0, the disturbance is introduced by following the transition probability matrices presented in Equation 5.23,

$$\begin{aligned}
 P_{NF}^{dis} &= \begin{matrix} & \begin{matrix} 0 & 1 \end{matrix} \\ \begin{matrix} 0 \\ 1 \end{matrix} & \begin{pmatrix} 0.9 & 0.1 \\ 0.2 & 0.8 \end{pmatrix} \end{matrix}, P_{leakage}^{dis} = \begin{matrix} & \begin{matrix} 0 & 1 \end{matrix} \\ \begin{matrix} 0 \\ 1 \end{matrix} & \begin{pmatrix} 0.1 & 0.9 \\ 0.8 & 0.2 \end{pmatrix}, \\
 P_{bias}^{dis} &= \begin{matrix} & \begin{matrix} 0 & 1 \end{matrix} \\ \begin{matrix} 0 \\ 1 \end{matrix} & \begin{pmatrix} 0.5 & 0.5 \\ 0.5 & 0.5 \end{pmatrix}, \tag{5.23}
 \end{aligned}$$

where P_M^{dis} represents the transition probability matrix of the introduced disturbance under mode M .

5.5.2 Diagnostic settings and results

A total of 600 evidence samples that correspond to the three modes are collected. The collected evidence data are divided into two portions to estimate the parameters, and for cross-validation. Table 5.4 summarizes the Bayesian diagnostic parameters.

Table 5.4: Summary of Bayesian diagnostic parameters

Discretizaion	$k_i = 2, K = 2^2 = 4$
Historical evidence data	A mixture of 360 samples
Prior samples	Uniformly distributed with prior samples, for single evidence space, evidence transition space, and mode transition space
Prior probabilities	$p(NF) = 0.5, p(m_{other}) = 0.25$
Evaluation data	A mixture of 240 independently generated cross-validation monitor readings

The diagnostic results in Figure 5.6 are obtained based on the cross-validation data. Due to inaccurate prior probabilities, the two Bayesian approaches that ignore mode dependency significantly overestimate the number of occurrences of the NF mode. Therefore, these approaches' overall correct diagnosis rates are much lower than the diagnosis rates of the methods that consider the mode dependency for the same prior probabilities, as shown in Table 5.5.

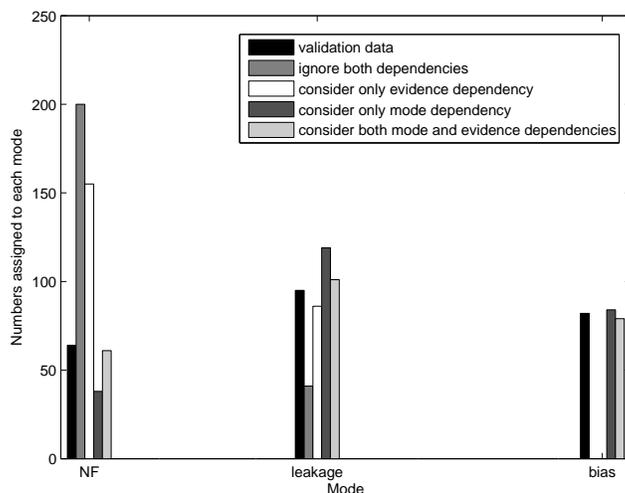


Figure 5.6: Numbers of occurrences diagnosed for each mode

Table 5.5: Overall diagnosis rates

		Evidence dependency	
		Ignore	Consider
Mode dependency	Ignore	37.76%	46.67%
	Consider	67.73%	75.93%

Figure 5.7 summarizes the diagnostic results in the form of average posterior probabilities. The title of each subplot denotes the true underlying mode from which the validation data come from, and the posterior probability corresponding to the true underlying mode is highlighted with light gray. The left-most panel summarizes the diagnostic results obtained by the approach that ignores all data dependencies; the next panel summarizes the diagnostic results obtained by the approach that considers evidence dependency only; the further next panel summarizes the diagnostic results by the approach that considers mode dependency only; and the right-most panel summarizes the diagnostic results obtained by the approach that considers both mode and evidence dependencies. The approach that considers both the mode and evidence dependencies always assigns the highest posterior probabilities to the true underlying modes, while the other approaches do not possess such diagnostic performance. Therefore, we can conclude that this last approach has a consistently better performance for all the modes, as also shown in Table 5.6.

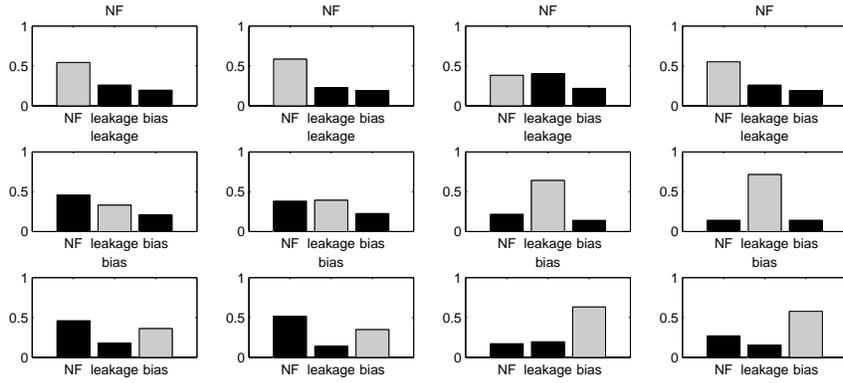


Figure 5.7: Average posteriors assigned to each mode

Table 5.6: Correct diagnosis rates for each single mode

	<i>NF</i>	<i>leakage</i>	<i>bias</i>
Ignoring both dependencies	98.44%	29.47%	0%
Considering evidence dependency	76.56%	66.32%	0%
Considering mode dependency	37.5%	81.05%	75.61%
Considering mode and evidence dependencies	67.19%	85.26%	71.95%

5.6 Conclusions

In this chapter, a data-driven approach based on the dynamic Bayesian model is presented to handle the temporal dependency problem in control loop diagnosis. Temporal dependencies of evidences as well as underlying modes are taken into consideration to achieve better diagnostic performance. A recursive solution for the mode posterior probability calculation is developed. The mode and evidence transition probabilities necessary for the recursive solution are estimated from the historical evidence data. The proposed method is applied to a simulated binary distillation column and a pilot scale experiment setup, where the performance of the proposed approach is shown to be better than approaches that ignore data dependencies.

Chapter 6

Control Loop Diagnosis with Sparse Historical Data

A major concern with the data-driven Bayesian approach is the intensive requirement for historical data. While in industry, the faulty data may be sparse. In extreme cases, a fault may only happen once, which causes difficulty for diagnosis. In this chapter, we propose an approach to estimate the statistical distribution of monitor readings when historical data is sparse. The monitor distributions are estimated with both analytical approaches and the bootstrap method. Applications of the proposed approach to the Tennessee Eastman Challenge problem and an experimental distillation column are presented to examine the performance of the proposed likelihood reconstruction methods.

6.1 Introduction

The Bayesian diagnostic methods discussed so far are all data-driven. To estimate the distribution of evidences, which is required by any Bayesian diagnostic method, sufficient historical samples must be collected. Otherwise poor diagnostic results are inevitable. A good example is shown in Chapter 4, where the number of historical evidences is much smaller than the dimension of the evidence transition space, and as a result, the estimation of evidence transition likelihood is poor. As discussed in previous chapters, the dimension of evidence space will grow exponentially as the number of monitors increases. Thus a large number of historical evidences are required for a reasonable diagnosis for a medium or large scale system. However, the faulty data in industry can be very sparse. In extreme cases, a fault may only

A portion of this chapter has been published in F. Qi, and B. Huang. Estimation of distribution function for control valve stiction estimation. *Journal of Process Control*, 21:1208-1216, 2011.

appear once. Estimating the evidence likelihood with limited number of historical samples is a challenging problem.

Knowing the nominal value of the monitor output, which is obtained from the sparse evidence samples, nonetheless, is still not sufficient for the Bayesian diagnosis. Recall that the core of the Bayesian diagnosis is the estimation of evidence likelihood. Only knowing nominal value of the evidence will result in a likelihood of one for the discrete bin which the nominal value falls in, and zeros for all the other bins. Such kind of likelihood completely removes the uncertainty, and will not fit into the Bayesian diagnostic framework. In order to have a reasonable diagnostic result, the uncertainty should be estimated and incorporated into the diagnostic framework.

The uncertainty of monitor output, either continuous or discrete, originates from disturbance and the infinite window length of monitor calculation: if there is no disturbance, or the segment window for monitor calculation contains infinite process data samples, the monitor output is solely determined by the underlying fault. To estimate the uncertainty, the distribution of continuous monitor output needs to be reconstructed. In this chapter, we will show how the monitor output distribution can be estimated using an analytical approach and a data-driven approach with limited number of historical evidence samples. It should be noted that there exists a large number of monitoring algorithms, and we will focus on some of the critical ones as examples to demonstrate the proposed distribution function estimation. The general idea, which is to establish relations between the nominal monitor output value and the monitor output distribution, can be extended to other monitoring algorithms.

The remainder of the chapter is organized as follows: in Section 6.2, the distributions of control performance monitor, sensor bias monitor and model validation monitor are determined with analytical approaches. Section 6.3 presents a bootstrap approach to estimate the distribution of valve stiction monitor output. The proposed distribution estimation approaches are applied to the Tennessee Eastman Challenge problem and an experimental distillation column setup in Sections 6.4 and 6.5. This chapter is concluded by Section 6.6.

6.2 Analytical estimation of monitor output distribution function

The idea to be discussed in this section is to derive the monitor output distribution as an analytical function of nominal monitor output (mean value), such that

the distribution function can be estimated with even only one historical evidence sample. Recall that each historical evidence sample is calculated from a section of process data sampled from physical process variables. Therefore, it is possible to derive the distribution of the corresponding evidence sample. In this section, several monitoring algorithms are selected for the analytical derivation of distribution function to illustrate the proposed approach. These algorithms include the minimum variance control performance monitor [36], process model monitor [3], and sensor bias monitor [77].

6.2.1 Control performance monitor

The univariate control performance assessment method adopted in this work is the minimum variance benchmark [36]. Filtering and CORrelation (FCOR) algorithm is employed to compute the control performance index of single controlled variable (CV).

A stable closed-loop process can be modeled as an infinite-order moving-average process:

$$y_t = (f_0 + f_1q^{-1} + \dots + f_{d-1}q^{-(d-1)} + f_dq^{-d} + \dots)a_t, \quad (6.1)$$

where d is the process delay, and a_t is white noise. Multiplying Equation 6.1 by $a_t, a_{t-1}, \dots, a_{t-d+1}$ respectively, and then taking the expectation of both sides of the equation yields

$$\begin{aligned} r_{ya}(0) &= E[y_t a_t] = f_0 \sigma_a^2 \\ r_{ya}(1) &= E[y_t a_{t-1}] = f_1 \sigma_a^2 \\ &\vdots \\ r_{ya}(d-1) &= E[y_t a_{t-d+1}] = f_{d-1} \sigma_a^2 \end{aligned}$$

The minimum variance or the invariant portion of output variance is [36]

$$\sigma_{mv}^2 = (f_0^2 + f_1^2 + \dots + f_{d-1}^2) \sigma_a^2. \quad (6.2)$$

Define the control performance index as

$$\eta(d) = \frac{\sigma_{mv}^2}{\sigma_y^2}, \quad (6.3)$$

where σ_y^2 can be extracted from the output data.

In Desborough and Harris (1992) [16], the values of the mean and variance of the performance index are estimated as

$$\text{mean}[\hat{\eta}(d)] = \eta(d) \quad (6.4)$$

$$\text{var}[\hat{\eta}(d)] = \frac{4}{n}(1 - \eta(d))^2 \left[\sum_{k=1}^{d-1} (\rho_k - \rho_{e,k})^2 + \sum_{k=d}^{\infty} \rho_k^2 \right], \quad (6.5)$$

where d is the process time delay; ρ_k and $\rho_{e,k}$ represent the output and residual auto-correlations respectively, and can be calculated from the available process data. Readers are referred to Desborough and Harris (1992) [16] for the derivation details. It is also shown in Desborough and Harris (1992) [16] that distribution of the control performance can be approximated by a normal distribution. Therefore, with the nominal value from one evidence available, the variance of the monitor output can be calculated to give an estimation of the monitor output distribution.

6.2.2 Process model monitor

A local Output Error (OE) method has been employed to validate the nominal process model [3].

For a multi-input single-output (MISO) subsystem, we have

$$Y(s) = G_1(s)e^{-\delta_1 s}U_1(s) + \dots + G_k(s)e^{-\delta_k s}U_k(s) + V(s), \quad (6.6)$$

where $U_i(s)$ is the the Laplace transform of the i th input, $G_i(s) = \frac{B_i(s)}{A_i(s)}$, $A_i(s) = a_{i,n}s^{n_i} + \dots + a_{i,1}s + a_{i,0}$, $B_i(s) = b_{i,n}s^{n_i} + \dots + b_{i,1}s + b_{i,0}$, $Y(s)$ is the Laplace transform of output, and the model parameters are given by

$$\theta = [\theta_1^T, \theta_2^T, \dots, \theta_k^T]^T,$$

$$\theta_i = [a_{i,n_i}, \dots, a_{i,0}, b_{i,n_i}, \dots, b_{i,0}, \delta_i]^T.$$

Define the overall model output $\hat{y}(t|\theta) = \mathcal{L}^{-1}[\hat{Y}(s)]$, and output corresponding to the i th input channel $\hat{y}_i(t|\theta) = \mathcal{L}^{-1}[\hat{G}_i(s)e^{-\delta_i s}U_i(s)]$. We have

$$\hat{y}(t|\theta) = \hat{y}_1(t|\theta_1) + \dots + \hat{y}_k(t|\theta_k). \quad (6.7)$$

Primary residuals and improved residuals are defined as

$$\rho(\theta, x_t) = \varphi(t)(y(t) - \hat{y}_k(t|\theta)) = \varphi(t)e(t, \theta) \quad (6.8)$$

$$\xi_N(\theta) = \frac{1}{\sqrt{N}} \sum_{t=1}^N \varphi(t)e(t, \theta), \quad (6.9)$$

where

$$\varphi(t|\theta) = -\frac{\partial \hat{y}(t|\theta)}{\partial \theta} = \begin{pmatrix} -\frac{\partial \hat{y}_1(t|\theta_1)}{\partial \theta_1} \\ \vdots \\ -\frac{\partial \hat{y}_k(t|\theta_k)}{\partial \theta_k} \end{pmatrix} = \begin{pmatrix} \varphi_1(t|\theta_1) \\ \vdots \\ \varphi_k(t|\theta_k) \end{pmatrix},$$

$$\varphi_i(t|\theta_i) = [\hat{y}_i^{(n_i)}(t), \dots, \hat{y}_i(t), -\hat{u}_i^{(n_i)}(t^{*i}), \dots, -\hat{u}_i^{(n_i)}(t^{*i}), \hat{y}_i^1(t)]^T,$$

$t^{*i} = t - \delta_i$, $\hat{u}_i^{(j)} = \mathcal{L}^{-1}[\frac{s^j}{A_i(s)}U_i(s|\theta)]$, and $\hat{y}_i^{(j)} = \mathcal{L}^{-1}[\frac{s^j}{A_i(s)}Y_i(s|\theta)]$. Then calculate the generalized likelihood ratio test as $I = \xi_N(\theta)^T \Sigma^{-1}(\theta) \xi_N(\theta)$, where $\Sigma(\theta) = \sum_{t=-\infty}^{\infty} cov(\rho(\theta, x_t), \rho(\theta, x_t))$.

The above I value can be used as model validation monitor of a MISO system. The model monitoring index is subject to χ^2 distribution when there is no model mismatch, and is subject to a non central χ^2 distribution when mismatch exists. A MIMO system can be divided into several MISO parts by different outputs. Model of each MISO part can be validated with the local OE approach, and therefore the overall MIMO model will be validated.

When there is a model mismatch, the improved residual, whose dimension is p , will be subject to a nonzero multivariate normal distribution,

$$\xi(\theta) \sim N(\mu, \Sigma) \quad (6.10)$$

where $\mu \neq 0$. As a result, the model monitor index, which is calculated as

$$I = \xi^T S^{-1} \xi, \quad (6.11)$$

will no longer follow χ^2 distribution. The distribution of I , has to be re-modeled.

Define $d = \Sigma^{-1/2} \xi$, $M = \Sigma^{-1/2} S \Sigma^{-1/2}$ [62] and

$$\alpha = (n-1)(d^T M^{-1} d / d^T d) d^T d = (n-1)I. \quad (6.12)$$

Knowing that M is subject to Wishart distribution [62],

$$M \sim W_p(\Sigma, n-1), \quad (6.13)$$

and its distribution is independent of that of d , the distribution of $\beta = d^T d / d^T M^{-1} d$ is χ_{n-p}^2 . Furthermore, the mean value of d is nonzero, therefore the distribution of $d^T d$ should be a noncentral χ^2 distribution with degree of p , and the noncentral parameter

$$\lambda = n\mu^T \Sigma^{-1} \mu. \quad (6.14)$$

Thus α can be represented as the ratio of a noncentral χ^2 variable and a central χ^2 variable, which in turn is a noncentral F distribution, with noncentral parameter λ ,

$$\alpha = (n-1)\chi_{p,\lambda}^2/\chi_{n-p}^2 = (n-1)p/(n-p)F_{p,n-p,\lambda}. \quad (6.15)$$

As such, I can be calculated as

$$I = \frac{1}{(n-1)}F_{p,n-p,\lambda}. \quad (6.16)$$

According to Evans (2000) [20], mean value the noncentral $F_{p,n-p,\lambda}$ distribution is

$$\mu_F = \frac{(n-p)(p+\lambda)}{p(n-p-2)} \quad (6.17)$$

Consequently, we can estimate the mean value of I ,

$$E[I] = \frac{1}{n-p} \frac{(n-p)(p+\lambda)}{p(n-p-2)} \quad (6.18)$$

In Equation 6.18, the only unavailable variable is λ and $E[I]$. We can substitute the $E[I]$ with the nominal monitor output. Once $E[I]$, n and p are all known, λ can be calculated with Equation 6.18, and thereafter the distribution of I can be reconstructed as per Equation 6.16.

6.2.3 Sensor bias monitor

The sensor bias monitor discussed here is the algorithm proposed in Qin and Li (2001) [77].

A system with sensor fault can be described by the following state-space model:

$$\begin{cases} x(t+1) = Ax(t) + Bu(t) + d \\ y(t) = Cx(t) + y^f + o \end{cases}. \quad (6.19)$$

Define $Y_s = [y(t-s), y(t-s-1), \dots, y(t)]^T$, where s is the observability index of the system. It should be noted that $y(t-i)$, $i = 0, \dots, s$ are actual outputs subtracted with normal operating value y_o . Therefore, knowing the nominal operating point y_o is a necessary condition for sensor fault diagnosis.

For simplicity, we can assume that the state-space model is a minimal realization, say, $s = n$ for a single output system. Equation 6.19 can be transformed as [77]

$$Y_s = \Gamma_x(t-s) + Y_s^f + H_s U_s + G_s D_s + O_s, \quad (6.20)$$

$$\text{where } \Gamma_x = \begin{pmatrix} C \\ CA \\ \vdots \\ CA^s \end{pmatrix}, H_s = \begin{pmatrix} 0 & \cdots & \cdots & 0 \\ CB & \ddots & & \vdots \\ \vdots & \ddots & \ddots & \vdots \\ CA^{s-1}B & \cdots & \cdots & 0 \end{pmatrix},$$

$$G_s = \begin{pmatrix} 0 & \cdots & \cdots & 0 \\ C & \ddots & & \vdots \\ \vdots & \ddots & \ddots & \vdots \\ CA^{s-1} & \cdots & \cdots & 0 \end{pmatrix}, \text{ and } U_s, D_s \text{ are all defined similarly as } Y_s. \text{ Denote}$$

$$Z_s = \begin{bmatrix} Y_s \\ U_s \end{bmatrix}. \quad (6.21)$$

Equation 6.20 can be written as

$$\begin{bmatrix} I & -H_s \end{bmatrix} Z_s = \Gamma_x(t-s) + Y_s^f + G_s D_s + O_s \quad (6.22)$$

Let characteristic polynomial of A be

$$|\lambda I - A| = \sum_{k=1}^n a_k \lambda^k \quad (6.23)$$

According to Cayley-Hamilton theorem, we know that

$$\sum_{k=1}^n a_k A^k = a_0 I + a_1 A + \cdots + a_n A^n = 0. \quad (6.24)$$

Defining $\Phi = [a_0 \ a_1 \ \cdots \ a_n]$, then Equation 6.24 gives

$$\Phi \Gamma_s = C(a_0 + a_1 A + \cdots + a_n A^n) = 0 \quad (6.25)$$

Multiplying both sides of Equation 6.22 with Φ yields

$$\begin{aligned} \Phi \begin{bmatrix} I & -H_s \end{bmatrix} Z_s &= \Phi \Gamma_x(t-s) + Y_s^f + G_s D_s + O_s \\ &= \Phi [Y_s^f + G_s D_s + O_s], \end{aligned} \quad (6.26)$$

so that the unknown state $x(t-s)$ is completely removed. Define $e^* = \Phi G_s D_s + \Phi O_s$, which is a scalar value. Then Equation 6.26 equals to

$$e = \Phi \begin{bmatrix} I & -H_s \end{bmatrix} Z_s = \Phi Y_s^f + e^* \quad (6.27)$$

D_s and O_s are process noise and output noise which are both Gaussian distributed, and e^* is linear combination of D_s and O_s , so it is readily known that e^* is also Gaussian distributed, $e^* \sim N(0, R_{e^*})$, where R_{e^*} is the variance of $e^*(t)$.

If there is a sensor bias fault $E(y^f) \neq 0$, $e = \Phi Y_s^f + e^*$ will not have zero mean. Therefore, we can use the index $d = e^T R_e^{-1} e$ as observation of the sensor bias monitor, where R_e is the variance of e .

Similar to the model monitoring monitor, the sensor bias index follows χ^2 distribution when there is no sensor bias, and non-central χ^2 distribution when sensor bias exists. The only difference lies in that the sensor bias monitor is calculated with a scalar residual, while the process model monitor is based on a vector residual. Thus we can consider the distribution of sensor bias monitor as a special case of the process model monitor when $p = 1$. Following the same procedure in Section 6.2.2, the distribution of sensor bias monitor output can be estimated.

6.3 Bootstrap approach to estimate monitor output distribution function

In this section, a bootstrap method for monitor output distribution estimation is introduced. In contrast to the analytical approaches discussed in the previous section, the bootstrap method provides an empirical estimation of the monitor distribution function. The valve stiction monitor, whose distribution cannot be derived with analytical approaches, is presented as an example of applying the bootstrap approach.

6.3.1 Valve stiction identification

Several valve stiction detection methods have been proposed in the past decades, for instance, the method based on cross-correlation analysis proposed by Horch (2009) [32], the curve fitting method of He et al. (2007) [30], the higher order statistics method of Choudhury et al. (2004) [14], and the area comparison method of Singhal and Salsbury (2005) [91]. A comprehensive review and comparison of these methods can be found in Jelali and Huang (2009) [43]. However, valve stiction quantification remains to be a challenging problem. Only a few publications are available. Choudhury et al. (2008) [13] proposed a method to quantify stiction using the ellipse-fitting method. The PV vs. OP plot is fitted to an ellipse and the amount of stiction is estimated as the maximum width of the ellipse in the OP direction. Jelali (2008) [42] presented a global optimization based method to quantify the valve stiction. Good initial values of the stiction parameters, which are obtained by using the ellipse-fitting method, play an essential role for an accurate

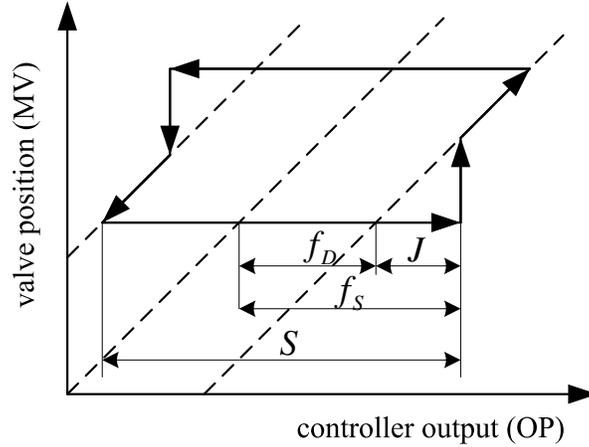


Figure 6.1: Operation diagram of sticky valve

estimation. A similar method was proposed by Srinivasan et al. (2005) [93]. The approach is based on identification of a Hammerstein model consisting of a sticky valve and a linear process. The stiction parameters and the model parameters are estimated simultaneously with a global grid optimization search method. Lee et al. (2009) [55] presented a stiction quantification approach. Given the stiction model structure, a feasible search domain of stiction model parameters is defined, and a constrained optimization is used in search of stiction parameters.

In Lee et al. (2009) [55], a method based on a type of Hammerstein model identification is proposed for the estimation of valve stiction parameters. The process (excluding sticky valve) is approximated by a linear transfer function model. The valve stiction model introduced by He et al. (2007) [30] is chosen to describe the nonlinearities invoked by the stiction valve.

Figure 6.1 shows the operation diagram of a sticky valve, where f_D is the kinetic friction band, f_S is the static friction band, S is the stick plus deadband,

$$S = f_S + f_D, \quad (6.28)$$

and J is the slip jump,

$$J = f_S - f_D. \quad (6.29)$$

If there is no stiction, the valve movement will follow the dashed line crossing the origin. Any change in the controller output, i.e., input to the valve, is matched

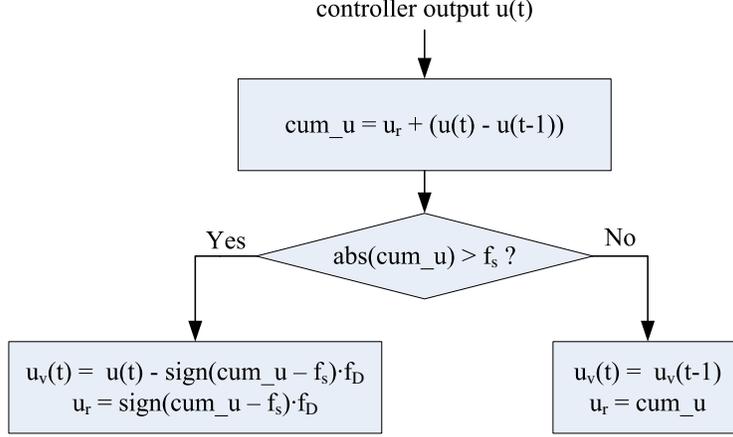


Figure 6.2: Stiction model flow diagram

exactly by the valve output, i.e., the valve movement. If there is stiction, the valve movement, will follow the solid line in Figure 6.1.

He et al. (2007) [30] proposed a valve stiction model as illustrated by the flow chart in Figure 6.2. u_r is the residual force applied to the valve which has not moved yet; cum_u is an intermediate variable describing the current force acting on the valve. If cum_u is larger than the static friction f_S , the valve position will equal to the controller output, subtracted by the dynamic stiction f_D ; if not, the valve will remain in the same position, and cum_u will be the residual force applied to the valve at the next time instant.

Stiction estimation can be considered as a Hammerstein model identification problem. The identification of the overall Hammerstein model is performed by a global optimization search for the stiction parameters, in conjunction with the identification of the linear transfer function model. In order to have an effective search for optimal estimation, the search space needs to be specified. The bounded search space can be defined by analyzing the collected operation data and the relationship of stiction parameters.

In view of Figure 6.1, it is noted that

$$f_D + f_S \leq S_{max}, \quad (6.30)$$

where S_{max} is the span of OP . Also consider the relation

$$f_S = f_D + J, \quad (6.31)$$

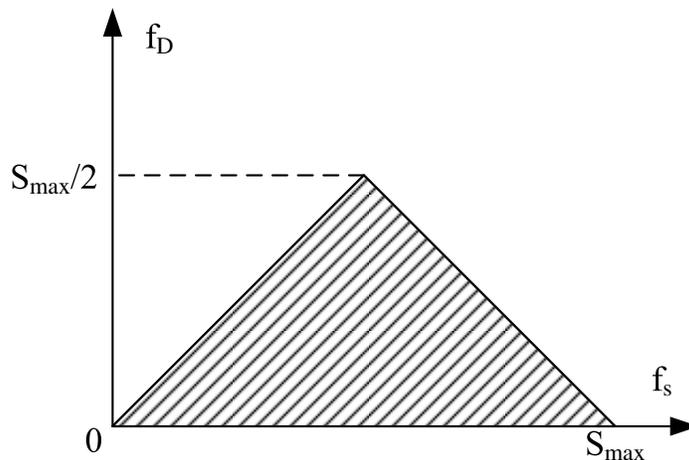


Figure 6.3: Bounded stiction parameter search space

so

$$2f_D + J \leq S_{max}. \quad (6.32)$$

Figure 6.3 shows the constrained search space for the stiction parameter set (f_s, f_D) .

For each point in the stiction parameter space, a series of intermediate valve output data (denoted as MV' , input to the linear dynamic model) can be calculated from the collected OP data as per the stiction model. With MV' and the collected PV data, the linear dynamic model, which is chosen as a first or second order plus dead time transfer function, is identified with the least squares method by minimizing the mean squared error. A grid optimization method is applied to search within the bounded stiction parameter space for the minimal mean square error so as to obtain the corresponding linear dynamic model and the stiction model parameters.

6.3.2 The bootstrap method

The quantification of uncertainty for nonlinear parameter estimation is nontrivial. Ninness et al. (2002) [70] investigated the accuracy of Hammerstein model estimation, and derived an approximated solution for the variance of identified nonlinear parameters. This result, however, requires that the nonlinear part of the Hammerstein model must be memoryless. In the valve stiction case, the current valve output (position) depends not only on the current controller output, but also on

the previous controller output series [30], and thus, the valve stiction model is a memory one. Extending the work of Ninness to nonlinear components with memory, for instance, a sticky valve, is an open problem [93]. As to the best of the author's knowledge, there is no work that has successfully handled the uncertainty quantification problem.

Bootstrap methods have been proven useful for a variety of statistical inferences, such as estimation of bias, variance, and distribution functions [9, 108]. Applications of bootstrap methods have been reported in medicine science, communication, image processing, physics, quality assurance, etc. [88, 8, 107, 33, 96]. Built on previous work of valve stiction quantification of Lee et al. (2009) [55], this section develops a bootstrap based approach to quantify the distribution of valve stiction parameter estimation.

The purpose of the bootstrap method is to estimate the distribution of parameter estimators [9]. Suppose that we have a set of data $x = \{x_1, x_2, \dots, x_N\}$ collected from the realizations of a random variable X , which follows the distribution F_X . Let θ be a parameter or statistics of the distribution F_X . $\hat{\theta}$ is a estimator of θ , which can be, for instance, the estimator for the mean value, $\hat{\theta} = \sum_{i=1}^N x_i/N$. If the distribution F_X is known and yet relatively simple, it will not be difficult to evaluate the distribution of the parameter estimator. A common example is the variance of the sample mean value of a normal distribution. However, if the distribution is unknown, or is too complicated to evaluate analytically, the bootstrap method provides a well suited alternative to estimate the distribution of the estimator by resampling the collected data.

Assume that the collected data samples $x = \{x_1, x_2, \dots, x_N\}$ are independently and identically distributed (i.i.d.). The collected samples $x = \{x_1, x_2, \dots, x_N\}$ are resampled to construct the bootstrap samples from the same distribution. Often the distribution is unknown and has to be estimated from the original samples. To construct the bootstrap samples is to estimate the kernel distribution of the collected sample set, and to generate the bootstrap samples from the estimated kernel distribution. The bootstrap sample sets have the same size as the original sample set, and can be denoted as

$$x_i^b = \{x_{i,1}^b, x_{i,2}^b, \dots, x_{i,N}^b\}, \quad (6.33)$$

where x_i^b is the i th bootstrap sample set, and $x_{i,j}^b$ is the j th sample in the i th

bootstrap sample set. Each bootstrap set is considered as a new set of data. Suppose that totally M sets of bootstrap samples are collected. For the i th bootstrap set, the bootstrap estimator of θ , $\hat{\theta}_i^b$, can be evaluated. Based on the M sets of bootstrap samples, we can have a group of bootstrap estimators,

$$\{\hat{\theta}_1^b, \hat{\theta}_2^b, \dots, \hat{\theta}_M^b\}. \quad (6.34)$$

With a sufficiently large M , the distribution of parameter estimator $\hat{\theta}$, $F_{\hat{\theta}}$, can be approximated by the distribution of $\hat{\theta}^b$, $F_{\hat{\theta}^b}$, which is determined from the bootstrap estimators $\hat{\theta}_i^b$, $i = 1, 2, \dots, M$.

An important assumption for the bootstrap is that the data samples that are to be bootstrapped must be i.i.d.. This assumption, however, has a clear limitation in control related applications. Once the collected data are dependent, the aforementioned bootstrap method will lead to incorrect estimation result of $F_{\hat{\theta}}$. A solution for the dependent data bootstrap is to whiten the data through a time series model, and then bootstrap the whitened data, but not the original data [108], as elaborated below.

Suppose that a set of dependent data $x = \{x_1, x_2, \dots, x_N\}$ is collected. To estimate the distribution of estimator $\hat{\theta}$, the first step is to fit the data into a time series model. With the fitted model, a set of simulated model output can be obtained, $\hat{x} = \{\hat{x}_1, \hat{x}_2, \dots, \hat{x}_N\}$. Subtracting the model output from the collected data, a set of i.i.d. residuals is obtained, $\hat{e} = \{e_1, e_2, \dots, e_N\}$. Perform bootstrap on the residuals, and we can generate M sets of bootstrap residual samples

$$e_i^b = \{e_{i,1}^b, e_{i,2}^b, \dots, e_{i,N}^b\}, \quad i = 1, \dots, M \quad (6.35)$$

where e_i^b is the i -th bootstrapped residual sample set, and $e_{i,j}^b$ is the j -th sample in the i -th bootstrapped residual sample set. The new sets of bootstrap sample for x_i can be obtained by adding the bootstrapped residual samples to the simulated model output,

$$x_i^b = \hat{x} + e_i^b. \quad (6.36)$$

It has been shown that with the above procedure, the bootstrap distribution is an asymptotically valid estimator of the distribution of parameter estimation [51, 4].

For a closed-loop system with sticky valve, the output data (PV) are almost always non-white. Thus we cannot apply the bootstrap method to the PV data

directly. Following the above data whitening procedure, we propose a residual bootstrap method based on the identification of both closed-loop model and the disturbance model.

Controller identification

The process model, which includes the linear dynamic model and the sticky valve model, can be identified by the method outlined in Section 6.3.1. In order to reconstruct the closed-loop model, the controller model is needed. If the controller is known, the closed-loop model can be readily constructed. However, in some scenarios when the controller is unknown, the controller has to be identified.

In this work, all controllers attached to the valves are assumed to be PI or PID controllers. Thus the controller model can be written as,

$$u(t) = K_p \epsilon(t) + K_i \sum_{i=1}^t \epsilon(i) + K_d (\epsilon(t) - \epsilon(t-1)), \quad (6.37)$$

where $u(t)$ is the controller output, and $\epsilon(t)$ is the controller error, $\epsilon(t) = r(t) - y(t)$. $r(t)$ is the control loop setpoint, and $y(t)$ is the process output.

The PID controller parameters

$$\theta = (K_p, K_i, K_d)^T$$

can be identified with the collected process data. Let

$$U = (u(2) \ u(3) \ \dots \ u(N))^T, \quad (6.38)$$

and

$$X = \begin{pmatrix} \epsilon(2) & \sum_{i=1}^2 \epsilon(i) & \epsilon(2) - \epsilon(1) \\ \epsilon(3) & \sum_{i=1}^3 \epsilon(i) & \epsilon(3) - \epsilon(2) \\ \vdots & \vdots & \vdots \\ \epsilon(N) & \sum_{i=1}^N \epsilon(i) & \epsilon(N) - \epsilon(N-1) \end{pmatrix}. \quad (6.39)$$

Accordingly we have

$$U = X \cdot \theta. \quad (6.40)$$

Following the least squares method, the PID parameters are calculated as

$$\theta = (X^T X)^{-1} X^T U. \quad (6.41)$$

In a closed control loop, two equations exist

$$y(t) = H(S, J, u(t))G_p(q^{-1}) + G_l(q^{-1})e(t), \quad (6.42)$$

$$u(t) = G_c(q^{-1})(r(t) - y(t)), \quad (6.43)$$

where $H(S, J, u(t))$ is the sticky valve model; $G_p(q^{-1})$ is the linear dynamic model; $G_l(q^{-1})$ is the disturbance model; $e(t)$ is white noise; $G_c(q^{-1})$ is the controller model; $r(t)$ is the setpoint signal, and in most valve stiction scenarios it is 0. These two equations can be written as

$$u(t) = F(S, J, (y(t) - G_l(q^{-1})e(t))G_p^{-1}(q^{-1})), \quad (6.44)$$

$$u(t) = -G_c(q^{-1})y(t), \quad (6.45)$$

where $F(S, J, \cdot)$ is the inverse function of $H(S, J, \cdot)$. The purpose of controller model identification is to find a model that can fit the process data $u(t)$ and $y(t)$,

$$u(t) = -\hat{G}_c(q^{-1})y(t) \quad (6.46)$$

In Equation 6.44, $F(S, J, \cdot)$ is a nonlinear function due to valve stiction, and the output $u(t)$ is corrupted by noise; in Equation 6.46, the controller model is linear and noise free, and the model structure is a perfect match of the real controller used if correctly selected. Identifiability of $G_c(q^{-1})$ can be easily proved.

Although only the PID controller is considered, the discussion of this work can be extended to controller of other structures by selecting the corresponding controller model structure. For example, some PID controllers have filters. In this case we can fit a second order or higher order model, with constraint that the denominator has a pole that equals one to account for the integrator. With the controller model, the sticky valve parameters and the linear dynamic model, the closed loop model can be built, and we are ready to generate bootstrap data, as discussed below.

Bootstrap of sticky valve parameters

When identifying the valve stiction parameters and linear dynamic model parameters, a set of residuals are generated from the identification. However, we cannot bootstrap the identification residuals directly. The reason lies in the fact that the valve stiction identification algorithm outlined in Section 6.3.1 only estimates the linear dynamic model as well as the valve stiction model, and the disturbance model has not been estimated. If the disturbance model is considered, the identification

of the linear dynamic model becomes a nonlinear optimization problem [60]. Considering the global grid search of stiction parameters and the need of identifying the linear dynamic model from each grid search, the overall computation will be formidable. The generated identification residuals are not temporally independent, and thus are not i.i.d.. Whitening of the residual is needed.

Let the residual be $e(t)$. Fit the residual by a time series model:

$$e(t) = \frac{C(q^{-1})}{B(q^{-1})}a(t), \quad (6.47)$$

where $a(t)$ is i.i.d. white noise. Since $a(t)$ is i.i.d., the bootstrap can be applied. Let,

$$a(t) \sim \mathcal{N}(\hat{\mu}, \hat{\sigma}^2), \quad (6.48)$$

where $\hat{\mu}$ and $\hat{\sigma}^2$ can be calculated from the filtered residuals $a(t)$. New sets of bootstrapped $a^b(t)$ are generated from the following normal distribution

$$a^b(t) \sim \mathcal{N}(\hat{\mu}, \hat{\sigma}^2), \quad (6.49)$$

The bootstrapped $a^b(t)$ are then passed to the model identified in Equation 6.47 to get bootstrap residuals $e^b(t)$.

The bootstrapped residuals $e^b(t)$ are added into the closed-loop model, to simulate new process data, including the process output (PV) and the controller output (OP), where the controller and model parameters for the simulation are the previously identified ones. It should be noted that the newly simulated bootstrap data set must have the same length as the original one. Use the valve stiction parameter identification method presented in Section 6.3.1, and a new set of stiction parameters can be estimated from the re-simulated closed-loop response data.

Repeat the above procedure for sufficient number of iterations, say, M times, and as a result, M different sets of stiction parameters are estimated. With the newly bootstrapped stiction parameters, the distribution of the valve stiction parameters can be determined. The procedure is summarized in Figure 6.4.

6.3.3 Illustrative example

To verify the proposed procedure, a SISO closed control loop with sticky valve is simulated. The linear part of the process model, together with the disturbance

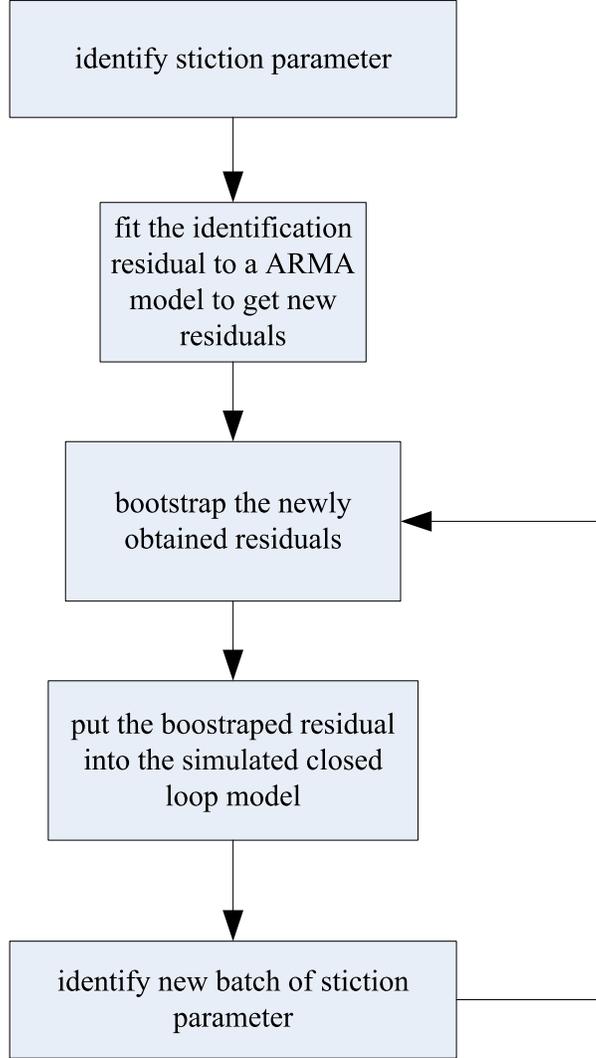


Figure 6.4: Bootstrap method flow diagram

model in Garatti and Arnaiz (2009) [23], is

$$y(t) = \frac{q^{-1}}{1 + q^{-1}}u_v(t) + (1 + q^{-1})a(t), \quad (6.50)$$

where $u_v(t)$ is the valve output (position), and $a(t)$ is Gaussian distributed white noise, $a(t) \sim \mathcal{N}(0, 0.01)$. A sticky valve model with the structure described in Section 6.3.1 is used to convert the controller output $u(t)$ into valve position $u_v(t)$. The valve stiction parameters are $S = 2$, and $J = 1$. The loop is controlled by a PI controller, with parameters $K_p = 0.5$, and $K_i = 0.1$.

Totally 2000 Monte-Carlo simulation runs are performed. Each simulation contains 1000 samples of process data. Based on each single simulation run, a set of

linear process model, disturbance model, valve stiction model, and controller parameters, are identified. The histograms of the 2000 sets of identified valve stiction parameter are shown in Figure 6.5.

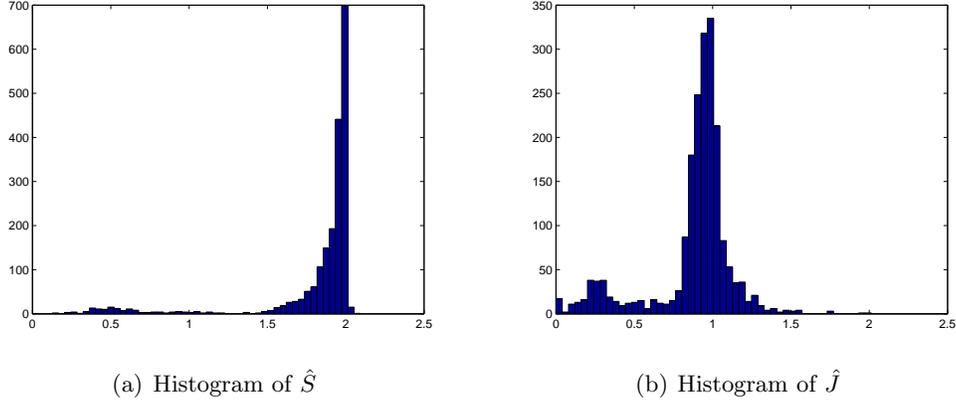


Figure 6.5: Histogram of simulated \hat{S} and \hat{J}

From each one of the simulations, a set of stiction and linear dynamic model parameters are estimated. With any set of identified linear dynamic model, stiction parameters, disturbance model, controller parameters, as well as the whitened residuals, we can generate bootstrap samples to determine the empirical distribution of the identified stiction parameters. Before that, we also need to assure that the whitened residuals are i.i.d.. The auto-correlation coefficients of the filtered 1000 residuals are shown in Figure 6.6 and the histogram of the residuals is shown in Figure 6.7. It can be observed that the whitened residuals are uncorrelated, and have a distribution close to normal. Thus the filtered residual can be used as the bootstrap variable to generate new samples.

The distributions of parameters based on the bootstrap from one set of the identification results are compared with the results from 1000 Monte-Carlo simulations based on true stiction parameters, as presented in Figure 6.8. In Figure 6.8, \hat{S} and \hat{J} are the stiction parameters estimated from the Monte-Carlo simulations, and \hat{S}^b and \hat{J}^b are the stiction parameters estimated from the bootstrap samples based on one of the 1000 Monte-Carlo simulations. As can be seen, the distributions of bootstrapped parameters \hat{S}^b and \hat{J}^b are close to the Monte-Carlo simulated results. This is also verified by comparing the sample standard deviations, as shown in Table 6.1.

To further quantify the accuracy of bootstrap estimation performance, Kullback-

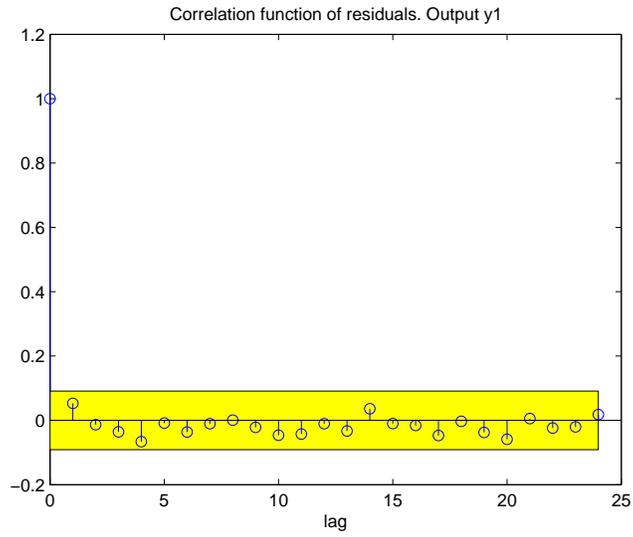


Figure 6.6: Auto-correlation coefficient of residuals

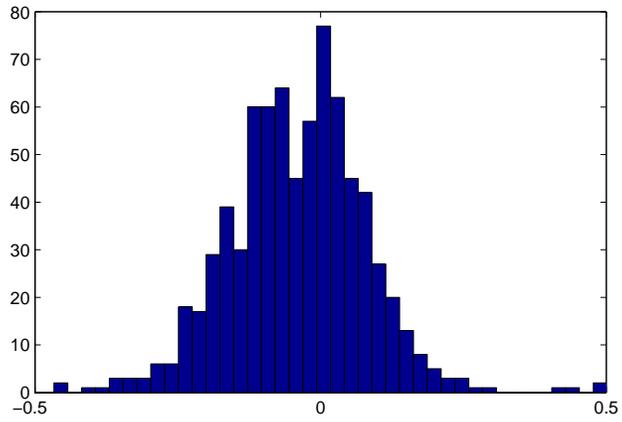
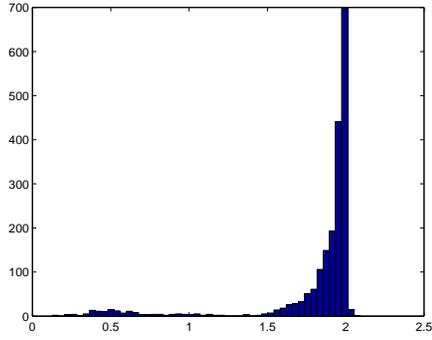
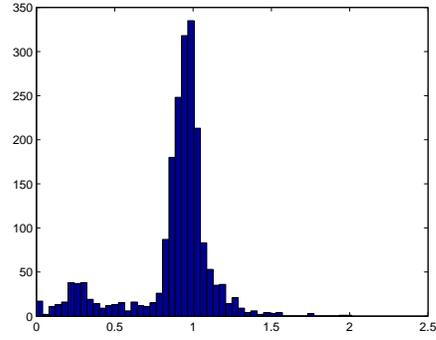


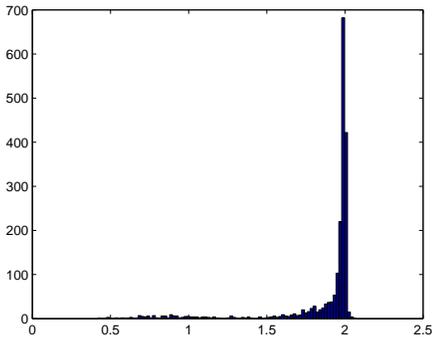
Figure 6.7: Histogram of residual distribution



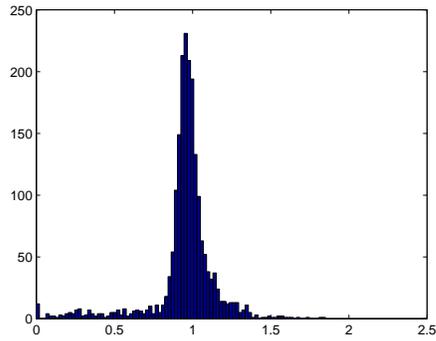
(a) Histogram of \hat{S}



(b) Histogram of \hat{J}



(c) Histogram of \hat{S}^b



(d) Histogram of \hat{J}^b

Figure 6.8: Comparison of parameter histograms

Table 6.1: Comparison of sample standard deviations

	$\sigma_{\hat{S}}$	$\sigma_{\hat{J}}$
simulated	0.3614	0.2645
bootstrapped	0.3013	0.2021

Leibler divergence is employed to measure the distance between the the bootstrap distribution and the simulated distributions. In information theory, the Kullback-Leibler (KL) divergence is a measure of the difference between two probability distributions P and Q [53].

For distributions P and Q of a continuous random variable, the KL divergency is defined as

$$D_{KL}(P||Q) = \int_{-\infty}^{\infty} p(x) \log \frac{p(x)}{q(x)} dx, \quad (6.51)$$

where p and q denote the densities of P and Q . While for probability distributions P and Q of a discrete random variable, the KL divergence of Q from P is defined as

$$D_{KL}(P||Q) = \sum_i P(i) \log \frac{P(i)}{Q(i)}. \quad (6.52)$$

Since the analytical density function is unavailable for both the simulated distribution and the bootstrapped distribution, the only way to calculate the KL divergence in the valve stiction case is to consider the two distributions as discrete distribution over a vector X . The elements of X are selected as a finite number of small consecutive intervals within the range of the stiction parameters. $P(i)$ or $Q(i)$ is the frequency over the i th interval. Choosing interval size of 0.05, the KL divergences of the two sets of distribution pairs are

$$D_{KL}(\hat{S}||\hat{S}^b) = 1.0285, \quad (6.53)$$

$$D_{KL}(\hat{J}||\hat{J}^b) = 0.8695, \quad (6.54)$$

which further confirm the good estimation performance of the bootstrap method.

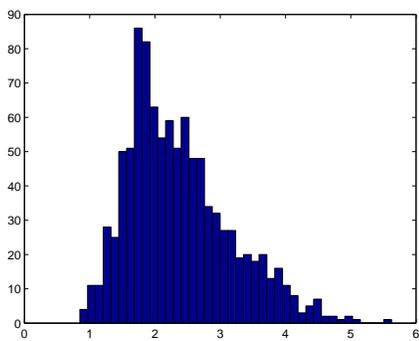
6.3.4 Applications

Several industrial data sets are selected to further investigate the performance of the proposed method. These data sets have been used in Jelali and Huang (2009) [43] and Thornhill et al. (2002) [99] to test the stiction detection performances.

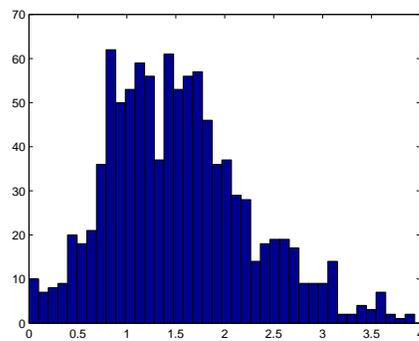
1000 bootstrap data sets are simulated for each loop. The histograms of the bootstrapped parameters are shown in Figures 6.9, 6.10, 6.11, and 6.12. The estimated stiction parameters and 95% confidence intervals (CIs) of the estimated parameters are summarized in Table 6.2. The distribution of stiction parameter estimates provides valuable information to determine the stiction.

Table 6.2: Confidence intervals of the identified stiction parameters

	chemical 55	chemical 60	paper 1	paper 9
Stiction?	yes	yes	yes	no
\hat{S}	1.6672	1.8809	4.4006	0.015
\hat{J}	0.6669	0.8996	3.5204	0.010
95% CI of \hat{S}^b	[1.1795, 4.1668]	[0.5860, 3.2911]	[0.2090, 10.4878]	[0, 0.1306]
95% CI of \hat{J}^b	[0.2852, 3.1174]	[0, 2.3051]	[0.0713, 9.6138]	[0, 0.1277]

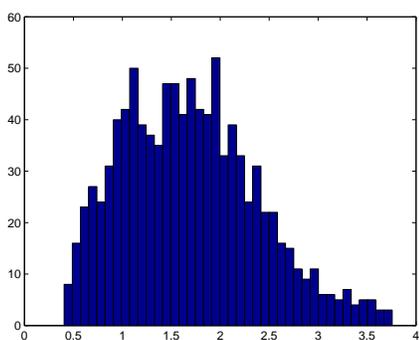


(a) Histogram of \hat{S}^b

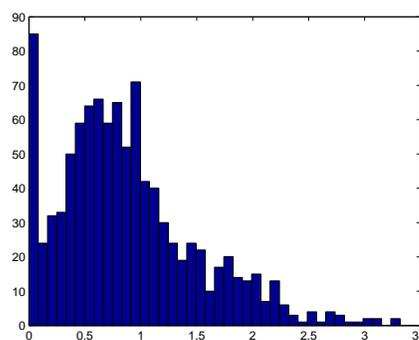


(b) Histogram of \hat{J}^b

Figure 6.9: Histogram of bootstrapped \hat{S}^b and \hat{J}^b for chemical 55

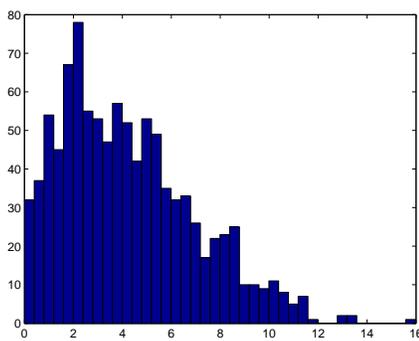


(a) Histogram of \hat{S}^b

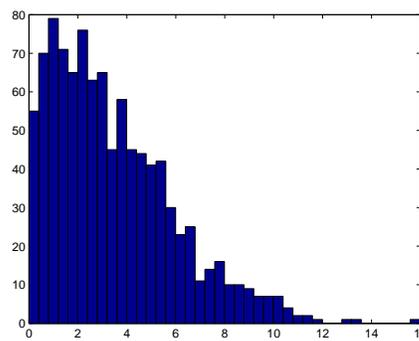


(b) Histogram of \hat{J}^b

Figure 6.10: Histogram of bootstrapped \hat{S}^b and \hat{J}^b for chemical 60



(a) Histogram of \hat{S}^b



(b) Histogram of \hat{J}^b

Figure 6.11: Histogram of bootstrapped \hat{S}^b and \hat{J}^b for paper 1

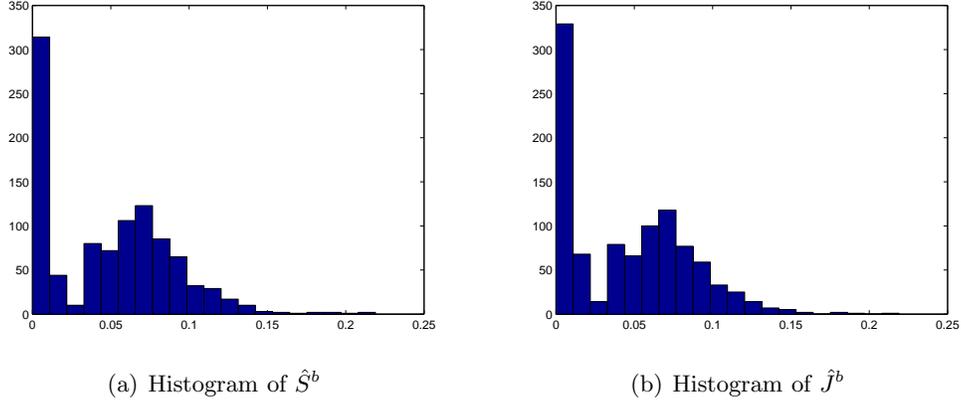


Figure 6.12: Histogram of bootstrapped \hat{S}^b and \hat{J}^b for paper 9

According to the data source [99], the valve has stiction problem in the “chemical 55” data set, and the estimated stiction parameters are nonzero for both S and J . The previous stiction identification methods only provide a point estimation but not a confidence interval. According to the distribution provided by the bootstrap method, we can observe that the 95% CIs do not include zero. Thus we can conclude that the valve do have stiction problem with 95% confidence.

For the “chemical 60” data set, the bootstrap results show that the 95% CI of \hat{J}^b includes the 0, although the point estimation suggests that \hat{J} is nonzero. Therefore we can conclude with 95% confidence that the valve only has a deadband problem, i.e., $S > 0$, but J might be zero

Both the two 95% CIs of “paper 5” do not include 0, indicating that the valve does have a stiction problem with 95% confidence. However, both the CIs of \hat{S}^b and \hat{J}^b are wide. The quantification of the valve stiction is of great uncertainty, i.e., the extent of stiction estimated from this set of data may be unreliable.

For the “paper 9” data set, the point estimation algorithm yields non-zero estimations for both \hat{S} and \hat{J} , which does not agree with the fact that there is no stiction in the loop [43]. However, in the 1000 bootstrap simulations, over 30% yield zero stiction parameters. The 95% CIs of both \hat{S} and \hat{J} cover 0. Thus 0 is still within the possible range of the stiction parameters. When diagnosing the valve problem, even though the point estimation may indicate stiction, it is important to check its confidence interval to avoid misleading conclusions.

It should be noted that the bootstrap method is not only applicable to the valve

stiction monitor. It can also be applied to other monitors. During the bootstrap procedures, the controller model, valve stiction model, and process model plus the disturbance model are all identified. A complete closed-loop model is built based on the identified models. Simulated samples can be generated to estimate the distribution of not only valve stiction monitors, but also the control performance monitor, process model monitor, and sensor bias monitor. The disadvantage of the bootstrap method is that the accuracy of the estimation may not be as good as the analytical solutions. The bootstrap method, however, estimates the distributions of the monitors in the same loop simultaneously, other than considering each monitor distribution individually. Therefore the cross monitor output dependency is taken into consideration. If cross monitor output dependency is of concern, the bootstrap method should be employed. As a rule of thumb, the bootstrap method shall be used on multiple monitors located in the same control loop. If monitors are located in different control loops, we can apply the analytical estimation techniques to each individual monitor.

With the distributions estimated by approaches presented in Sections 6.2 and 6.3, the likelihoods of single monitors can be calculated with the method outlined in Chapter 2 using the simulated data samples. Further assuming that the monitor distributions are cross-independent between control loops, the overall evidence likelihood can be calculated as

$$\begin{aligned}
 p(E|M) &= p(\pi_1, \pi_2, \dots, \pi_k|M) \\
 &= \prod_{i=1}^L p(\pi_{i,1}, \dots, \pi_{i,k_i}|M),
 \end{aligned} \tag{6.55}$$

where L is the number of control loops; $\pi_{i,j}$ is the j th monitor in the i th control loop; and k_i is the number of monitors in the i th control loop.

6.4 Simulation example

The Tennessee Eastman (TE) Challenge problem is selected to evaluate the monitor distribution estimation algorithms. TE problem provides a realistic industrial process for process control community [18] to test developed algorithms. It has been widely used as a benchmark to exam the performances of different control monitoring and diagnostic approaches [52, 64, 10, 78, 54, 61]. The structure of the process

is shown in Figure 6.4.

The gaseous reactants A, C, D, and E and the inert B are fed to the reactor where the liquid products G and H are formed. The reactions in the reactor are irreversible, exothermic, and approximately first-order with respect to the reactant concentrations. The reactor product stream is cooled through a condenser and then fed to a vapor/liquid separator. The vapor exiting the separator is recycled to the reactor feed through the compressor. A portion of the recycle stream is purged to keep the inert and byproducts from accumulating in the process. The condensed components from the separator are pumped to the stripper to strip the remaining reactants, which are combined with the recycle stream. The products G and H exiting the base of the stripper are sent to a downstream process which is not included in this process. The simulation code allows 15 known preprogrammed major process faults.

The decentralized control strategy outlined by Ricker (1996) [79] was adopted in this work to provide control to the TE process. In Ricker (1996) [79], six variables are selected as the key process variables to be controlled. They are: production rate, mole % G in product, reactor pressure, reactor liquid level, separator liquid level, and stripper level. The simulation codes are available in [2].

The monitoring algorithms discussed in Sections 6.2 and 6.3 are used to construct a diagnostic system. Six univariate control performance monitors are commissioned to monitor the control performance of the six key PVs. In addition to the control performance monitors, three additional model validation monitors are commissioned to monitor the model change of reactor level, separator level and stripper level. According to Downs and Vogel (1993) [18], the reactor cooling water valve and the condenser cooling water valve both have potential sticky valve problem. Two valve stiction monitors are commissioned to monitor the problems. Each valve stiction monitor yields two outputs: the stiction parameter S and J , and thus there are 13 monitor outputs in total. Out of the 13 monitors, three pairs of monitors are located in the same loop: the control performance monitors and model validation monitors of reactor level, separator level, and stripper level. For the three pairs, the bootstrap method is used to model the the joint distributions of monitor outputs. For the rest seven monitors, their distributions are modeled individually using analytical methods (for control performance monitor/model validation monitor) or bootstrap method (for valve stiction monitor).

6.4.1 Diagnostic settings and results

Seven faulty modes are selected for simulation in addition to the normal operation (NF) mode. Thus there are eight modes in total. The eight modes and the corresponding problematic variables are listed in Table 6.3.

Table 6.3: List of simulated modes

Variable number	Process variable	Type
NF	N/A	N/A
IDV 1	A/C feed ratio B composition constant (stream 4)	Step
IDV 2	B composition, A/C ratio constant (stream 4)	Step
IDV 7	C header pressure loss, reduced availability	Step
IDV 8	A, B, C feed composition (stream 4)	Variation
IDV 9	D feed temperature (stream 2)	Variation
IDV 12	Reactor cooling water inlet temperature	Variation
IDV 14	Reactor cooling water valve	Sticking

For each mode, 500 historical data samples, and 500 cross validation data samples are simulated. The diagnosis is performed first using all available historical data. The diagnostic settings are summarized in Table 6.4.

Table 6.4: Summary of Bayesian diagnostic parameters

Discretizaion	$k_i = 2, K = 2^{13} = 8192$
Historical data	500 samples for each mode
Prior samples	Uniformly distributed with prior sample, $a_j = 1, A = 8192$
Prior probabilities	Uniform distributed for all modes
Evaluation data	500 samples for each mode, from training modes

The diagnostic results in terms of average posterior probability are shown in Figure 6.14. The title of each figure indicates the underlying mode from which the validation data are from. The highest posterior probabilities from this simulation are all assigned to the true underlying modes, which are highlighted in gray.

Further assume that there is only one sample of evidence data for IDV 8 fault (stream 4 random variation). The evidence likelihood needs to be estimated from the one evidence sample with the approaches developed in this chapter. The diagnostic

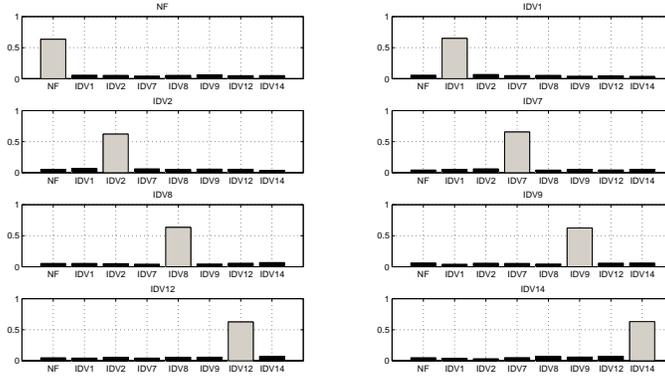


Figure 6.14: TE problem diagnosis with all historical data

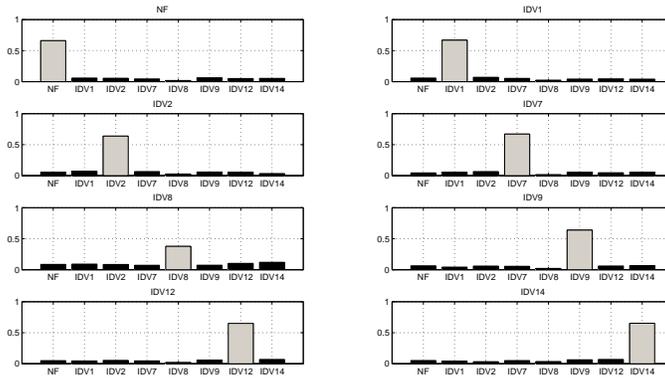


Figure 6.15: TE diagnosis with only one sample from mode IDV 8

results for the eight modes are shown in Figure 6.15. Although the probability assigned to IDV 8 when the underlying fault is IDV 8 is lower than the one calculated in Figure 6.14, the eight modes are correctly diagnosed despite the sparse historical evidence data.

The IDV 8 fault is also selected to examine the likelihood estimation performance. In Figure 6.16, the reconstructed likelihood distribution does not completely agree with the one calculated from the original data. It, however, tracks the trend. Most of the peaks or valleys are captured. Thus the above method provides a good estimation of the likelihood distribution even though the dependency of monitors across different control loops is not considered.

The diagnostic performances are also examined when the historical data is sparse

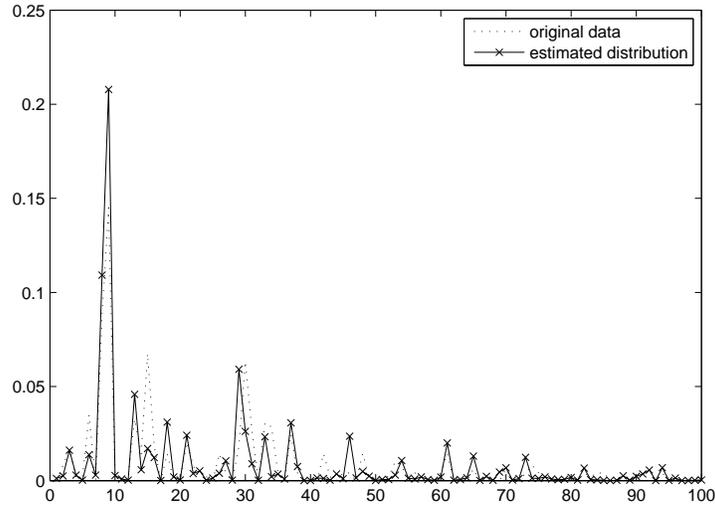


Figure 6.16: Comparison of likelihood distributions

under mode IDV 2, IDV 7, and IDV 14. The results are summarized in Figures 6.17 to 6.19. All the modes are correctly diagnosed even with only one historical evidence sample.

6.4.2 Weighting of historical data

An interesting problem is how to estimate likelihood when there are multiple but relatively sparse historical evidence samples. The key issue of the likelihood estimation problem is how to extract information effectively from a limited number of historical samples. There are two different ways: the first one is to use the average of multiple monitor readings as the nominal monitor output value to calculate the likelihood distribution using analytical approaches or bootstrap; the second one is to estimate the distribution directly from the historical data samples if the number of historical evidences is sufficient. Apparently when there are insufficient historical readings, more weight should be given to the calculated distribution; on the other hand, as the historical data number grows, more weight should be given to the historical data samples; and eventually, we will only consider historical data when a large number of samples are available. Therefore, the weight of the historical data should be a function of the historical sample number. Here we propose using exponential function as the weight function of likelihood estimated from historical

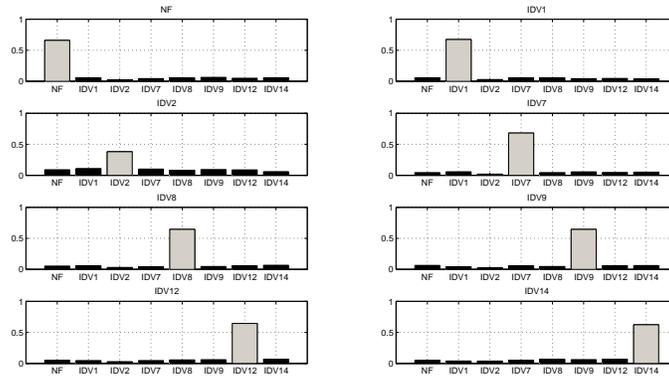


Figure 6.17: TE diagnosis with only one sample from mode IDV 2

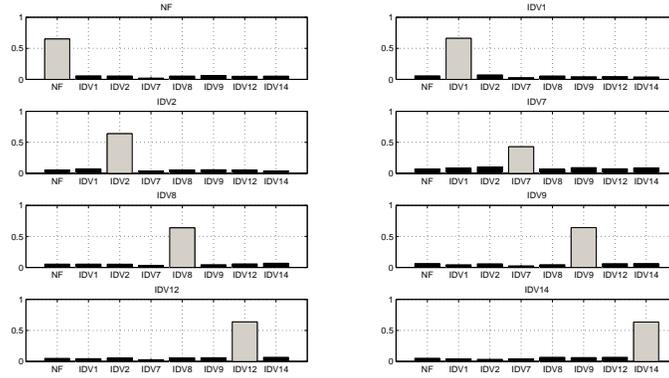


Figure 6.18: TE diagnosis with only one sample from mode IDV 7

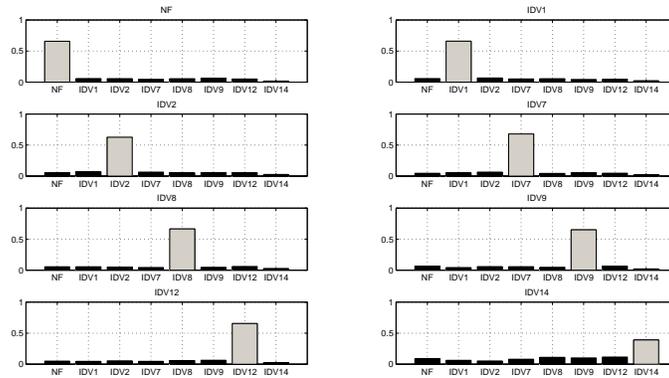


Figure 6.19: TE diagnosis with only one sample from mode IDV 14

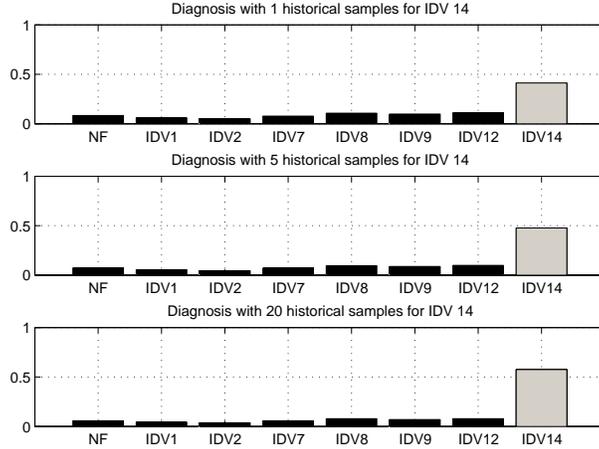


Figure 6.20: Diagnosis with different number of historical samples for IDV 14

data,

$$w = 1 - \exp(-n/\beta), \quad (6.56)$$

where n is the number of historical samples; β is a tuning parameter. The final likelihood is calculated as

$$p = p_{calc} * (1 - w) + p_{hist} * w, \quad (6.57)$$

where p_{calc} is the likelihood calculated with analytical approaches or bootstrap, and p_{hist} is the distribution estimated from historical data directly using the data-driven methods presented in Chapter 2.

We apply the weighting function to the TE problem. As an example, IDV 14 fault is selected. Figure 6.20 summarizes the diagnostic results with different numbers of historical samples for IDV 14 fault. It can be seen that as the data sample number grows, better diagnostic performance is achieved.

6.5 Experimental example

In order to demonstrate practicality of the developed monitor distribution estimation technique, we investigate the proposed approach experimentally on a real distillation column.

6.5.1 Process description

A distillation column with 0.3 m diameter is used to separate a methanol and isopropanol mixture. A schematic diagram of the experimental setup is shown in Figure 6.5.1. The column contains five identical sieve trays spaced 0.457 m apart. Each tray is made of stainless steel and equipped with thermocouple and liquid sampling point at the outlet of the tray. The column is made of Pyrex glass to enable observation of the vapor/liquid phenomena. Detailed dimensions of the column and tray are shown in Table 6.5. The total pressure drop for two trays is measured using a Rosemount differential pressure cell. A total condenser and a thermosiphon partial reboiler complete the distillation system. The column is instrumented for continuous unattended operation. An Opto-22 process I/O subsystem interface with a personal computer running LabView (Version 7.1) software is used for process control and data acquisition.

Table 6.5: Detail dimensions of the column and trays

Column Diameter	0.3m
Tray active area	0.0537 m^2
Hole diameter	4.76 mm
Open hole area	0.00537 m^2
Tray thickness	3.0 mm
Outlet weir height	0.063 m
Inlet weir height	0.051 m
Weir length	0.213 m
Liquid path length	0.202 m
Tray spacing	0.457 m

The column is started with total reflux operation and is then switched to continuous mode by introducing feed to the column and withdrawal of two products from the top and bottom of the column. In this study, a total of five different steady state operating modes are carried out under ambient pressure using methanol/isopropanol mixture [71]. For each operating mode, the column bottom level and the top reflux drum level are kept constant while the other variables, including feed rate, reflux rate, top pressure, and steam rate are varied. Table 6.6 shows the operating variables for the five steady state operating modes. In Table 6.6, all the process data are normalized for easier computation and comparison. When the flow rate and temperature profiles shown by the software (LabView) remain constant for a period

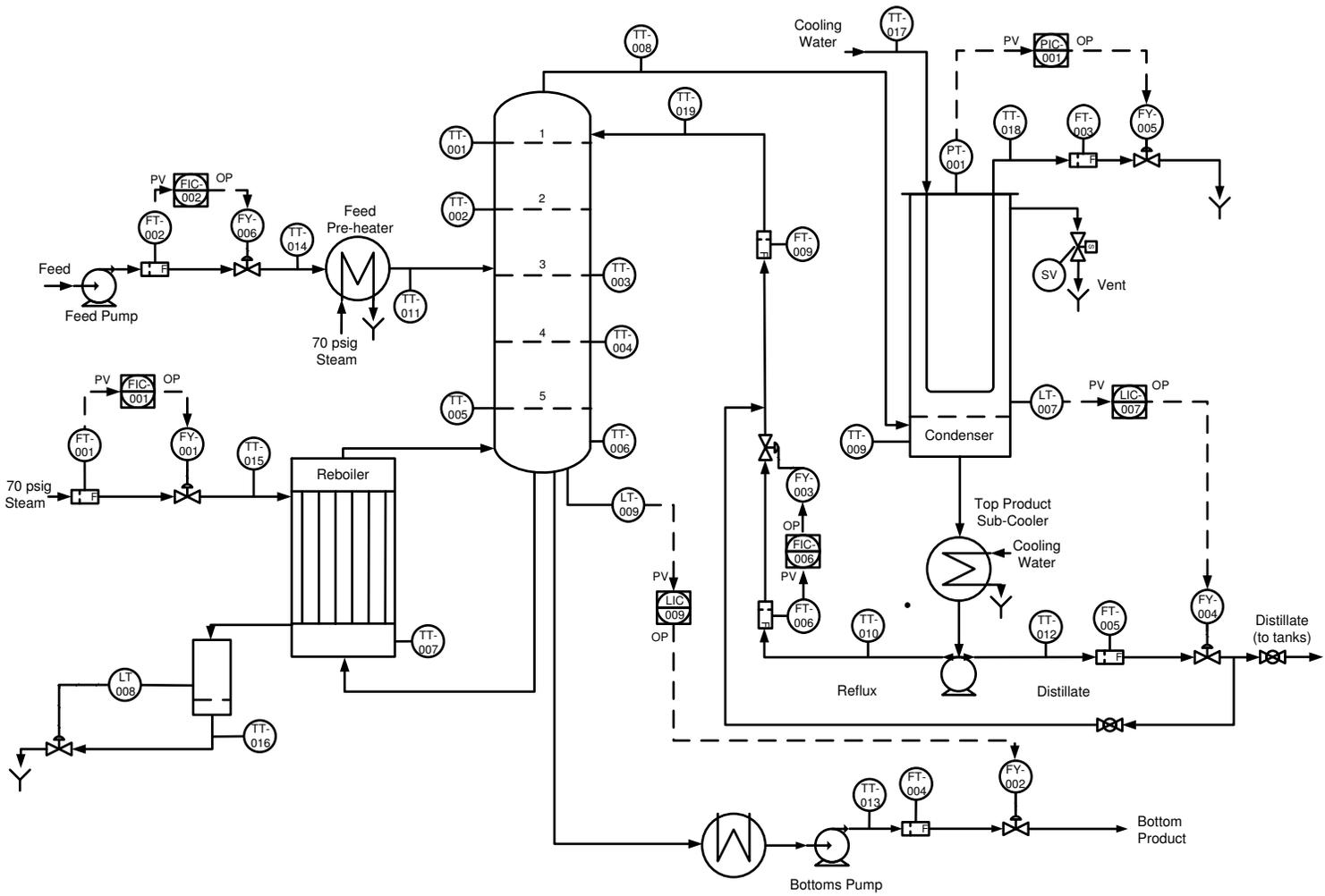


Figure 6.21: Schematic diagram of the distillation column

of 30 minutes, steady state condition is assumed for that particular mode. Liquid samples from each tray outlet and condenser bottom as well as one from the reboiler are taken and analyzed to minimize the measurement uncertainty. During the steady state operation, the sampling period is set to 3 seconds.

Table 6.6: Operating modes for the column

Variables	NF(benchmark)	m_1 (Feed)	m_2 (Reflux)	m_3 (Pressure)	m_4 (Steam)
Feed	0.8	0.95	0.8	0.8	0.8
Reflux	0.15	0.15	0.3	0.15	0.15
Pressure	0.3	0.3	0.3	0.4	0.3
Steam	0.2	0.2	0.2	0.2	0.35
Bottom Level	0.45	0.45	0.45	0.45	0.45
Top Level	0.45	0.45	0.45	0.45	0.45

6.5.2 Diagnostic settings and results

Six monitors are commissioned to detect any changes in the process. As for the model monitor and sensor bias monitor, we use the variable that has the most significant direct impact to the CV as the model input. For example, consider the reflux flow rate as the input for tray 1 (top) temperature. The designed monitors are presented in Table 6.7.

Table 6.7: Commissioned monitors for the column

Monitor	Description
π_1	control performance monitor for tray 5 temperature
π_2	control performance monitor for tray 1 temperature
π_3	control performance monitor for tray 3 temperature
π_4	control performance monitor for cooling water flow rate
π_5	model monitor between steam flow rate and tray 5 temperature
π_6	sensor bias monitor between reflux rate and tray 1 temperature

For each mode, steady state process data of 5 hours are collected. In total there are 6000 process data samples available for each mode. Every 50 process data samples are segmented for a calculation of one evidence/monitor data sample, resulting 120 evidence samples for each mode. Out of the 120 evidence samples, 80 of them are designated as historical samples, and the other 40 are used as cross validation samples. Detailed diagnostic settings are summarized in Table 6.8.

Table 6.8: Summary of Bayesian diagnostic parameters

Discretizaion	$k_i = 2, K = 2^6 = 64$
Historical data	80 samples for each mode
Prior samples	Uniformly distributed with prior sample, $a_j = 1, A = 64$
Prior probabilities	Uniform distributed for all modes
Evaluation data	40 samples for each mode

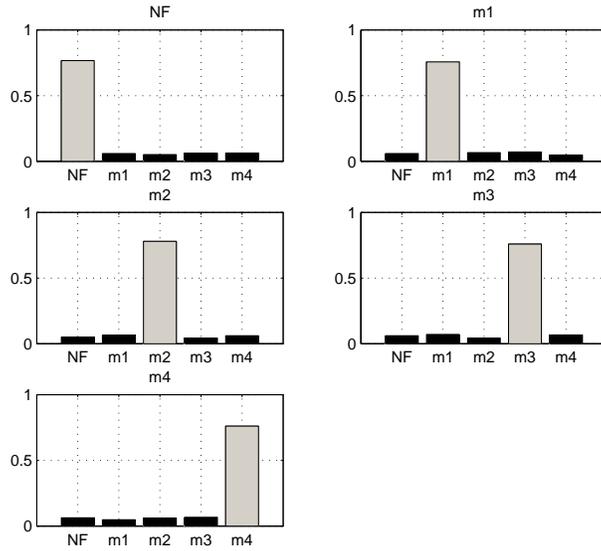


Figure 6.22: Distillation column diagnosis with all historical data

Figure 6.22 shows the diagnostic results in terms of average posterior probability when all historical data samples are available for the five modes. It is observed that all the modes are assigned with the largest posterior probabilities, which are highlighted with gray.

Now assuming that only one historical evidence sample is available for each faulty mode, the proposed estimation techniques for monitor distributions are applied to the six monitors to generate diagnostic results. For the monitors that involve the same process variables, we consider them to be in the same control loop or cross-dependent, i.e., the pair of monitors π_1 and π_5 , and the pair of monitors π_2 and π_6 . The bootstrap approach is employed to estimate the joint distribution of the two pairs. For monitors π_3 and π_4 , which do not have direct correlation with other

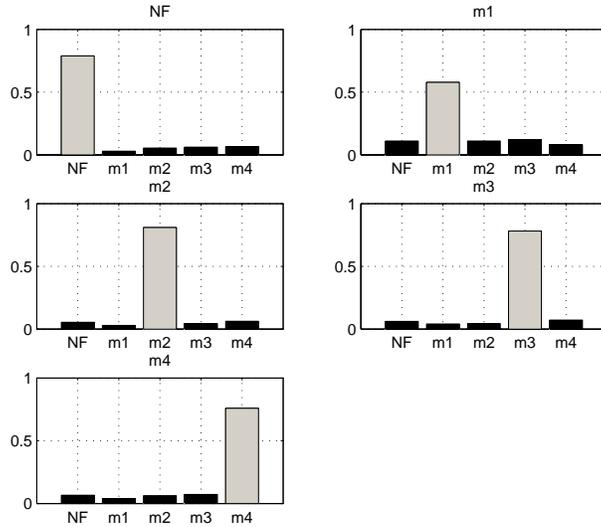


Figure 6.23: Distillation column diagnosis with only one sample from mode m_1

monitors, the analytical approaches are applied.

Figures 6.23 to 6.26 show the diagnostic results when only one historical sample is available for modes m_1 to m_4 . The title of each sub figure indicates the mode from which the validation data are from. It can be seen that even though the average posterior probabilities assigned to the true underlying mode, which are highlighted in gray, are lower than when all historical data are available, the correct diagnosis is still made for all the five modes.

6.6 Conclusions

In this chapter, monitor distribution reconstruction techniques in the presence of sparse historical evidence samples are proposed for control loop diagnosis. The approaches are classified into two categories: the analytical ones and the bootstrap based method. Several monitoring algorithms are selected to illustrate how the distributions can be estimated with the nominal monitor output values from the sparse samples. With the reconstructed monitor distributions, additional data samples are generated using a Monte-Carlo method such that the evidence likelihood can be calculated. The proposed approaches have been applied to the TE problem and a distillation column experiment setup, where the diagnostic performance of the proposed approaches is demonstrated.

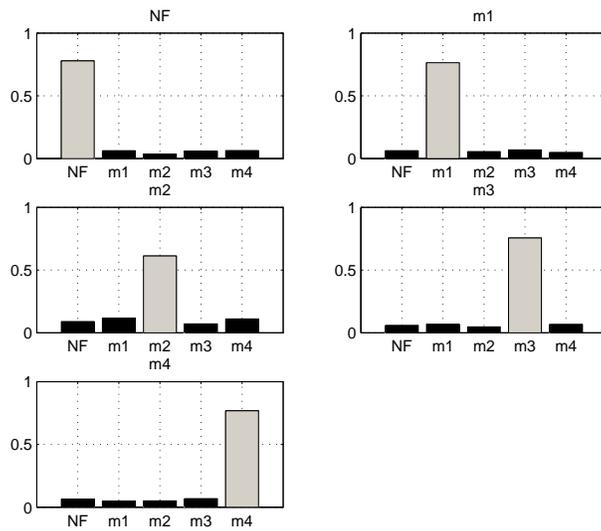


Figure 6.24: Distillation column diagnosis with only one sample from mode m_2

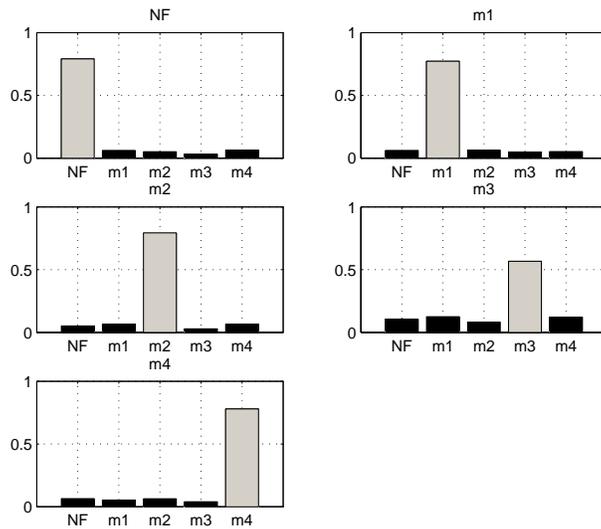


Figure 6.25: Distillation column diagnosis with only one sample from mode m_3

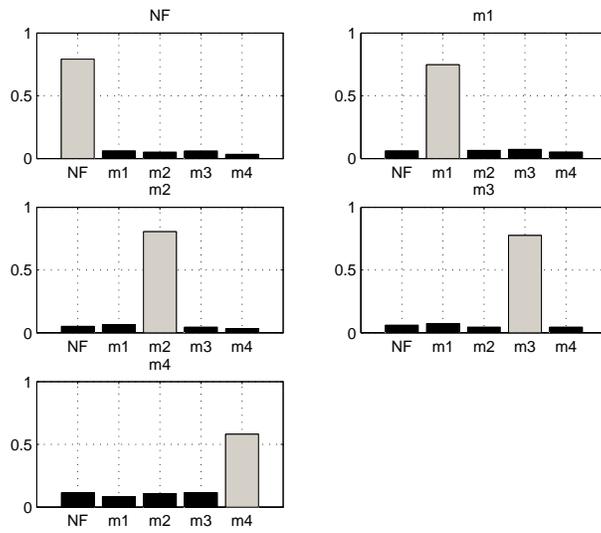


Figure 6.26: Distillation column diagnosis with only one sample from mode m_4

Chapter 7

Concluding Remarks and Future Work

7.1 Concluding remarks

The main objectives of the work reported in this thesis are to develop data-driven Bayesian diagnostic strategies for control loops. Related topics, including incomplete historical evidence, temporal evidence and mode dependency, and sparse historical data, have been addressed. The main contributions of this thesis can be summarized as follows:

- A data-driven Bayesian framework is developed for control loop diagnosis. Information from various monitor readings is synthesized to provide an effective diagnosis of problem source of poor control performance.
- The incomplete historical evidence problem is effectively handled. By introducing the missing pattern concept, the incomplete evidence problems are categorized into single missing pattern problems and multiple missing pattern ones. The likelihood of each incomplete evidence is marginalized as per the underlying complete evidence matrix (UCEM), such that these incomplete evidence samples can also be utilized to facilitate the diagnosis.
- The evidence temporal dependency is taken into consideration to improve Bayesian control loop diagnostic performance. The temporal correlation ratio of monitors is analyzed to alleviate the intensive requirement for historical data when calculating evidence transition probability.
- In addition to the evidence dependency, mode dependency is also considered. A hidden Markov model is built to address the temporal mode dependency

problem in control loop diagnosis. The new solution to mode dependency is then further synthesized with the solution to evidence dependency to develop a recursive auto-regressive hidden Markov model for the online control loop diagnosis.

- An approach to estimate the statistical distributions of monitor readings with sparse historical data is proposed. The proposed approach has the potential to generate evidence likelihood estimation with as few as only one evidence sample. The distribution functions of several monitoring algorithm outputs are analyzed to generate simulated data for likelihood estimation.
- A bootstrap approach is proposed to estimate the statistical distributions of valve stiction parameters. The identification residuals are filtered through a disturbance model to generate i.i.d. residuals for bootstrapping. With the obtained bootstrap residuals, identified closed loop model and the identified disturbance model, the distributions of stiction parameters can be estimated.

The newly proposed techniques have been evaluated on a variety of simulation, experimental and industrial examples.

- The data-driven Bayesian approach (in Chapter 2), incomplete evidence handling strategies (in Chapter 3), and evidence/mode dependency analysis (in Chapters 4 and 5) have been successfully applied to a simulated distillation column model built upon the industrial process presented in Volk et al. (2004) [102].
- The performance of the data-driven Bayesian approach (in Chapter 2), and incomplete evidence handling strategies (in Chapter 3) are evaluated in an industrial oil pre-heater process.
- The evidence/mode dependency analysis (in Chapters 4 and 5) is studied on a pilot tank process in the Computer Process Control Laboratory at the University of Alberta.
- The diagnostic strategy dealing with sparse historical data (Chapter 6) is successfully demonstrated on the Tennessee Eastman Challenge problem and a distillation column set up in the Process Engineering Lab at the University of Alberta.

- Several industrial data sets in Jelali and Huang (2009) [43] and Thornhill et al. (2002) [99] are used as benchmarks to demonstrate the bootstrap approach for valve stiction parameter distribution estimation.

7.2 Recommendations for future work

The following topics/problems are worthy of future investigations:

- A solution to multiple missing pattern problem is provided in this thesis. The likelihood calculation, however, is computationally intensive, and may not be suitable for large scale problems. An approximate solution is desired to simplify the calculation process.
- In Chapter 6, when dealing with the sparse historical data problem, an assumption that monitors across different control loops are independent is made. The overall evidence likelihood is then calculated as the product of likelihood of these independent monitor groups. This assumption may be restrictive in practice, since the material/energy flow can propagate from one loop to another. A cross control loop monitor dependency analysis will further improve the diagnostic performance.
- In this thesis, all discussions are based on discrete monitor output. Actual monitor outputs, however, could be continuous. The diagnosis performance can be affected by the discretization thresholds. A Bayesian diagnosis framework built on continuous monitor outputs will be beneficial to achieve better diagnostic performance.
- Another interesting topic is the economic impact of diagnosis. With several possible problem sources and their probabilities being suggested by the Bayesian approach, a maintenance order will be created to check each one of them. The service of each fault is associated with a price tag, and the removing of each fault will bring different economic benefits. How to design a maintenance sequence, to minimize the maintenance cost and to maximize the economic benefits will be of great interest to practitioners.
- A significant disadvantage of the approaches discussed in this work is the reliance on availability of fault signatures. The evidence likelihood of fault

cannot be estimated if no historical data is available. In order to have a correct diagnosis, a fault must have occurred beforehand, at least once. Developing a method that can identify unknown fault will make the overall Bayesian diagnostic framework more comprehensible.

Bibliography

- [1] The bugs (bayesian inference using gibbs sampling) project website. <http://www.mrc-bsu.cam.ac.uk/bugs/>.
- [2] Tennessee eastman challenge archive. <http://depts.washington.edu/control/LARRY/TE/download.html>.
- [3] S. Ahmed, B. Huang, and S.L. Shah. Validation of continuous-time models with delay. *Chemical Engineering Science*, 64(3):443–454, 2009.
- [4] M. Allen and S. Datta. A note on bootstrapping m-estimators in arma models. *Journal of Time Series Analysis*, 20:365–379, 2002.
- [5] A.S. Badwe, R.D. Gudi, R.S. Patwardhan, S.L. Shah, and S.C. Patwardhan. Detection of model-plant mismatch in mpc applications. *Journal of Process Control*, 19:1305–1313, 2009.
- [6] A.S. Badwe, R.S. Patwardhan, S.L. Shah, S.C. Patwardhan, and R.D. Gudi. Quantifying the impact of model-plant mismatch on controller performance. *Journal of Process Control*, 20:408–425, 2010.
- [7] L. Barford, K. Kanevsky, and L. Kamas. Bayesian fault diagnosis in large-scale measurement systems. In *Proceedings of IMTC 2004 - Instrumentation and Measurement Technology Conference*, 2004.
- [8] J. Carpenter and J. Bithell. Bootstrap confidence intervals: When, which, what? a practical guide for medical statisticians. *Statistics in Medicine*, 19:1141–1164, 2000.
- [9] M.R. Chernick. *Bootstrap methods: a guide for practitioners and researchers*, 2nd ed. Wiley-Interscience, 2008.
- [10] L.H. Chiang, E.L. Russell, and R.D. Braatz. Fault diagnosis in chemical processes using fisher discriminant analysis, discriminant partial least squares, and principal component analysis. *Chemometrics and Intelligent Laboratory Systems*, 50:243–252, 2000.
- [11] Saneer B. Chitralkha, Sirish L. Shah, and J. Prakash. Detection and quantification of valve stiction by the method of unknown input estimation. *Journal of Process Control*, 20:206–216, 2010.
- [12] M. A. A. S. Choudhury, V. Kariwala, N. F. Thornhill, H. Douke, S. L. Shah, H. Takada, and J. F. Forbes. Detection and diagnostics of plant-wide oscillations. *The Canadian Journal of Chemical Engineering*, 85:208–219, 2007.
- [13] M.A.A.S. Choudhury, M. Jain, and S.L. Shah. Stiction - definition, modelling, detection and quantification. *Journal of Process Control*, 18(3-4):232–243, 2008.

- [14] M.A.A.S. Choudhury, N.F. Thornhill, and S.L. Shah. Diagnosis of poor control loop performance using higher order statistics. *Automatica*, 40:1719–1728, 2004.
- [15] D.R. Cox and D.V. Hinkley. *Theoretical statistics*. Chapman & Hall, 1974.
- [16] L.D. Desborough and T.J. Harris. Performance assessment measure for univariate feedback control. *Canadian Journal of Chemical Engineering*, 70:1186–1197, 1992.
- [17] S. Dey and J.A. Stori. A bayesian network approach to root cause diagnosis of process variations. *International Journal of Machine Tools*, 45(1):75–91, 2005.
- [18] J.J. Downs and E.F. Vogel. A plant-wide industrial process control problem. *Computers and Chemical Engineering*, 17:245–255, 1993.
- [19] J.B. Duluc, T. Zimmer, N. Milet, and J.P. Dom. A wafer level reliability method for short-loop processing. *Microelectron reliab.*, 36(11/12):1859–1862, 1996.
- [20] M. Evans, N.A.J. Hastings, and J.B. Peacock. *Statistical distributions*. Wiley: New York, 2000.
- [21] Z. Flores-Loredo, P.H. Ibarguengoytia, and E.F. Morales. On line diagnosis of gas turbines using probabilistic and qualitative reasoning. In *Proceedings of the International Conference on Intelligent Systems Application to Power Systems*, 2005.
- [22] J. Gao, R.S. Patwardhan, K. Akamatsu, Y. Hashimoto, G. Emoto, and S. L. Shah. Performance evaluation of two industrial mpc controllers. *Control Engineering Practice*, 11:1371–1387, 2003.
- [23] S. Garatti and R.R. Bitmead. On resampling and uncertainty estimation in linear system identification. In *Proceeding of 15th IFAC symposium on system identification*, Saint-Malo, France, July 2009.
- [24] S. Geman and D. Geman. Stochastic relaxation, gibbs distributions, and the bayesian restoration of images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 6:721–741, 1984.
- [25] J. P. Gerry. Prioritizing and optimizing problem loops using a loop monitoring. In *Proceedings of the ISA*, 2002.
- [26] Z. Ghahramani. An introduction to hidden markov models and bayesian networks. *International Journal of Pattern Recognition and Artificial Intelligence*, 15(1):9–42, 2002.
- [27] E. Gilabert and A. Arnaiz. Intelligent automation systems for predictive maintenance: A case study. *Robotics and Computer-Integrated Manufacturing*, 22:543–549, 2006.
- [28] T. Harris, F. Boudreau, and J. F. MacGregor. Performance assessment using of multivariable feedback controllers. *Automatica*, 32:1505–1518, 1996.
- [29] T.J. Harris, C.T. Seppala, and L.D. Desborough. A review of performance monitoring and assessment techniques for univariate and multivariate control systems. *Journal of Process Control*, 9(1):1–17, 1999.
- [30] Q.P. He, J. Wang, M. Pottmann, and S.J. Qin. A curve fitting method for detecting valve stiction in oscillation control loops. *Industrial Engineering Chemistry Research*, 46:4549–4560, 2007.

- [31] K.A. Hoo, M.J. Piovoso, P.D. Sheneller, and D.A. Rowan. Process and controller performance monitoring: overview with industrial applications. *International Journal of Adaptive Control and Signal Processing*, 17:635–662, 2003.
- [32] A. Horch. A simple method for detection of stiction in control valves. *Control Engineering Practice*, 7:1221–1231, 1999.
- [33] W.A. Houlberg, K.C. Shaing, S.P. Hirshman, and M.C. Zarnstorff. Bootstrap current and neoclassical transport in tokamaks of arbitrary collisionality and aspect ratio. *Phys. Plasmas*, 4:3230, 1997.
- [34] B. Huang. On-line closed-loop model validation and detection of abrupt parameter changes. *Journal of Process Control*, 11:699–715, 2001.
- [35] B. Huang. Bayesian methods for control loop monitoring and diagnosis. *Journal of Process Control*, 18:829–838, 2008.
- [36] B. Huang and S.L. Shah. *Performance assessment of control loops*. Springer-Verlag, 1999.
- [37] B. Huang, S.L. Shah, and E.Z. Kwok. Good, bad or optimal? performance assessment of multivariable process. *Automatica*, 33:1175–1183, 1997.
- [38] B. Huang, S.L. Shah, and K.Z. Kwok. On-line control performance monitoring of mimo processes. In *Proceedings of American Control Conference*, 1995.
- [39] A.J. Hugo. Performance assessment of single-loop industrial controllers. *Journal of Process Control*, 16:784–794, 2006.
- [40] S.A. Imtiaz and S.L. Shah. Treatment of missing values in process data analysis. *Canadian Journal of Chemical Engineering*, 86:838–858, 2008.
- [41] M. Jelali. An overview of control performance assessment technology and industrial applications. *Control Engineering Practice*, 14(5):441–466, 2006.
- [42] M. Jelali. Estimation of valve stiction in control loops using separable least-squares and global search algorithms. *Journal of Process Control*, 18:632–642, 2008.
- [43] M. Jelali and B. Huang, editors. *Detection and Diagnosis of Stiction in Control Loops: State of the Art and Advanced Methods*. Springer London, 2009.
- [44] H. Jiang, B. Huang, and S.L. Shah. Closed-loop model validation based on the two-model divergence method. *Journal of Process Control*, 19:644–655, 2009.
- [45] M. Kano, H. Maruta, Kugemoto H, and K. Shimizu. Practical model and detection algorithm for valve stiction. In *Proceeding of IFAC Symposium on Dynamics and Control of Process Systems*, Boston, USA, July 2004.
- [46] Y. Kawahara, T. Yairi, and K. Machida. Fault diagnosis for spacecraft using probabilistic reasoning and statistical learning with dynamic bayesian networks. In *Proceedings of International Astronautical Federation - 56th International Astronautical Congress*, 2005.
- [47] S.J. Kendra and A. Cinar. Controller performance assessment by frequency domain techniques. *Journal of Process Control*, 7:181–194, 1997.
- [48] B. Ko and T.F. Edgar. Assessment of achievable pi control performance for linear processes with dead time. In *Proceedings of American Control Conference*, 1998.

- [49] B.S. Ko and T. F. Edgar. Performance assessment of cascade control loops. *AIChE Journal*, 46:281–291, 2000.
- [50] K.B. Korb and A.E. Nicholson. *Bayesian artificial intelligence*. Chapman & Hall/CRC, 2004.
- [51] J.-P. Kreiss and J. Franke. Bootstrapping stationary autoregressive moving-average models. *Journal of Time Series Analysis*, 13:297–316, 1992.
- [52] W. Ku, R.H. Storer, and C. Georgakis. Disturbance detection and isolation by dynamic principal component analysis. *Chemometrics and Intelligent Laboratory Systems*, 30:179–196, 1995.
- [53] S. Kullback. *Information Theory and Statistics*. John Wiley & Sons New York, 1959.
- [54] J.-M Lee and I.-B. Lee S.J. Qin. Fault detection and diagnosis based on modified independent component analysis. *AIChE Journal*, 52:3501–3514, 2006.
- [55] K.H. Lee, Z. Ren, and B. Huang. Stiction estimation using constrained optimisation and contour map. In *Detection and Diagnosis of Stiction in Control Loops: State of the Art and Advanced Methods*. Springer London, 2010.
- [56] R. W. Lewis and R.S. Ransing. A semantically constrained bayesian network for manufacturing diagnosis. *International Journal of Production Research*, 35:2171–2187, 1997.
- [57] Q. Li, J. R. Whiteley, and R. R. Rhinehart. A relative performance monitor for process controllers. *International Journal of Adaptive Control and Signal Processing*, 17:685–708, 2003.
- [58] R. J. A. Little. Modeling the dropout mechanism in repeated-measures studies. *Journal of the American Statistical Association*, 90:1112–1121, 1995.
- [59] R. J. A. Little and D. B. Rubin. *Statistical analysis with missing data*. New York: Wiley, 1987.
- [60] L. Ljung. *System Identification: Theory for the User (2nd Edition)*. Prentice Hall, 1999.
- [61] S. Mahadevan and S. L. Shah. Fault detection and diagnosis in process data using one-class support vector machines. *Journal of Process Control*, 19:1627–1639, 2009.
- [62] K.V. Mardia, J.T. Kent, and J.M. Bibby. *Multivariate Analysis*. Academic Press, 1979.
- [63] L. Matthaus, P. Trilleberg, T. Fadini, M. Finke, and A. Schweikard. Brain mapping with transcranial magnetic stimulation using a refined correlation ratio and kendall’s τ . *Statistics in Medicine*, 27:5252–5270, 2008.
- [64] T.J. McAvoy and N. Ye. Base control for the tennessee eastman problem. *Computers and Chemical Engineering*, 18:383–413, 1994.
- [65] C. A. McNabb and S. J. Qin. Projection based mimo control performance monitoring: Ii-measured disturbances and setpoint changes. *Journal of Process Control*, 15:89–102, 2005.
- [66] M. Mehranbod and M. Soroush. Probabilistic model for sensor fault detection and identification. *AIChE Journal*, 49:1787–1802, 2003.

- [67] N. Mehranbod, M. Soroush, and C. Panjapornpon. A method of sensor fault detection and identification. *Journal of Process Control*, 15(3):321–339, 2005.
- [68] K.P. Murphy. *Dynamic Bayesian networks: representation, inference, and learning*. PhD thesis, University of California, Berkeley, 2002.
- [69] U. Nallasivama, S. Babjib, and R. Rengaswamy. Stiction identification in nonlinear process control loops. *Computers and Chemical Engineering*, 2010. in press.
- [70] B. Ninness and S. Gibson. Quantifying the accuracy of hammerstein model estimation. *Automatica*, 38:2037–2051, 2002.
- [71] M.J. Olanrewaju, B. Huang, and A. Afacan. Online composition estimation and experiment validation of distillation processes with switching dynamics. *Chemical Engineering Science*, 66:1597–1608, 2010.
- [72] R.S. Patwardhan and S.L. Shah. Issues in performance diagnostics of model-based controllers. *Journal of Process Control*, 12(3):413–427, 2002.
- [73] A. Pernestål. *A Bayesian approach to fault isolation with application to diesel engine diagnose*. PhD thesis, KTH School of Electrical Engineering, 2007.
- [74] Anna Pernestål and Mattias Nyberg. Using prior information in bayesian inference - with application to diagnosis. MaxEnt, Bayesian Inference and Maximum Entropy Methods in Science and Engineering, Albany, NY, USA, 2007.
- [75] F. Qi, B. Huang, and E.C. Tamayo. A bayesian approach for control loop diagnosis with missing data. *AIChE Journal*, 56:179–195, 2010.
- [76] S.J. Qin. Control performance monitoring - a review and assessment. *Computers and Chemical Engineering*, 23(2):173–186, 1998.
- [77] S.J. Qin and W. Li. Detection and identification of faulty sensors in dynamic process. *AIChE Journal*, 47(7):1581–1593, 2001.
- [78] A. Raich and A. Inar. Statistical process monitoring and disturbance diagnosis in multivariable continuous processes. *AIChE Journal*, 42:995–1009, 1996.
- [79] N.L. Ricker. Decentralized control of the tennessee eastman challenge process. *Journal of Process Control*, 6:205–221, 1996.
- [80] C. Romessis and K. Mathioudakis. Bayesian network approach for gas path fault diagnosis. *Journal of Engineering for Gas Turbines and Power*, 64-72:64–72, 2006.
- [81] D. B. Rubin. *Multiple imputation for noresponse in surveys*. New York: Wiley, 1987.
- [82] F. Sahin, M.C. Yavuz, Z. Arnavut, and O. Uluyol. Fault diagnosis for airplane engines using bayesian networks and distributed particle swarm optimization. *Parallel Computing*, 33:124–143, 2007.
- [83] Claudio Scali and Claudio Ghelardoni. An improved qualitative shape analysis technique for automatic detection of valve stiction in flow control loops. *Control Engineering Practice*, 16:1501–1508, 2008.
- [84] J. Schafer and A. Cinar. Multivariable mpc system performance assessment, monitoring, and diagnosis. *Journal of Process Control*, 14(2):113–129, 2006.
- [85] J. L. Schafer. *Analysis of incomplete multivariate data*. London: Chapman & Hall, 1997.

- [86] S. Selvanathan and A. K. Tangirala. Diagnosis of poor control loop performance due to model-plant mismatch. *Industrial Engineering Chemistry Research*, 49:4210–4229, 2010.
- [87] C. T. Seppala, T. J. Harris, and D. W. Bacon. Time series methods for dynamic analysis of multiple controlled variables. *Journal of Process Control*, 12:257–276, 2002.
- [88] T. Seppala, H. Moskowitz, R. Plante, and J. Tang. Statistical process control via the subgroup bootstrap. *Journal of Quality Technology*, 27:139–153, 1995.
- [89] S. L. Shah, R. Patwardhan, and B. Huang. Multivariate controller performance analysis: Methods, applications and challenges. In *Proceedings of the Chemical Process Control Conference*, 2001.
- [90] Y. A.W. Shardt, Y. Zhao, F. Qi, K. H. Lee, X. Yu, B. Huang, and S. L. Shah. A review of performance assessment to determine the state of a process control system. *Canadian Journal of Chemical Engineering*, in press, 2011.
- [91] A. Singhal and T.I. Salsbury. A simple method for detecting valve stiction in oscillating control loops. *Journal of Process Control*, 15:371–382, 2005.
- [92] S. Sinharay, H. S. Stern, and D. Russell. The use of multiple imputation for the analysis of missing data. *Psychological Methods*, 6:317–329, 2001.
- [93] R. Srinivasan, R. Rengaswamy, S. Narasimhan, and R. Miller. A curve fitting method for detecting valve stiction in oscillation control loops. *Industrial Engineering Chemistry Research*, 44:6719–6728, 2005.
- [94] N. Stanfelj, Marlin T., and MaGregor J.F. Monitoring and diagnosing process control performance: the single-loop case. *Industrial Engineering Chemistry Research*, 32:301–314, 1993.
- [95] M. Steinder and A.S. Sethi. Probabilistic fault localization in communication systems using belief networks. *IEEE/ACM Transactions on Networking*, 12(5):809–822, 2004.
- [96] C.V. Stewart, C.-L. Tsai, and B. Roysam. The dual-bootstrap iterative closest point algorithm with application to retinal image registration. *IEEE Transactions on Medical Imaging*, 22:1379–1394, 2003.
- [97] N. F. Thornhill, B. Huang, and S. L. Shah. Controller performance assessment in set point tracking and regulatory control. *International Journal of Adaptive Control and Signal Processing*, 17:709–727, 2003.
- [98] N.F. Thornhill and A. Horch. Advances and new directions in plant-wide controller performance assessment. In *Proceeding of International Symposium on Advanced Control of Chemical Processes*, Gramado, Brazil, April 2006.
- [99] N.F. Thornhill, S.L. Shah, B. Huang, and A. Vishnubhotla. Spectral principal component analysis of dynamic process data. *Control Engineering Practice*, 10:833–846, 2002.
- [100] G. Venkataramanan, V. Shukla, R. Saini, and R. R. Rhinehart. An automated on-line monitor of control system performance. In *Proceedings of the American Control Conference*, 1997.
- [101] G. Verbeke and G. Molenberghs. *Linear mixed models for longitudinal data*. New York: Springer-Verlag, 2000.
- [102] U. Volk, D.W. Kniese, R. Hahn, R. Haber, and U. Schmitz. Optimized multivariable predictive control of an industrial distillation column considering hard and soft constraints. *Control Engineering Practice*, 13(7):809–822, 2004.

- [103] E. Wolbrechet, B. D'Ambrosio, R. Paasch, and D. Kirby. Monitoring and diagnosis of a multistage manufacturing process using bayesian networks. *Artificial Intelligence for Engineering Design, Analysis and Manufacturing*, 14:53–67, 2000.
- [104] Yoshiyuki Yamashita. An automatic method for detection of valve stiction in process control loops. *Control Engineering Practice*, 14:503–510, 2006.
- [105] J. Yu and S. J. Qin. Statistical mimo controller performance monitoring. part ii: Performance diagnosis. *Journal of Process Control*, 18:297–319, 2008.
- [106] S. L. Zeger, K. Y. Liang, and P. S. Albert. Models for longitudinal data: A generalized estimating equation approach. *Biometrics*, 44:1049–1060, 1988.
- [107] A.M. Zoubir and B. Boashash. The bootstrap and its application in signal processing: An attractive tool for assessing the accuracy of estimators and testing hypothesis for parameters in small data-sample situations. *IEEE Signal Processing Magazine*, 15:56–76, 1998.
- [108] A.M. Zoubir and D.R. Iskander. Bootstrap methods and applications. *IEEE Singal Processing Magazine*, pages 10–19, 2007.

Appendix A

Estimation of Likelihood with Incomplete Evidences

A.1 Evidence likelihood estimation

Suppose that the likelihood of evidence $\epsilon_{s,r}$ is about to be calculated. According to Equation 2.5, the likelihood probability can be computed by marginalization over all possible likelihood parameter sets Θ ,

$$\begin{aligned}
& p(\epsilon_{s,r}|M, \mathcal{D}) \\
&= \int_{\Omega} p(\epsilon_{s,r}|\Theta, M, \mathcal{D}) f(\Theta|M, \mathcal{D}) d\Theta \\
&= \int_{\Omega} \theta_{s,r} \frac{c}{p(\mathcal{D}|M)} \prod_{i=1}^S \left[\prod_{j=1}^R \theta_{i,j}^{\eta_{i,j}+a_{i,j}-1} \cdot \left(\sum_{k=1}^R \theta_{i,k} \right)^{\eta_i} \right] d\Theta \\
&= \frac{c}{p(\mathcal{D}|M)} \int_{\Omega} \theta_{s,r} \prod_{i=1}^S \left[\prod_{j=1}^R \theta_{i,j}^{\eta_{i,j}+a_{i,j}-1} \cdot \left(\sum_{k=1}^R \theta_{i,k} \right)^{\eta_i} \right] d\Theta \\
&= \frac{c}{\int_{\Omega} p(\mathcal{D}|\theta, M) f(\theta|M)} \cdot \int_{\Omega} \theta_{s,r} \prod_{i=1}^S \left[\prod_{j=1}^R \theta_{i,j}^{\eta_{i,j}+a_{i,j}-1} \cdot \left(\sum_{k=1}^R \theta_{i,k} \right)^{\eta_i} \right] d\Theta \\
&= \left(\int_{\Omega} \theta_{s,r} \prod_{i=1}^S \left[\prod_{j=1}^R \theta_{i,j}^{\eta_{i,j}+a_{i,j}-1} \cdot \left(\sum_{k=1}^R \theta_{i,k} \right)^{\eta_i} \right] d\Theta \right) \\
&\quad / \left(\int_{\Omega} \prod_{i=1}^S \left[\prod_{j=1}^R \theta_{i,j}^{\eta_{i,j}+a_{i,j}-1} \cdot \left(\sum_{k=1}^R \theta_{i,k} \right)^{\eta_i} \right] d\Theta \right). \tag{A.1}
\end{aligned}$$

The denominator of Equation A.1 is

$$\begin{aligned}
& \int_{\Omega} \prod_{i=1}^S \left[\prod_{j=1}^R \theta_{i,j}^{\eta_{i,j} + a_{i,j} - 1} \cdot \left(\sum_{k=1}^R \theta_{i,k} \right)^{\eta_i} \right] d\Theta \\
&= \int_{\Omega} \prod_{i=1}^S \prod_{j=1}^R \theta_{i,j}^{\eta_{i,j} + a_{i,j} - 1} \cdot \sum_{t_{i,1}, \dots, t_{i,R}} \binom{\eta_i}{t_{i,1}, \dots, t_{i,R}} \theta_{i,1}^{t_{i,1}} \dots \theta_{i,R}^{t_{i,R}} d\Theta \\
&= \int_{\Omega} \prod_{i=1}^S \sum_{t_{i,1}, \dots, t_{i,R}} \binom{\eta_i}{t_{i,1}, \dots, t_{i,R}} \cdot \theta_{i,1}^{t_{i,1} + \eta_{i,1} + a_{i,1} - 1} \dots \theta_{i,R}^{t_{i,R} + \eta_{i,R} + a_{i,R} - 1} d\Theta, \quad (\text{A.2})
\end{aligned}$$

where

$$\begin{aligned}
\binom{\eta_i}{t_{i,1}, \dots, t_{i,R}} &= \binom{\eta_i}{t_{i,1}} \binom{\eta_i - t_{i,1}}{t_{i,2}} \dots \binom{\eta_i - t_{i,1} - t_{i,2} - \dots - t_{i,R}}{t_{i,R}} \\
&= \frac{\eta_i!}{t_{i,1}! t_{i,2}! \dots t_{i,R}!}, \quad (\text{A.3})
\end{aligned}$$

and $\sum_{t_{i,1}, \dots, t_{i,R}}$ is summation over all possible sequences of nonnegative integer indices $t_{i,1}$ through $t_{i,R}$ such that $\sum_{j=1}^R t_{i,j} = \eta_i$.

Similarly, the numerator in Equation A.1 is

$$\int_{\Omega} \theta_{s,r} \prod_{i=1}^S \sum_{t_{i,1}, \dots, t_{i,R}} \binom{\eta_i}{t_{i,1}, \dots, t_{i,R}} \cdot \theta_{i,1}^{t_{i,1} + \eta_{i,1} + a_{i,1} - 1} \dots \theta_{i,R}^{t_{i,R} + \eta_{i,R} + a_{i,R} - 1} d\Theta \quad (\text{A.4})$$

Denote all sets of $t_{i,1}, \dots, t_{i,R}$ as \mathcal{T}_i , combination of all \mathcal{T}_i as \mathcal{T} , $\binom{\eta_i}{t_{i,1}, \dots, t_{i,R}}$ as $\mathbf{C}_{\eta_i}^{\mathcal{T}_i}$, and $\eta_{i,j} + a_{i,j} - 1$ as $q_{i,j}$. Then Equation A.2 is

$$\begin{aligned}
& \int_{\Omega} \prod_{i=1}^S \left[\sum_{\mathcal{T}_i} \mathbf{C}_{\eta_i}^{\mathcal{T}_i} \theta_{i,1}^{t_{i,1} + q_{i,1}} \dots \theta_{i,R}^{t_{i,R} + q_{i,R}} \right] d\Theta \\
&= \sum_{\mathcal{T}} \prod_{i=1}^S \mathbf{C}_{\eta_i}^{\mathcal{T}_i} \cdot \int_{\Omega} \prod_{i=1}^S \left[\theta_{i,1}^{t_{i,1} + q_{i,1}} \dots \theta_{i,R}^{t_{i,R} + q_{i,R}} \right] d\Theta \\
&= \sum_{\mathcal{T}} \frac{\prod_{i=1}^S \mathbf{C}_{\eta_i}^{\mathcal{T}_i} \prod_{j=1}^R \Gamma(t_{i,j} + q_{i,j} + 1)}{\Gamma(N + A + 1)}, \quad (\text{A.5})
\end{aligned}$$

where $N = \sum_i \eta_i + \sum_i \sum_j \eta_{i,j}$ is the total number of historical data samples for mode M , including both complete and incomplete samples; $A = \sum_i \sum_j a_{i,j}$ is the total number of prior samples, which is, however, only applicable to the complete evidences.

With the new notations, Equation A.4 can be rewritten as

$$\begin{aligned} & \int_{\Omega} \theta_{s,r} \prod_{i=1}^S \left[\sum_{\mathcal{T}_i} \mathbf{C}_{\eta_i}^{\mathcal{T}_i} \theta_{i,1}^{t_{i,1}+q_{i,1}} \dots \theta_{i,R}^{t_{i,R}+q_{i,R}} \right] d\Theta \\ &= \sum_{\mathcal{T}} \frac{\prod_{i=1}^S \mathbf{C}_{\eta_i}^{\mathcal{T}_i} \Gamma(t_{s,r} + q_{s,r} + 2) \prod_{j \neq r} \Gamma(t_{i,j} + q_{i,j} + 1)}{\Gamma(N + A + 1)} \end{aligned} \quad (\text{A.6})$$

The likelihood of evidence $\epsilon_{s,r}$ can be calculated as

$$\begin{aligned} & p(\epsilon_{s,r} | M, \mathcal{D}) \\ &= \frac{1}{N + A} \cdot \frac{\sum_{\mathcal{T}} \prod_{i=1}^S \mathbf{C}_{\eta_i}^{\mathcal{T}_i} (t_{s,r} + q_{s,r} + 1)! \prod_{j \neq r} (t_{i,j} + q_{i,j})!}{\sum_{\mathcal{T}} \prod_{i=1}^S \mathbf{C}_{\eta_i}^{\mathcal{T}_i} \prod_{j=1}^R (t_{i,j} + q_{i,j})!}. \end{aligned} \quad (\text{A.7})$$

As discussed previously, there is no replication between different rows in a UCEM. Any possible underlying complete evidences of an incomplete one can only be located uniquely in the corresponding row. Thus the assignments of $t_{i,1}, \dots, t_{i,R}$ are independent for different rows i , and Equation A.7 can be simplified to

$$\begin{aligned} & p(\epsilon_{s,r} | M, \mathcal{D}) \\ &= \frac{1}{N + A} \cdot \frac{\sum_{\mathcal{T}_s} \mathbf{C}_{\eta_s}^{\mathcal{T}_s} (t_{s,r} + q_{s,r} + 1)! \prod_{j \neq r} (t_{s,j} + q_{s,j})!}{\sum_{\mathcal{T}_s} \mathbf{C}_{\eta_s}^{\mathcal{T}_s} \prod_{j=1}^R (t_{s,j} + q_{s,j})!} \\ & \quad \cdot \frac{\sum_{\mathcal{T} \setminus \mathcal{T}_s} \prod_{i \neq s} \mathbf{C}_{\eta_i}^{\mathcal{T}_i} \prod_{j=1}^R (t_{s,j} + q_{s,j})!}{\sum_{\mathcal{T} \setminus \mathcal{T}_s} \prod_{i \neq s} \mathbf{C}_{\eta_i}^{\mathcal{T}_i} \prod_{j=1}^R (t_{s,j} + q_{s,j})!} \\ &= \frac{1}{N + A} \cdot \frac{\sum_{\mathcal{T}_s} \mathbf{C}_{\eta_s}^{\mathcal{T}_s} (t_{s,r} + q_{s,r} + 1)! \prod_{j \neq r} (t_{s,j} + q_{s,j})!}{\sum_{\mathcal{T}_s} \mathbf{C}_{\eta_s}^{\mathcal{T}_s} \prod_{j=1}^R (t_{s,j} + q_{s,j})!}, \end{aligned} \quad (\text{A.8})$$

where $\mathcal{T} \setminus \mathcal{T}_s$ is the combination of all the possible set of \mathcal{T}_i with $i \neq s$.

As a result, the likelihood equation is greatly simplified by completely removing other possible \mathcal{T}_i with $i \neq s$. Only the assignment of $t_{s,1}, \dots, t_{s,R}$ to evidences $\epsilon_{s,1}, \dots, \epsilon_{s,R}$ needs to be considered. However, the computation load is still heavy, and will grow exponentially with the increase of R and η_s . Further simplification is needed to make the likelihood computation feasible. This is developed below.

Lemma A.1.

$$\sum_{\mathcal{T}_s} \binom{\eta_s}{t_1, \dots, t_R} \prod_{j=1}^R (x_j + t_j)! = \prod_{k=1}^R x_k! \cdot \prod_{i=0}^{\eta_s-1} \left(\sum_{j=1}^R x_j + i + R \right) \quad (\text{A.9})$$

See appendix A.2 for details of the proof.

With Lemma A.1, Equation A.8 can be further simplified. The denominator is

$$\begin{aligned} & (N + A) \sum_{\mathcal{I}_s} \binom{\eta_s}{t_1, \dots, t_R} \prod_{j=1}^R (t_j + \eta_{s,j} + a_{s,j} - 1)! \\ &= (N + A) \prod_{k=1}^R (\eta_{s,k} + a_{s,k} - 1)! \cdot \prod_{i=0}^{\eta_s-1} \left(\sum_{j=1}^R [\eta_{s,j} + a_{s,j}] + i \right), \end{aligned} \quad (\text{A.10})$$

and the numerator is

$$\begin{aligned} & \sum_{\mathcal{I}_s} \binom{\eta_s}{t_1, \dots, t_R} (t_r + \eta_{s,r} + a_{s,r})! \cdot \prod_{j \neq r} (t_j + \eta_{s,j} + a_{s,j} - 1)! \\ &= (\eta_{s,r} + a_{s,r})! \prod_{k \neq r} (\eta_{s,k} + a_{s,k} - 1)! \cdot \prod_{i=0}^{\eta_s-1} \left(\sum_{j=1}^R [\eta_{s,j} + a_{s,j}] + i + 1 \right) \end{aligned} \quad (\text{A.11})$$

By substituting Equation A.10 and Equation A.11 in Equation A.8, the following result is obtained for the likelihood of evidence $\epsilon_{s,r}$,

$$\begin{aligned} p(\epsilon_{s,r} | M, \mathcal{D}) &= \frac{\eta_{s,r} + a_{s,r}}{N + A} \cdot \frac{\sum_{j=1}^R (\eta_{s,j} + a_{s,j}) + \eta_s}{\sum_{j=1}^R (\eta_{s,j} + a_{s,j})} \\ &= \frac{\eta_{s,r} + a_{s,r}}{N + A} \cdot \left(1 + \frac{\eta_s}{\sum_{j=1}^R (\eta_{s,j} + a_{s,j})} \right) \end{aligned} \quad (\text{A.12})$$

A.2 Proof of Lemma A.1

Lemma A.1

$$\sum_{\mathcal{I}_s} \binom{\eta_s}{t_1, \dots, t_R} \prod_{j=1}^R (x_j + t_j)! = \prod_{k=1}^R x_k! \prod_{i=0}^{\eta_s-1} \left(\sum_{j=1}^R x_j + i + R \right) \quad (\text{A.13})$$

Use mathematical induction to prove the above equation.

First, let $n_s = 1$, $R = 2$, which is the most simplest data missing case, i.e., one incomplete data sample with two possible evidence values.

$$RHS = \prod_{j=1}^2 x_j! \prod_{i=0}^0 \left(\sum_{j=1}^2 x_j + i + 2 \right) = x_1! x_2! (x_1 + x_2 + 2) \quad (\text{A.14})$$

$$\begin{aligned} LHS &= \sum_{\mathcal{I}} \binom{1}{t_1, t_2} \prod_{j=1}^2 (x_j + t_j)! = (x_1 + 1)! x_2! + x_1! (x_2 + 1)! \\ &= x_1! x_2! (x_1 + x_2 + 2) = RHS \end{aligned} \quad (\text{A.15})$$

There are two parameters which need to be inducted, n_s and R . Suppose the equation holds for n_s and R .

1. Induction on R

Note that Equation A.13 can be written as

$$\sum_{\mathcal{I}_s} \binom{\eta_s}{t_1, \dots, t_R} \prod_{j=1}^R \prod_{r=1}^{t_j} (x_j + r) = \prod_{i=0}^{n_s-1} \binom{R}{\sum_{j=1}^R x_j + i + R}. \quad (\text{A.16})$$

Thus

$$\begin{aligned} & \prod_{i=0}^{n_s-1} \binom{R+1}{\sum_{j=1}^{R+1} x_j + i + R + 1} \\ &= \prod_{i=0}^{n_s-1} \binom{R-1}{\sum_{j=1}^{R-1} (x_j + 1) + [(x_R + x_{R+1} + 1) + 1] + i} \\ &= \sum_{\mathcal{I}_s} \binom{\eta_s}{t_1, \dots, t_{R-1}, T} \prod_{j=1}^{R-1} \prod_{r=1}^{t_j} (x_j + r) \\ & \quad \cdot (x_R + x_{R+1} + 1 + 1) \cdots (x_R + x_{R+1} + 1 + T) \\ &= \sum_{\mathcal{I}_s} \binom{\eta_s}{t_1, \dots, t_{R-1}, T} \prod_{j=1}^{R-1} \prod_{r=1}^{t_j} (x_j + r) \\ & \quad \cdot (x_R + 1 + x_{R+1} + 1) \cdots (x_R + 1 + x_{R+1} + 1 + T - 1) \\ &= \sum_{\mathcal{I}_s} \binom{\eta_s}{t_1, \dots, t_{R-1}, T} \prod_{j=1}^{R-1} \prod_{r=1}^{t_j} (x_j + r) \cdot \binom{T}{t_R, t_{R+1}} \prod_{j=R}^{R+1} \prod_{r=1}^{t_j} (x_R + r) \quad (\text{A.17}) \end{aligned}$$

In Equation A.17,

$$\begin{aligned} & \binom{\eta_s}{t_1, \dots, t_{R-1}, T} \cdot \binom{T}{t_R, t_{R+1}} \\ &= \frac{n_s}{t_1! \cdots t_{R-1}! T!} \cdot \frac{T!}{t_R! t_{R+1}!} = \binom{\eta_s}{t_1, \dots, t_{R+1}}, \quad (\text{A.18}) \end{aligned}$$

so

$$\prod_{i=0}^{n_s-1} \binom{R+1}{\sum_{j=1}^{R+1} x_j + i + R + 1} = \sum_{\mathcal{I}_s} \binom{\eta_s}{t_1, \dots, t_{R+1}} \prod_{j=1}^{R+1} \prod_{r=1}^{t_j} (x_j + r). \quad (\text{A.19})$$

By multiplying Equation A.19 with $\prod_{j=1}^{R+1} x_j!$, we can get

$$\sum_{\mathcal{I}_s} \binom{\eta_s}{t_1, \dots, t_{R+1}} \prod_{j=1}^{R+1} (x_j + t_j)! = \prod_{j=1}^{R+1} x_j! \prod_{i=0}^{n_s-1} \binom{R+1}{\sum_{j=1}^{R+1} x_j + i + R + 1}. \quad (\text{A.20})$$

2. Induction on n_s

RHS of equation A.16 is

$$\begin{aligned}
& \prod_{i=0}^{n_s} \left(\sum_{j=1}^R x_j + i + R \right) \\
&= \prod_{i=0}^{n_s-1} \left(\sum_{j=1}^R x_j + i + R \right) \cdot \left(\sum_{j=1}^R x_j + n_s + R \right) \\
&= \sum_{\mathcal{J}_s} \binom{\eta_s}{t_1, \dots, t_R} \prod_{j=1}^R (x_j + t_j)! \cdot \left(\sum_{j=1}^R x_j + n_s + R \right) \\
&= \sum_{\mathcal{J}_s} \binom{\eta_s}{t_1, \dots, t_R} \prod_{j=1}^R (x_j + t_j)! \sum_{k=1}^R (x_k + 1 + t_k) \\
&= \sum_{k=1}^R \sum_{\mathcal{J}_s} \binom{\eta_s}{t_1, \dots, t_R} \prod_{j=1}^R (x_j + t_j)! (x_k + t_k + 1) \tag{A.21}
\end{aligned}$$

Following the multinomial theorem, which indicates that

$$\begin{aligned}
(x_1 + x_2 + \dots + x_m)^{n+1} &= \sum_{\mathcal{J}_s} \binom{n}{t_1, \dots, t_m} x_1^{t_1} \dots x_m^{t_m} \cdot (x_1 + \dots + x_m) \\
&= \sum_{j=1}^m \sum_{\mathcal{J}_s} \binom{n}{t_1, \dots, t_m} x_1^{t_1} \dots x_m^{t_m} x_j \\
&= \sum_{\mathcal{J}_s} \binom{n}{t_1, \dots, t_m} x_1^{t_1} \dots x_m^{t_m}, \tag{A.22}
\end{aligned}$$

Equation A.21 can be written as

$$\sum_{k=1}^R \sum_{\mathcal{J}_s} \binom{\eta_s}{t_1, \dots, t_R} \prod_{j=1}^R (x_j + t_j)! (x_k + t_k + 1) = \sum_{\mathcal{J}_s} \binom{\eta_s}{t_1, \dots, t_R} \prod_{j=1}^R (x_j + t_j)! \tag{A.23}$$

By multiplying Equation A.19 with $\prod_{k=1}^{R+1} x_k!$, we can get

$$\sum_{\mathcal{J}_s} \binom{\eta_s + 1}{t_1, \dots, t_R} \prod_{j=1}^R (x_j + t_j)! = \prod_{k=1}^R x_k! \cdot \prod_{i=0}^{n_s} \left(\sum_{j=1}^{R+1} x_j + i + R + 1 \right), \tag{A.24}$$

which completes the induction and the proof.

Appendix B

Estimation of Mode Transition Probability

Suppose that the mode transition probability from $M^{t-1} = m_u$ to $M^t = m_v$, i.e., $p(m_v|m_u)$ is about to be estimated from the composite mode data:

$$p(M^t|M^{t-1}, \mathcal{D}_M) = p(m_v|m_u, \mathcal{D}_M) \quad (\text{B.1})$$

where

$$m_v, m_u \in \mathcal{M} = \{m_1, \dots, m_Q\}. \quad (\text{B.2})$$

This transition probability can be calculated by marginalization over all possible mode transition parameters,

$$\begin{aligned} & p(m_v|m_u, \mathcal{D}_M) \\ &= \int_{\Sigma_1, \dots, \Sigma_Q} p(m_v|m_u, \mathcal{D}_M, \Upsilon_1, \dots, \Upsilon_Q) f(\Upsilon_1, \dots, \Upsilon_Q|m_u, \mathcal{D}_M) d\Upsilon_1 \dots \Upsilon_Q, \end{aligned} \quad (\text{B.3})$$

where $\Upsilon_i = \{\varphi_{i,1}, \dots, \varphi_{i,Q}\}$ is the probability parameter set for mode transition starting from mode $M^{t-1} = m_i$; Q is the total number of possible modes; for instance, $\varphi_{i,j} = p(m_j|m_i)$; Σ_i is the space of all possible parameter sets Υ_i , where $\sum_{j=1}^Q \varphi_{i,j} = 1$.

The transition probability from m_u to m_v depends only on the parameter set Υ_u ,

$$p(m_v|m_u, \mathcal{D}_M, \Upsilon_1, \dots, \Upsilon_Q) = p(m_v|m_u, \Upsilon_u). \quad (\text{B.4})$$

Thus, Equation B.3 can be written as

$$\begin{aligned}
& p(m_v|m_u, \mathcal{D}_M) \\
&= \int_{\Sigma_1, \dots, \Sigma_Q} p(m_v|m_u, \mathcal{D}_M, \Upsilon_1, \dots, \Upsilon_Q) f(\Upsilon_1, \dots, \Upsilon_Q|m_u, \mathcal{D}_M) d\Upsilon_1 \cdots d\Upsilon_Q \\
&= \int_{\Sigma_1, \dots, \Sigma_Q} p(m_v|m_u, \Upsilon_u) f(\Upsilon_1, \dots, \Upsilon_Q|m_u, \mathcal{D}_M) d\Upsilon_1 \cdots d\Upsilon_Q \\
&= \int_{\Sigma_1, \dots, \Sigma_Q} \varphi_{u,v} f(\Upsilon_1, \dots, \Upsilon_Q|m_u, \mathcal{D}_M) d\Upsilon_1 \cdots d\Upsilon_Q. \tag{B.5}
\end{aligned}$$

The second term in the integration can be calculated from a Bayesian perspective,

$$f(\Upsilon_1, \dots, \Upsilon_Q|m_u, \mathcal{D}_M) = \frac{p(\mathcal{D}_M|\Upsilon_1, \dots, \Upsilon_Q, m_u) f(\Upsilon_1, \dots, \Upsilon_Q|m_u)}{p(\mathcal{D}_M|m_u)}, \tag{B.6}$$

where $p(\mathcal{D}_M|m_u)$ is the scaling factor,

$$p(\mathcal{D}|m_u) = \int_{\Sigma_1, \dots, \Sigma_Q} p(\mathcal{D}_M|\Upsilon_1, \dots, \Upsilon_Q, m_u) f(\Upsilon_1, \dots, \Upsilon_Q|m_u) d\Upsilon_1 \cdots d\Upsilon_Q. \tag{B.7}$$

In Equation B.6, the first term in the numerator is the likelihood of historical composite mode data. It is solely determined by the parameter set $\{\Upsilon_1, \dots, \Upsilon_Q\}$, and is independent of m_u , i.e.

$$p(\mathcal{D}_M|\Upsilon_1, \dots, \Upsilon_Q, m_u) = p(\mathcal{D}_M|\Upsilon_1, \dots, \Upsilon_Q) = \prod_{i=1}^Q \prod_{j=1}^Q \varphi_{i,j}^{\hat{n}_{i,j}}, \tag{B.8}$$

where $\hat{n}_{i,j}$ is the number of mode transitions from $M^{t-1} = m_i$ to $M^t = m_j$ in the historical composite mode data set.

In accordance with the common assumption that the priors of different parameter sets Υ_i and Υ_j , where $i \neq j$, are independent [73],

$$f(\Upsilon_1, \dots, \Upsilon_Q|m_u) = f(\Upsilon_1|m_u) \cdots f(\Upsilon_Q|m_u). \tag{B.9}$$

Dirichlet distribution is usually used for the priors of the mode transition parameters Υ_i with parameters $c_{i,1}, \dots, c_{i,Q}$,

$$f(\Upsilon_i|m_u) = \frac{\Gamma(\sum_{j=1}^Q c_{i,j})}{\prod_{j=1}^Q \Gamma(c_{i,j})} \prod_{j=1}^Q \varphi_{i,j}^{c_{i,j}-1}. \tag{B.10}$$

As a result, we have

$$f(\Upsilon_1, \dots, \Upsilon_Q|m_u) = \prod_{i=1}^Q \frac{\Gamma(\sum_{j=1}^Q c_{i,j})}{\prod_{j=1}^Q \Gamma(c_{i,j})} \prod_{j=1}^Q \varphi_{i,j}^{c_{i,j}-1}, \tag{B.11}$$

where $c_{i,j}$ can be interpreted as the number of prior samples for mode transition from m_i to m_j . $\Gamma(\cdot)$ is the gamma function, defined as

$$\Gamma(x) = \int_0^\infty t^{x-1} e^{-t} dt. \quad (\text{B.12})$$

Since in this thesis all the independent variables x of gamma functions are counts of mode transitions, which are all positive integers, so

$$\Gamma(x) = (x-1)!. \quad (\text{B.13})$$

Substituting Equation B.11 and B.8 in Equation B.6, we have

$$\begin{aligned} & f(\Upsilon_1, \dots, \Upsilon_Q | m_u, \mathcal{D}_M) \\ &= \frac{p(\mathcal{D}_M | \Upsilon_1, \dots, \Upsilon_Q, m_u) f(\Upsilon_1, \dots, \Upsilon_Q | m_u)}{p(\mathcal{D}_M | m_u)} \\ &= \frac{1}{p(\mathcal{D}_M | m_u)} \cdot \prod_{i=1}^Q \frac{\Gamma(\sum_{j=1}^Q c_{i,j})}{\prod_{j=1}^Q \Gamma(c_{i,j})} \prod_{j=1}^Q \varphi_{i,j}^{c_{i,j}-1} \cdot \prod_{i=1}^Q \prod_{j=1}^Q \varphi_{i,j}^{\hat{n}_{i,j}} \end{aligned} \quad (\text{B.14})$$

Let

$$\rho = \prod_{i=1}^Q \frac{\Gamma(\sum_{j=1}^Q c_{i,j})}{\prod_{j=1}^Q \Gamma(c_{i,j})}, \quad (\text{B.15})$$

and then Equation B.14 can be written as

$$\begin{aligned} & f(\Upsilon_1, \dots, \Upsilon_Q | m_u, \mathcal{D}_M) \\ &= \frac{\rho}{p(\mathcal{D}_M | m_u)} \cdot \prod_{i=1}^Q \prod_{j=1}^Q \varphi_{i,j}^{c_{i,j}-1} \cdot \prod_{i=1}^Q \prod_{j=1}^Q \varphi_{i,j}^{\hat{n}_{i,j}} \\ &= \frac{\rho}{p(\mathcal{D}_M | m_u)} \prod_{i=1}^Q \prod_{j=1}^Q \varphi_{i,j}^{\hat{n}_{i,j} + c_{i,j} - 1}. \end{aligned} \quad (\text{B.16})$$

Therefore, the transition probability from evidence m_u to m_v can be derived as

$$\begin{aligned} & p(m_v | m_u, \mathcal{D}_M) \\ &= \int_{\Sigma_1, \dots, \Sigma_Q} \varphi_{u,v} f(\Upsilon_1, \dots, \Upsilon_Q | m_u, \mathcal{D}_M) d\Upsilon_1 \dots \Upsilon_Q \\ &= \int_{\Sigma_1, \dots, \Sigma_Q} \varphi_{u,v} \frac{\rho}{p(\mathcal{D}_M | m_u)} \prod_{i=1}^Q \prod_{j=1}^Q \varphi_{i,j}^{\hat{n}_{i,j} + c_{i,j} - 1} d\Upsilon_1 \dots \Upsilon_Q \\ &= \frac{\rho}{p(\mathcal{D}_M | m_u)} \int_{\Sigma_1} \prod_{j=1}^Q \varphi_{1,j}^{\hat{n}_{1,j} + c_{1,j} - 1} d\Upsilon_1 \dots \int_{\Sigma_u} \varphi_{u,v}^{\hat{n}_{u,v} + c_{u,v}} \prod_{j \neq v} \varphi_{u,j}^{\hat{n}_{u,j} + c_{u,j} - 1} d\Upsilon_u \\ & \quad \dots \int_{\Sigma_Q} \prod_{j=1}^Q \varphi_{A,j}^{\hat{n}_{Q,j} + c_{K,j} - 1} d\Upsilon_Q, \end{aligned} \quad (\text{B.17})$$

In the above equation, $p(\mathcal{D}_M|m_u)$ is the scaling factor as defined in Equation B.6. According to Equation B.7,

$$\begin{aligned}
& p(\mathcal{D}_M|m_u) \\
&= \int_{\Sigma_1, \dots, \Sigma_Q} p(\mathcal{D}_M|\Upsilon_1, \dots, \Upsilon_Q, m_u) f(\Upsilon_1, \dots, \Upsilon_Q|m_u) d\Upsilon_1 \cdots d\Upsilon_Q \\
&= \int_{\Sigma_1, \dots, \Sigma_Q} \prod_{i=1}^Q \frac{\Gamma(\sum_{j=1}^Q c_{i,j})}{\prod_{j=1}^Q \Gamma(c_{i,j})} \prod_{j=1}^Q \varphi_{i,j}^{c_{i,j}-1} \cdot \prod_{i=1}^Q \prod_{j=1}^Q \varphi_{i,j}^{\hat{n}_{i,j}} d\Upsilon_1 \cdots d\Upsilon_Q \\
&= \rho \cdot \int_{\Sigma_1} \prod_{j=1}^Q \varphi_{1,j}^{\hat{n}_{1,j}+c_{1,j}-1} d\Upsilon_1 \cdots \int_{\Sigma_Q} \prod_{j=1}^Q \varphi_{Q,j}^{\hat{n}_{Q,j}+c_{Q,j}-1} d\Upsilon_Q \\
&= \rho \cdot \prod_{i=1}^Q \frac{\prod_{j=1}^Q \Gamma(\hat{n}_{i,j} + c_{i,j})}{\Gamma(\hat{N}_i + C_i)}, \tag{B.18}
\end{aligned}$$

where $\hat{N}_i = \sum_j \hat{n}_{i,j}$ is the total number of mode transitions, namely the number of historical composite mode samples, from previous mode m_i , and $C_i = \sum_j c_{i,j}$ is the corresponding total number of prior composite mode samples.

Similarly, we can derive

$$\begin{aligned}
& \int_{\Sigma_1} \prod_{j=1}^Q \varphi_{1,j}^{\hat{n}_{1,j}+c_{1,j}-1} d\Upsilon_1 \cdots \int_{\Sigma_u} \varphi_{u,v}^{\hat{n}_{u,v}+c_{u,v}} \prod_{j \neq v} \varphi_{u,j}^{\hat{n}_{u,j}+c_{u,j}-1} d\Upsilon_u \\
& \quad \cdots \int_{\Sigma_Q} \prod_{j=1}^Q \varphi_{Q,j}^{\hat{n}_{Q,j}+c_{Q,j}-1} d\Upsilon_Q \\
&= \frac{\Gamma(\hat{n}_{u,v} + c_{u,v} + 1)}{\Gamma(\hat{N}_u + C_u + 1)} \cdot \frac{\prod_{i,j \neq u,v} \Gamma(\hat{n}_{i,j} + c_{i,j})}{\prod_{i \neq u} \Gamma(\hat{N}_i + C_i)}. \tag{B.19}
\end{aligned}$$

Thus, Equation B.17 can be simplified as

$$\begin{aligned}
& p(m_v|m_u, \mathcal{D}_E) \\
&= \rho \cdot \frac{\Gamma(\hat{n}_{u,v} + c_{u,v} + 1)}{\Gamma(\hat{N}_u + C_u + 1)} \cdot \frac{\prod_{i,j \neq u,v} \Gamma(\hat{n}_{i,j} + c_{i,j})}{\prod_{i \neq u} \Gamma(\hat{N}_i + C_i)} \cdot \frac{\prod_{i=1}^Q \Gamma(\hat{N}_i + C_i)}{\rho \cdot \prod_{i=1}^Q \prod_{j=1}^Q \Gamma(\hat{n}_{i,j} + c_{i,j})} \\
&= \frac{\hat{n}_{u,v} + c_{u,v}}{\hat{N}_u + C_u}. \tag{B.20}
\end{aligned}$$