AIChE

A Bayesian Approach for Control Loop Diagnosis with Missing Data

Fei Qi and Biao Huang
Dept. of Chemical and Materials Engineering, University of Alberta, Edmonton, AB, T6G 2G6, Canada

Edgar C. Tamayo
Syncrude Canada Ltd., Fort McMurray, AB, T9H 3L1, Canada

DOI 10.1002/aic.11964
Published online September 25, 2009 in Wiley InterScience (www.interscience.wiley.com)

This article 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. A main issue encountered in the application of the data-driven approach is the problem of missing data or missing monitor reading. By introducing the concept of missing pattern, data missing problems are classified into single and multi missing patterns. A novel method based on marginalization over underlying complete evidence matrix is proposed to circumvent missing data problems. Performance of the proposed Bayesian approach is examined through simulations as well as an industrial application example to verify its ability of information synthesis. © 2009 American Institute of Chemical Engineers *AIChE J,* 56: 179–195, 2010
Keywords: process control, performance monitoring, performance assessment, bayesian method, control loop diagnosis, missing data, fault diagnosis

Introduction

The research on control loop performance monitoring and diagnosis remains one of the most active research areas in process control community. The main objective 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 loop,[1] as well as to suggest possible problem source and troubleshooting sequence.

A number of benchmarks for control performance monitoring have been proposed, including minimum variance control (MVC), linear quadratic Gaussian control (LQG), historical data based, and user-specified control.[1–4] A number of successful applications of control performance monitoring algorithms have been reported. Significant progress has also been made in the development of instrument and process monitors, including sensor monitors, actuator monitors, and model validation monitors. However, most of the monitoring algorithms target only one specific problem in a control loop with the assumption that other problems do not occur simultaneously. Obviously, 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.[1,5–7] Continuous improvement in control performance must be accompanied by constantly monitoring the performance of the control loops and diagnosing the source of poor performance, such as poor tuning, sticky valve, major disturbance upset, or other root causes.[1,7] Therefore, there is a strong incentive to develop methods that not only monitor control performance but also to isolate the underlying source of problematic control performance in the system.

There are several challenging issues in synthesizing performance monitors.[8] First, although problem sources may be different, the symptoms can be similar. For example, valve

Correspondence concerning this article should be addressed to B. Huang at biao.huang@ualberta.ca

stiction can cause loop oscillation. So does an over-tuned controller. Second, all processes operate in uncertain environment to some extent, and there are uncertainties in the links between problem sources and monitor readings. For example, a sensor monitor does not necessarily issue an alarm even if there is indeed a problem in the sensor. Thus, a solution should be built on a probabilistic framework. Last but not least, how to incorporate *a priori* knowledge in the diagnostic system to improve diagnosis 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 may also be necessary for an accurate diagnosis.[8]

Bayesian methods have 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.[9–12] It is one of the most widely applied techniques in probabilistic inferencing.[10] Built upon previous work in Bayesian fault diagnosis[9] and a framework outlined in Huang (2008),[8] this article develops a data-driven algorithm for control loop diagnosis based on the Bayesian framework with consideration of missing data.

The remainder of this article is organized as follows: 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 presented. An approach based on marginalization over underlying complete evidence matrix (UCEM) is then proposed to handle the missing data problem. Simulations for a binary distillation column are presented. The diagnosis approach is applied to an industry process. The final section concludes this article with discussion on the results obtained.

Data-Driven Bayesian Approach for Control Loop Diagnosis

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 malfunctionings. For example, a valve acting as an actuator may suffer from 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, and loop oscillation.

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 1, where the monitors marked with gray may respond to the valve stiction problem. It is challenging for the operator 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.

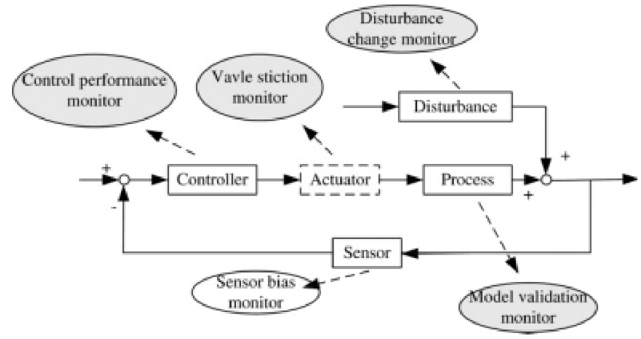

Figure 1. Typical control system structure.

In the presence of disturbances, Bayesian inference provides a well suited way to solve the diagnosis problem, quantifying the uncertainty in its conclusion. In the work of Pernestal,[9] a Bayesian approach for diesel engine diagnosis based on complete sensor readings is studied. This section applies this approach[9] to control loop diagnosis problem based on complete readings of control loop monitors.

Preliminaries

To adopt the Bayesian method for control loop diagnosis, several notations are introduced in this section.

Mode M. Assume that a control loop under diagnosis consists of P components of interest: C_1, C_2,…,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 diagnosis problem is to determine the operating status of all these components in the case of problematic loop performance, that is to locate the underlying problem source of problematic control performance. An assignment of operating status to all the components of interest in the control loop is called a mode, and denoted as M; M can take different values and its specific value is denoted by m. For example, $m = (C_1 =$ well-tuned controller, $C_2 =$ valve with stiction, …). Mode representing normal operation is denoted 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

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

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

Often, the monitor readings, which are generally continuous, are discretized according to 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, $e = (\pi_1 =$ optimal control performance, $\pi_2 =$ no sensor bias,...). Suppose that the single monitor output π_i has k_i different discrete values. Then there are totally

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

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

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

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

$$\mathcal{D} = \{d^1, d^2, \ldots, d^{\tilde{N}}\},$$

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

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

and

$$\mathcal{D}_{m_i} = \{d^1_{m_i}, \ldots, d^{N_{m_i}}_{m_i}\}$$

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 $\sum_i N_{m_i} = \tilde{N}$.

Different historical data samples may be autocorrelated or independent depending 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 work, all the historical data samples are assumed to be independent, i.e.,

$$p(\mathcal{D}) = p(d^1, d^2, \ldots, d^{\tilde{N}}) = p(d^1)p(d^2)\ldots p(d^{\tilde{N}}). \quad (1)$$

However, even for a medium-sized system, there are many possible combinations of abnormalities, so it is infeasible to collect data from all possible modes of a large-scale system. Further, there may be some abnormalities, or combination of abnormalities, which have not occurred before, and hence no corresponding historical data are available. These modes are all categorized as the unconsidered (UC) mode.

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

$$p(M|E,\mathcal{D}) = \frac{p(E|M,\mathcal{D})p(M|\mathcal{D})}{p(E|\mathcal{D})}, \quad (2)$$

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 probabilities of the abnormalities, and as a result, $p(M|\mathcal{D}) = p(M)$.[9]

Estimation of likelihood probability

Because prior probabilities are determined by a *priori* information, and the scaling factor $p(E|\mathcal{D})$ can be calculated readily by

$$p(E|\mathcal{D}) = \sum_M p(E|M,\mathcal{D})p(M|\mathcal{D}),$$

the main task of building a Bayesian diagnostic system is the estimation of the likelihood probabilities with historical data $\mathcal{D}$, whose objective is to make the estimated likelihood probabilities be consistent with historical data $\mathcal{D}$. Pernestal[9] presented a data-driven Bayesian algorithm to estimate the likelihood. The 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

$$e_i \in \mathcal{E} = \{e_1, \ldots, e_L\},$$

and

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

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 $M = m_j$,

$$p(e_i|m_j,\mathcal{D}) = p(e_i|m_j,\mathcal{D}_{m_j},\mathcal{D}_{\neg m_j}) = p(e_i|m_j,\mathcal{D}_{m_j}), \quad (3)$$

where $D_{\neg m_j}$ is the data set whose underlying mode is not m_j.

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

$$p(e_i|m_j,\mathcal{D}_{m_j}) = \int_\Omega p(e_i|\Theta_{m_j},m_j,\mathcal{D}_{m_j})f(\Theta_{m_j}|m_j,\mathcal{D}_{m_j})d\Theta_{m_j}, \quad (4)$$

where $\Theta_{m_j} = \{\theta_{e1|mj}, \theta_{e2|mj},\ldots,\theta_{eK|mj}\}$ are the likelihood parameters for all possible evidences of mode m_j, K is the total number of possible evidences; for example, $\theta_{ei|mj} = p(e_i|m_j)$ 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 Eq. 4, $f(\Theta_{m_j}|m_j,\mathcal{D}_{m_j})$ can be calculated according to Bayes' rule:

$$f(\Theta_{m_j}|m_j,\mathcal{D}_{m_j}) = \frac{p(\mathcal{D}_{m_j}|\Theta_{m_j},m_j)f(\Theta_{m_j}|m_j)}{p(\mathcal{D}_{m_j}|m_j)}. \quad (5)$$

In Eq. 5, Dirichlet distribution is commonly used to model priors of the likelihood parameters with Dirichlet parameters $a_{1|m_j},\cdots,a_{K|m_j}$,[9]

$$f(\Theta_{m_j}|m_j) = \frac{\Gamma(\sum_{i=1}^K a_{i|m_j})}{\prod_{i=1}^K \Gamma(a_{i|m_j})}\prod_{i=1}^K \theta_{i|m_j}^{a_{i|m_j}-1}, \quad (6)$$

where $a_{i|mj}$ 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,

$$\Gamma(x) = \int_0^\infty t^{x-1}e^{-t}dt. \quad (7)$$

Here, all the independent variables x of the gamma functions take positive integers, so

$$\Gamma(x) = (x-1)!. \quad (8)$$

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

$$p(\mathcal{D}_{m_j}|\Theta_{m_j},m_j) = \prod_{t=1}^{N_{m_j}} p(d_{m_j}^t|\Theta_{m_j},m_j). \quad (9)$$

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 $e^t = e_i$,

$$p(d_{m_j}^t|\Theta_{m_j},m_j) = \theta_{i|m_j}. \quad (10)$$

Combining Eqs. 9 and 10, we have

$$p(\mathcal{D}_{m_j}|\Theta_{m_j},m_j) = \prod_{i=1}^K \theta_{i|m_j}^{n_{i|m_j}}, \quad (11)$$

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 Eqs. 11 and 5 in Eq. 4, the following result can be obtained for the likelihood:[9]

$$p(E=e_i|M=m_j,\mathcal{D}) = \frac{n_{i|m_j}+a_{i|m_j}}{N_{m_j}+A_{m_j}}, \quad (12)$$

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 data 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.[9] 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 noninformative 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 2a. 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 2a. 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 Eq. 12, as presented in Figure 2b.

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 Eq. 2. 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 to generate posterior probabilities for diagnosis.

Diagnosis with Missing Data

Although a Bayesian procedure for data-driven control loop diagnosis has been discussed in the previous section,

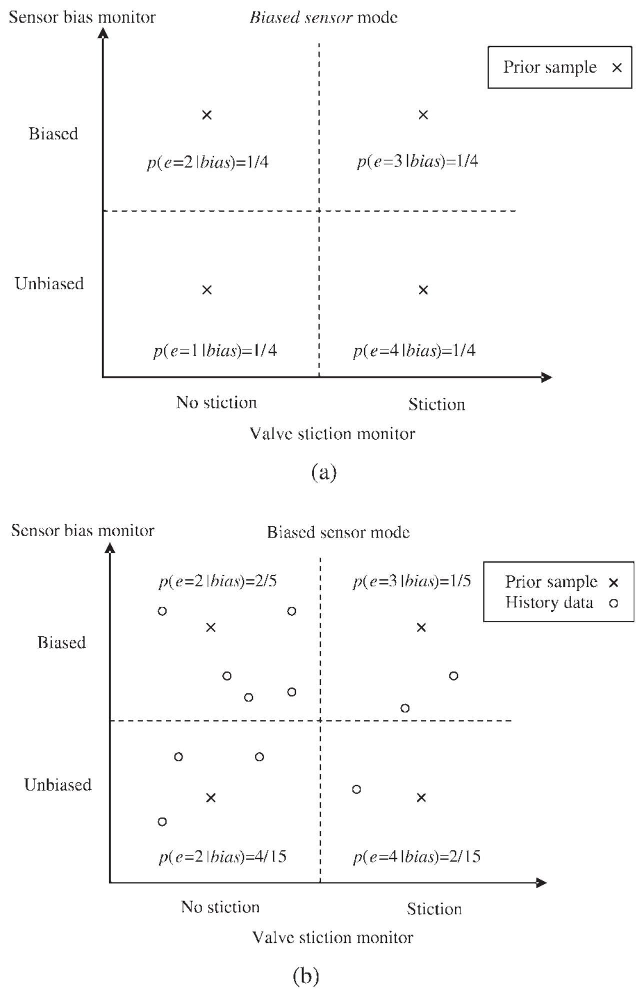

Figure 2. Likelihood updating.

(a) Likelihood with only prior samples.

(b) Updated likelihood with historical data.

main problems remain. An outstanding problem of the procedure described earlier is its inability to handle missing data. It is not uncommon that some monitor readings are not available in part of historical data record. However, in the method described so far, only complete historical data 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 section.

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

Problem formulation

The historical data set includes two parts, the complete data samples and the incomplete ones,

$$\mathcal{D} = \{\mathcal{D}_{\rm c}, \mathcal{D}_{\rm ic}\},$$

where $\mathcal{D}_{\rm c}$ is the data set with complete evidences, and the $\mathcal{D}_{\rm ic}$ 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,

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

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 data missing problems.

Missing pattern is determined by the locations of the missing data, 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 (×,×,0) and (×,×,1) belong to the same missing pattern, where × denotes the missing value. Otherwise, they are said to fall into different missing patterns. For instance, the two evidence readings (×,1,0) and (1,×,1) belong to two different missing patterns; it should be noted that two evidence readings (×,×,0) and (1,×,1) are also from two distinct missing patterns. By enumerating the numbers of missing patterns in the historical data set, the data missing problems can be classified into single missing pattern ones and multi missing pattern ones.

In a multi 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

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

Each of the available monitor readings $\pi_{q+1},\ldots,\pi_L$ 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 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

$$\begin{matrix}(\times,\times,0)\\(\times,\times,1)\end{matrix}\begin{bmatrix}(0,0,0) & (0,1,0) & (1,0,0) & (1,1,0)\\(0,0,1) & (0,1,1) & (1,0,1) & (1,1,1)\end{bmatrix}. \quad (13)$$

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 $\varepsilon_{i,j/k}$; 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 $\varepsilon_{i/k}$; 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 Eq. 13, which is assumed to be the first UCEM, the evidence (0,0,0) is denoted as $\varepsilon_{1,1/1}$; the number of historical samples with this evidence is denoted as $\eta_{1,1/1}$. The corresponding incomplete evidence, (×,×,0), is denoted as $\varepsilon_{1/1}$; the historical sample number of (×,×,0) is $\eta_{1/1}$, and its likelihood is $\lambda_{1/1}$.

The solutions

Consider Eq. 5,

$$f(\Theta|M,\mathcal{D}) = \frac{p(\mathcal{D}|\Theta,M)f(\Theta|M)}{p(\mathcal{D}|M)}.$$

The numerator can be calculated by integration over the likelihood space Ω,

$$p(\mathcal{D}|M) = \int_{\Omega} p(\mathcal{D}|\Theta,M)f(\Theta|M)d\Omega. \quad (14)$$

Our interest lies in how to derive the likelihood of historical data $p(\mathcal{D}|\Theta,M)$ in the presence of missing data.

When there are only complete historical samples, $p(\mathcal{D}|\Theta,M)$ can be calculated with Eq. 11. 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 $d_{ic}^{t} = (\varepsilon_{i/k}, m)$ equals the likelihood of the incomplete evidence $\varepsilon_{i/k}$,

$$p(d_{\text{ic}}^{t}|\Theta,M) = p((\varepsilon_{i/k},M)|\theta,M) = p(\varepsilon_{i/k}|\Theta,M) = \lambda_{i/k}, \quad (15)$$

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,

$$p((\varepsilon_{i/k},M)|\theta,M) = \sum_{j=1}^{R} p(\varepsilon_{i,j/k}|\Theta,M) = \sum_{j=1}^{R}\theta_{i,j/k}. \quad (16)$$

Take the UCEM in Eq. 13 as an example. The likelihood of incomplete evidence (×,×,0) equals the summation of likelihood from (0,0,0) to (1,1,0) in the first row,

$$p((\varepsilon_{1/1},M)|\Theta,M) = \lambda_{1/1} = \sum_{j=1}^{4}\theta_{1,j/1}. \quad (17)$$

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

$$\begin{aligned}p(\mathcal{D}|\Theta,M) &= p(\mathcal{D}_{\text{c}}|\Theta,M)p(\mathcal{D}_{\text{ic}}|\Theta,M)\\ &= \prod_{i=1}^{S}\prod_{j=1}^{R}\theta_{i,j}^{\eta_{i,j}}\cdot\prod_{i=1}^{S}\lambda_i^{\eta_i} = \prod_{i=1}^{S}\left[\prod_{j=1}^{R}\theta_{i,j}^{\eta_{i,j}}\cdot\left(\sum_{k=1}^{R}\theta_{i,k}\right)^{\eta_i}\right].\end{aligned} \quad (18)$$

Note that because only one UCEM exits 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 Dicichlet distributed with parameters $a_{i,j}$ for the complete evidences,

$$f(\Theta|M) = \frac{\Gamma(\sum_{i=1}^{S}\sum_{j=1}^{R}a_{i,j})}{\prod_{i=1}^{S}\prod_{j=1}^{R}\Gamma(a_{i,j})}\prod_{i=1}^{S}\prod_{j=1}^{R}\theta_{i,j}^{a_{i,j}-1}. \quad (19)$$

Then,

$$\begin{aligned}f(\Theta|M,\mathcal{D}) &= \frac{p(\mathcal{D}|\Theta,M)f(\Theta|M)}{p(\mathcal{D}|M)}\\ &= \frac{1}{p(\mathcal{D}|M)}\cdot\prod_{i=1}^{S}\left[\prod_{j=1}^{R}\theta_{i,j}^{\eta_{i,j}}\cdot\left(\sum_{k=1}^{R}\theta_{i,k}\right)^{\eta_i}\right]\\ &\quad\cdot\frac{\Gamma(\sum_{i=1}^{S}\sum_{j=1}^{R}a_{i,j})}{\prod_{i=1}^{S}\prod_{j=1}^{R}\Gamma(a_{i,j})}\prod_{i=1}^{S}\prod_{j=1}^{R}\theta_{i,j}^{a_{i,j}-1}.\end{aligned} \quad (20)$$

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 Eq. 20 can be written as

$$\begin{aligned}&f(\Theta|M,\mathcal{D})\\ &= \frac{c}{p(\mathcal{D}|M)}\cdot\prod_{i=1}^{S}\left[\prod_{j=1}^{R}\theta_{i,j}^{\eta_{i,j}}\cdot\left(\sum_{k=1}^{R}\theta_{i,k}\right)^{\eta_i}\right]\cdot\prod_{i=1}^{S}\prod_{j=1}^{R}\theta_{i,j}^{a_{i,j}-1}\\ &= \frac{c}{p(\mathcal{D}|M)}\cdot\prod_{i=1}^{S}\left[\prod_{j=1}^{R}\theta_{i,j}^{\eta_{i,j}+a_{i,j}-1}\cdot\left(\sum_{k=1}^{R}\theta_{i,k}\right)^{\eta_i}\right]\\ &= \frac{c}{\int_{\Omega}p(\mathcal{D}|\Theta,M)f(\Theta|M)d\Theta}\cdot\prod_{i=1}^{S}\left[\prod_{j=1}^{R}\theta_{i,j}^{\eta_{i,j}+a_{i,j}-1}\cdot\left(\sum_{k=1}^{R}\theta_{i,k}\right)^{\eta_i}\right].\end{aligned} \quad (21)$$

After some derivations, as presented in Appendix the likelihood of evidence $\varepsilon_{s,r}$ can be determined as

$$p(\varepsilon_{s,r}|M,\mathcal{D}) = \frac{\eta_{s,r}+a_{s,r}}{N+A}\cdot\frac{\sum_{j=1}^{R}(\eta_{s,j}+a_{s,j})+\eta_s}{\sum_{j=1}^{R}(\eta_{s,j}+a_{s,j})} = \frac{\eta_{s,r}+a_{s,r}}{N+A}\cdot\left(1+\frac{\eta_s}{\sum_{j=1}^{R}(\eta_{s,j}+a_{s,j})}\right). \tag{22}$$

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

$$p((0,0)|M,\mathcal{D}) = \frac{\eta_{(0,0)}+a_{(0,0)}}{N_c+A}, \tag{23}$$

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 Eq. 22, the likelihood becomes

$$p((0,0)|m,\mathcal{D}) = \frac{\eta_{(0,0)}+a_{(0,0)}}{N+A}\cdot\left(1+\frac{\eta_{(0,\times)}}{\sum_{\pi_2}\eta_{(0,\pi_2)}+a_{(0,\pi_2)}}\right), \tag{24}$$

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

The estimated likelihood in Eq. 22 has an intuitive explanation. It can be rewritten as

$$p(\varepsilon_{s,r}|M,\mathcal{D}) = \frac{\eta_{s,r}+a_{s,r}+\eta'_{s,r}}{N+A}, \tag{25}$$

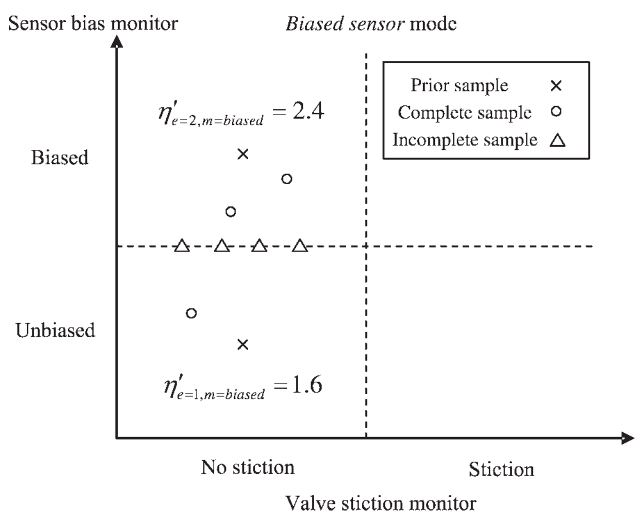

Figure 3. Estimation of expected numbers of complete evidences in the incomplete samples.

Table 1. Summary of Historical and Prior Samples

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

where

$$\eta'_{s,r} = \eta_s\cdot\frac{\eta_{s,r}+a_{s,r}}{\sum_{j=1}^{R}(\eta_{s,j}+a_{s,j})} \tag{26}$$

can be interpreted as the expected number of samples with evidence $\varepsilon_{s,r}$ in the incomplete data set. Interestingly, this number equals the expected value of a variable subject to Dirichlet distribution with parameters ($\eta_{s,1}+a_{s,1}$, $\eta_{s,2}+a_{s,2}$,…,$\eta_{s,R}+a_{s,R}$). 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 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 unbias. 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 using Eq. 26, as shown in Figure 3. 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 Eq. 25.

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

1. *Construct an UCEM for the incomplete data samples*;
2. *Calculate the expected numbers of all possible underlying evidences* $\eta'_{s,r}$ *according to Eq.* 26;
3. *Calculate the likelihood according to Eq.* 25.

By considering the incomplete samples, how will the likelihood be changed? Consider a data set with incomplete samples. Table 1 is the detail of the data set. The prior samples are uniformly distributed with one for each complete evidence.

Compare the two different missing data 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 2.

With the difference being significant, it is demonstrated that the information from the incomplete evidences is useful for the

Table 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

Table 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

estimation of likelihood. Simply 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 Eq. 22, if the missing ratio

$$\rho_s^{\text{miss}} = \frac{\eta_s}{\sum_{j=1}^{R}(\eta_{s,j} + a_{s,j})} \tag{27}$$

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

$$\begin{aligned} p(\varepsilon_{s,r}|M,\mathcal{D}) &= \frac{1}{N+A}\cdot\left[(\eta_{s,r}+a_{s,r}) + \eta_s\cdot\frac{\eta_{s,r}+a_{s,r}}{\sum_{j=1}^{R}(\eta_{s,j}+a_{s,j})}\right] \\ &= \frac{\eta_{s,r}+a_{s,r}}{N+A}\cdot\left[1+\rho_s^{\text{miss}}\right] \\ &= \frac{\eta_{s,r}+a_{s,r}}{\sum_{s,j}(\eta_{s,j}+a_{s,j}) + \rho_s^{\text{miss}}\sum_{s,j}(\eta_{s,j}+a_{s,j})}\cdot\left[1+\rho_s^{\text{miss}}\right] \\ &= \frac{\eta_{s,r}+a_{s,r}}{\sum_{s,j}(\eta_{s,j}+a_{s,j})} = \frac{\eta_{s,r}+a_{s,r}}{N_c + A}, \end{aligned} \tag{28}$$

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, with one prior sample being assigned to each complete evidence.

According to the definition of missing ratio in Eq. 27,

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

and

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

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.

Multi Missing Pattern Problem. The solution for the multi-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 patterns 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 $(\times,\pi_2)$ and $(\pi_1, \times)$. Let both π_1 and π_2 have two possible values 0 and 1. Two UCEMs are constructed for the historical data set:

$$\begin{matrix}(\times,0)\\(\times,1)\end{matrix}\begin{bmatrix}(0,0) & (1,0)\\(0,1) & (1,1)\end{bmatrix} \tag{29}$$

and

$$\begin{matrix}(0,\times)\\(1,\times)\end{matrix}\begin{bmatrix}(0,0) & (0,1)\\(1,0) & (1,1)\end{bmatrix}. \tag{30}$$

Both UCEMs need to be taken into consideration when calculating the likelihood of historical data, including both complete and incomplete samples. Because 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 Eq. 29.

In general, assume there are P missing patterns in the historical data set, the likelihood of the data set is calculated as

$$\begin{aligned} p(\mathcal{D}|\Theta,M) &= p(\mathcal{D}_c|\Theta,M)p(\mathcal{D}_{ic}|\Theta,M) \\ &= \prod_{i=1}^{S_1}\prod_{j=1}^{R_1}\theta_{i,j/1}^{\eta_{i,j/1}}\cdot\prod_{k=1}^{P}\prod_{i=1}^{S_k}\lambda_{i/k}^{\eta_{i/k}} \\ &= \prod_{i=1}^{S_1}\prod_{j=1}^{R_1}\theta_{i,j/1}^{\eta_{i,j/k}}\cdot\prod_{k=1}^{P}\prod_{i=1}^{S_k}\left(\sum_{j=1}^{R_k}\theta_{i,j/k}\right)^{\eta_{i/k}}, \end{aligned} \tag{31}$$

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 $\varepsilon_{i,j/k}$ 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 $\varepsilon_{s,r/1}$ can be derived, as

$$p(\varepsilon_{s,r/1}|M,\mathcal{D}) = \frac{1}{N+A}\cdot\frac{\sum_{\mathcal{T}^1\cdots\mathcal{T}^P}\prod_{k=1}^{P}\prod_{i=1}^{S_k}\left[\mathbf{C}_{\eta_{i/k}}^{\mathcal{T}_i^k}\Gamma(\sum_{j=1}^{P}\mathcal{Z}_{s,r}+1)\prod_{(u,v)\neq(s,r)}\Gamma(\mathcal{Z}_{u,v})\right]}{\sum_{\mathcal{T}^1\cdots\mathcal{T}^P}\prod_{k=1}^{P}\prod_{i=1}^{S_k}\left[\mathbf{C}_{\eta_{i/k}}^{\mathcal{T}_i^k}\prod_{u=1}^{S_1}\prod_{v=1}^{R_1}\Gamma(\mathcal{Z}_{u,v})\right]}, \tag{32}$$

where $\mathcal{T}^i$ is all the combination of $\mathcal{T}_j$ for the i-th missing pattern; $\mathcal{Z}_{s,r} = \sum_{j=1}^{P}\tilde{t}_{s,r/j} + \eta_{s,r/1} + \alpha_{s,r/1}$; $\tilde{t}_{s,r/j}$ is the possible number of complete evidence $\varepsilon_{s,r/1}$ 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 $\varepsilon_{s,r/1}$ 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 multi missing patterns presented in Eq. 32, 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.

However it is worthy to point out that the single missing pattern solution does have its practical merit whether the data have single missing pattern or multi missing patterns. In general, the multi 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 data missing problem. If data missing does not occur simultaneously across all the problematic variables, a multi 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 multi missing patterns can be fitted into this singlemissing 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., multi 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 multi missing pattern problem.

Simulation Example

Process description

To investigate diagnosis performance of the proposed Bayesian approach for process with advanced process control

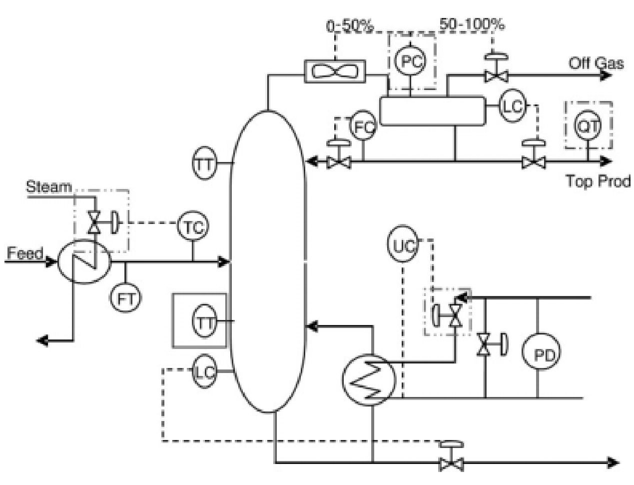

Figure 4. Distillation column simulation system.

Table 4. 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 and 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

scheme, we apply the algorithm to a simulated binary distillation column (Figure 4).[13]

The column has five inputs, four of which are MVs operated by a model predictive controller (MPC). Of the 10 outputs, three are controlled quality variables. They are as follows: 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 4.

Monitor selection

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

Control Performance Monitor. The MVC benchmark is adopted to evaluate control performances for both univariate and multivariate cases. The FCOR algorithm[2] is used 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 the sinusoid form, whereas 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[14] is used 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 earlier, 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

Table 5. 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
π_{10}, π_{11}, π_{12}	Disturbance change monitors for the three quality variables
π_{13}, π_{14}, π_{15}	Sensor bias detection monitors for the three quality variables

process model with inverse of the disturbance model yields the input to the disturbance model $\tilde{e}(t)$,

$$\tilde{e}(t) = G_l^{-1}[y(t) - \hat{y}(t)], \tag{33}$$

where $y(t)$ is the process output, and $\hat{y}(t)$ is the simulated output. 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.[15]

Diagnosis settings and results

Because the three quality CVs are of the main interest, the selected monitors mainly target these CVs, as shown in Table 5.

The parameter settings of the Bayesian diagnostic system are summarized in Table 6. 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.

Diagnosis results in Figure 5 are obtained from evaluation (cross-validation) data, which are generated independently of historical samples. In Figure 5, the title of each plot denotes the true underlying mode, and the numbers on the horizontal axis stand for the diagnosed 11 possible modes numbered according to the sequence shown in the first column in Table 4. In each plot, the posterior probability corresponding to the true underlying mode is highlighted with gray bars, whereas

Table 6. Summary of Bayesian Diagnosis Parameters

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

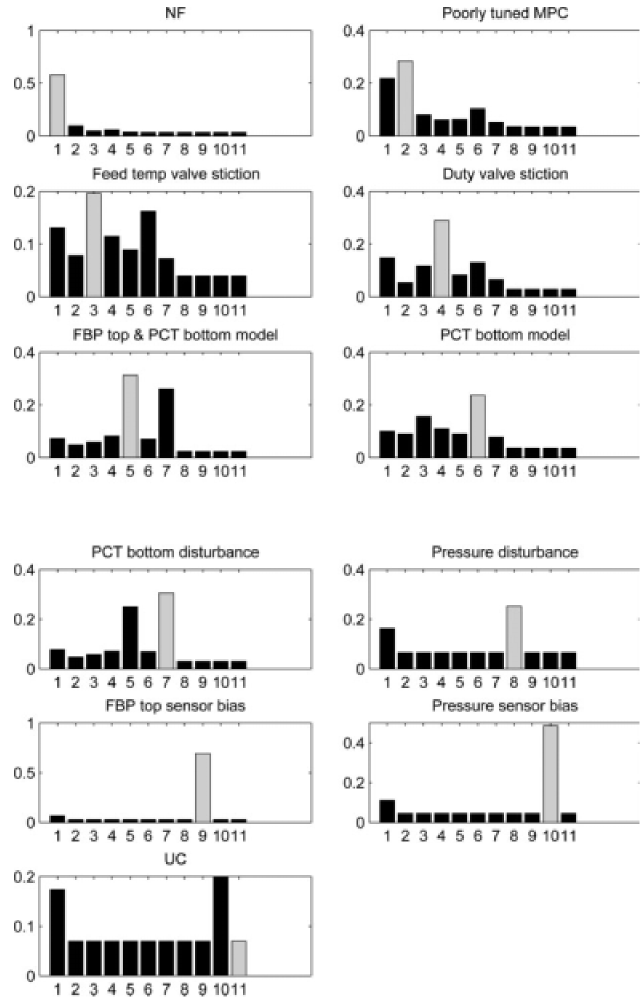

Figure 5. Posterior probability assigned to each mode.

the others are in dark bars. The diagnostic conclusion is determined by picking up 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 5, 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 diagnosis 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.

Missing data handling

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

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

In the simulation, set that π_6 and π_9 tend to be missing when the discrete output of the pressure sensor bias monitor reading is "high". A total of 90% of π_6 and π_9 are missing 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 and π_9 are missing is 10%. 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". Although most of such data samples are incomplete, the distribution of the complete evidences for m_9 is distorted as shown in Figure 6, which will deteriorate the diagnosis 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 Eq. 26. 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 data problem), the set with missing data, and the recovered data set are shown in Figure 6. It should be noted that the missing evidences that do not appear in the set with missing data are now recovered and displayed in Figure 6, together with the original ones.

Clearly, the evidence distribution is seriously distorted if only complete samples in the set with missing data are considered. Although the proposed approach cannot recover all underlying missing evidences, i.e., to make the dotted line in 6 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 diagnosis results of m_9 mode using the original data set, and the diagnosis result from the set with missing data 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 7. By ignoring all the incomplete samples, the posterior probability assigned to mode m_9, which is the true underlying mode, is only

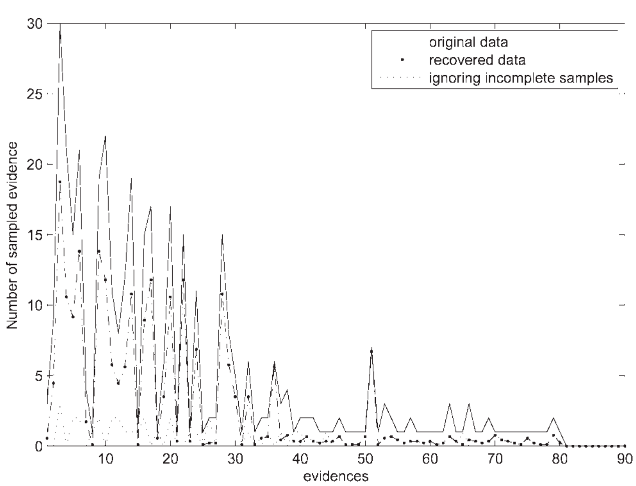

Figure 6. Comparison of complete evidence numbers.

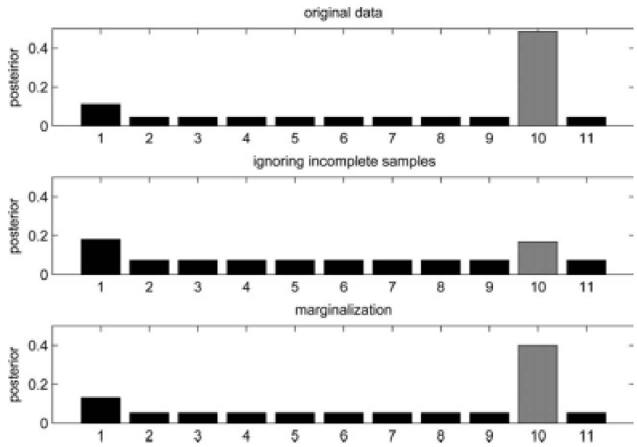

Figure 7. Diagnosis results with different data set.

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.

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 missing data handling strategies as discussed earlier are displayed in Figure 8. By comparing Figure 7 and 8, 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 9. The small standard deviations of the Monte Carlo simulations indicate once again quite consistent results of each simulation.

In the histogram of the posterior probabilities assigned to mode m_9 shown in Figure 10, one can see that the posteriors obtained from original data set are always the highest. Although the posteriors calculated from incomplete data set

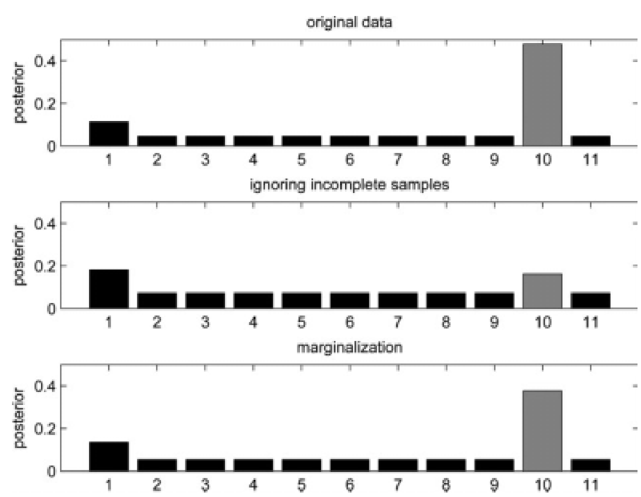

Figure 8. Mean value of posteriors with Monte Carlo simulations.

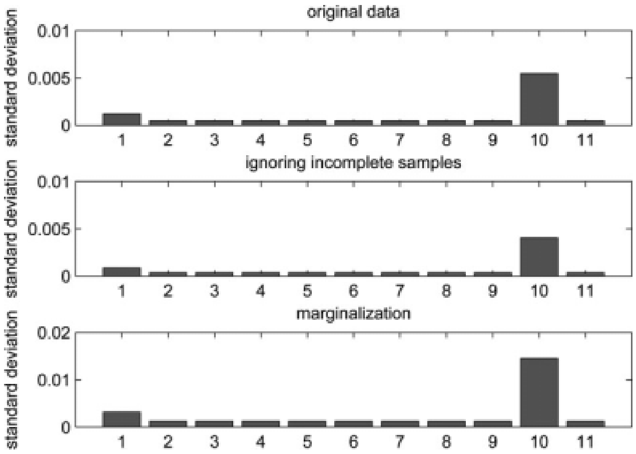

Figure 9. Standard deviations of posteriors with Monte Carlo simulations.

with 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 diagnosis performance in the presence of missing data.

Gibbs Sampling Method. The Gibbs sampling method is selected as an example to compare the performance of the proposed missing data handling strategy with those of traditional ones. Gibbs sampling is a Monte Carlo Markov chain (MCMC) method introduced by Geman[16] and Geman. It is capable of estimating the parameters in case of missing data, and thereby reconstruct the missing data according to the estimated parameters. Readers can refer to Ref. 17 for details of the algorithm. A software named BUGS (Bayesian inference Using Gibbs Sampling)[18] is used 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 proposed approach, and the recovered data with Gibbs sampling are shown in Figure 11. It can be observed that the evidence distribution

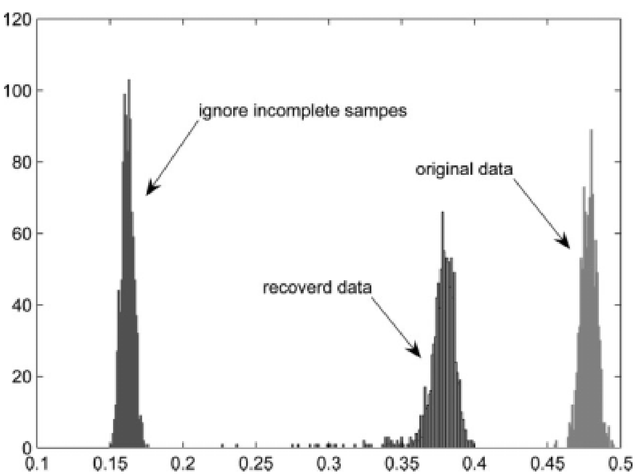

Figure 10. Histogram of posteriors assigned to m_9.

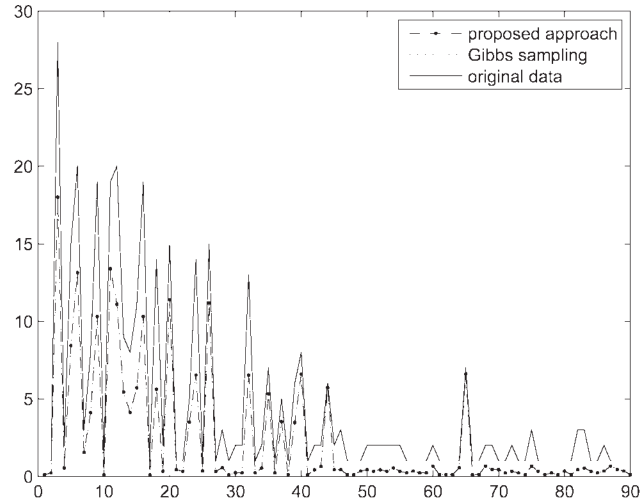

Figure 11. Comparison of complete evidence numbers.

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 Gibbs sampling method will deviate more from that obtained by the proposed method. Figure 12 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 to zero, the better the reconstruction performance is. It is observed that with insufficient iterations (10 iterations in this example), Gibbs sampling method yields 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.

Approximation of multi missing pattern solution

As discussed in previous sections, the solution of single missing pattern problem provides a viable alternative in the

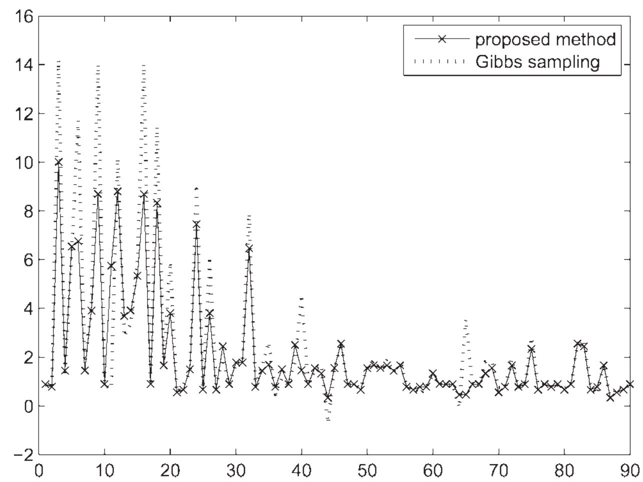

Figure 12. Reconstruction error.

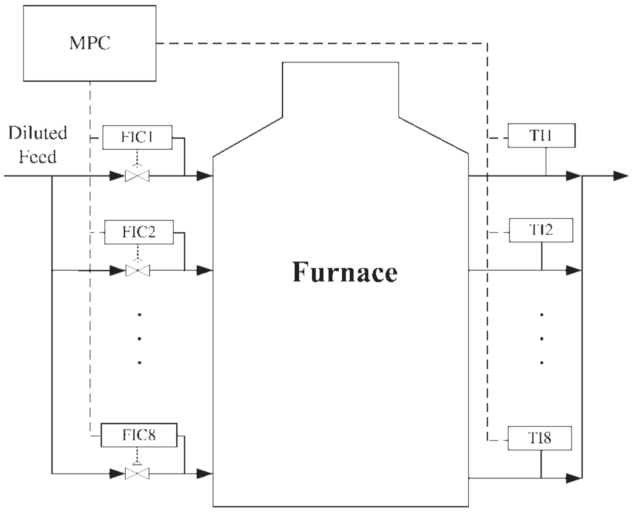

Figure 13. Structure of diluted bitumen preheater.

presence of multi 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 multi 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*"; 90% of π_6 or π_9 are missing 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 singlemissing pattern method, we only need to deal with the singlemissing 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 $(\pi_1,\ldots,\pi_5,\times,\pi_7,\pi_8,\times,\pi_{10},\ldots,\pi_{15})$. Not surprisingly, the diagnosis results by this approximation are the same as those presented in Figures 6 and 7, and it has been shown that they are considerably better than simply ignoring an entire evidence when one of the monitor reading is not available.

Industry Evaluation

The data-driven Bayesian diagnosis approach has been tested on an industry diluted bitumen preheater process.

Process description

The schematic diagram of the process is presented in Figure 13. The function of this process is to heat the diluted bitumen from the upstream process to a desired temperature with a furnace. The diluted bitumen is fed into the furnace through eight passes, and the bitumen is heated within the coils in the furnace. The eight bitumen passes are mixed at the outlet and then fed into downstream process. Flow control of the bitumen 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 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 phenomenons, so as to enhance the stiction detection and control performance monitoring.

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 used to monitor the control performance of the difference between TI4 and COT_{ave}. The model validation monitor 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 1 min. Each window of data consists

Table 7. Summary of Bayesian Diagnosis Parameters

Evidence	$E = (\pi_1, \pi_2, \pi_3)$
Discretization	$k_i = 2, K = 2^3 = 8$
Historical data	41 samples for *NF* mode, 23 samples for *SP* mode, and 8 samples for *sticky* 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 *NF* and *SP* mode, and 2 samples for *sticky* mode

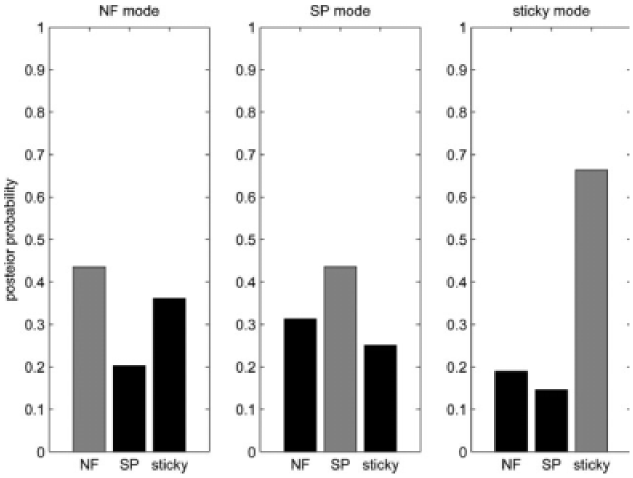

Figure 14. Posterior probability assigned to each mode.

of ∼8-h 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 crossvalidation of the Bayesian diagnostic system. Table 7 summarizes parameters for the Bayesian diagnosis.

With the data-driven Bayesian approach, the diagnosis results shown in Figure 14 are obtained for the crossvalidation data. In each plot, the posterior probability corresponding to the true underlying mode is shown with gray bars, whereas others are in dark bars. Thus, if the gray bar is highest then the correct diagnosis is obtained. From Figure 14, we can see that all the true underlying modes are assigned with the largest probabilities, indicating correct diagnosis of the three modes.

Missing data handling

Furthermore, we 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 data 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,

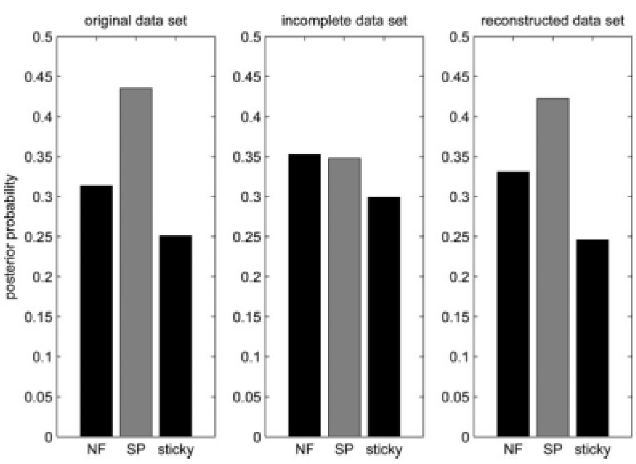

Figure 15. Diagnosis results with different data set.

Table 8. Realizations of Evidences with Different Data Set

	Numbers of Occurrences		
Evidences	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

π_1 is more likely to be unavailable when the control performance is poor. Column 2 of 8 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 data, and so does the likelihood of evidences. If only the samples with complete evidence are used to calculate the likelihood, the diagnosis results of *SP* mode are shown in the middle panel of Figure 15, where the highest probability is not assigned to the true underlying mode.

With the proposed approach, the recovered historical samples are presented in the last column of Table 8. 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 diagnosis results obtained with the recovered data, shown in the right panel of Figure 15, 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.

Conclusions

In this article, a novel data-driven Bayesian approach for control loop diagnosis with missing data is presented. Bayesian methods are used to synthesize control loop monitors and isolate the underlying problem sources. Data missing problem is handled via marginalization of evidences over the UCEM. The proposed method is verified by a simulated binary distillation column and an industry process, where the features of the Bayesian approach to the synthesis of a variety of monitors are demonstrated.

Notation

a_{jlm} = Dirichlet parameter, or prior sample number for the j-th evidence value under mode m
c_i = the i-th component in the control loop
d^t = the t-th historical data sample
$d^t_{m_i}$ = the t-th historical data sample for mode m_i
e_i = the i-th possible discrete evidence value
e^t = evidence of the t-th historical data sample
$\tilde{e}(t)$ = input of nominal disturbance model at time stamp t
k_i = number of discrete outputs of the i-th monitor
m_i = the i-th possible mode
m^t = underlying mode of t-th historical data sample
n_{jlm} = number of historical samples with evidence e_j and mode m
q_i = number of status of the i-th component
$y(t)$ = process output at time stamp t
$\hat{y}(t)$ = simulated process output at time stamp t
A_m = total number of prior sample for mode m
E = evidence
G_l = nominal disturbance model
K = total number of evidence values

M = mode of the control system
N_c = complete historical sample number (implicitly for mode m)
N_m, N = total number of historical samples for mode m
$\tilde{N}$ = total number of historical samples for all modes
P = number of components in the control system
Q = total number of modes
S_k = number of possible incomplete evidences of the k-th UCEM
R_k = number of possible complete evidence corresponding to an incomplete one in the k-th UCEM
$R_{\tilde{e}}$ = variance of $\tilde{e}(t)$
$\mathcal{D}$ = overall historical data set
$\mathcal{D}_c$ = historical data set with complete evidences
$\mathcal{D}_{ic}$ = historical data set with incomplete evidences
$\mathcal{D}_m$ = historical data set for mode m
$\mathcal{D}_{\neg m}$ = historical data set for modes other than m
$\mathcal{E}$ = set of all possible evidences
$\mathcal{M}$ = set of all possible modes
π_i = output of the i-th monitor
Θ_m = likelihood parameters for mode m
$\theta_{e_i|m}$,$\theta_{i|m}$ = likelihood of evidence e_i under mode m
$\theta_{i,j/k}$ = likelihood of evidence $\varepsilon_{s,r}$ in the k-th UCEM
Ω = space of all possible likelihood parameters
$\varepsilon_{i,j/k}$ = complete evidence with location (i,j) in the k-th UCEM
$\eta_{i/k}$ = numbers of historical data samples with evidence ε_i in the k-th UCEM
$\eta_{i,j/k}$ = numbers of historical data samples with evidence $\varepsilon_{i,j}$ in the k-th UCEM
$\eta'_{i,j}$ = expected number of samples with evidence $\varepsilon_{i,j}$ in the incomplete data set
$\varepsilon_{i/k}$ = incomplete evidence corresponding to the complete evidence $\varepsilon_{i,j}$ in the k-th UCEM
$\lambda_{i/k}$ = likelihood for incomplete evidence ε_i in the k-th UCEM
ρ_i^{miss} = data missing ratio for the i-th row evidences in a UCEM

Literature Cited

1. Jelali M. An overview of control performance assessment technology and industrial applications. *Control Eng Pract*. 2007;14:441–466.
2. Huang B, Shah S. *Performance Assessment of Control Loops*. Springer-Verlag, London, UK, 1999.
3. Qin S. Control performance monitoring—a review and assessment. *Comput Chem Eng*. 1998;23:173–186.
4. Harris T, Seppala C, Desborough L. A review of performance monitoring and assessment techniques for univariate and multivariate control systems. *J Process Control*. 1999;9:1–17.
5. Schafer J, Cinar A. Multivariable MPC system performance assessment, monitoring, and diagnosis. *J Process Control*. 2006;14:113–129.
6. Patwardhan R, Shah S. Issues in performance diagnostics of model-based controllers. *J Process Control*. 2002;12:413–427.
7. Thornhill N, Horch A. Advances and new directions in plant-wide controller performance assessment. Proceeding of International Symposium on Advanced Control of Chemical Processes, Gramado, Brazil, 2006.
8. Huang B. Bayesian methods for control loop monitoring and diagnosis. *J Process Control*. 2008;18:826–838.
9. Pernestal A. Ph.D Thesis, KTH School of Electrical Engineering, 2007.
10. Dey S, Stori J. A Bayesian network approach to root cause diagnosis of process variations. *Int J Mach Tools*. 2005;45:75–91.
11. Mehranbod N, Soroush M, Panjapornpon C. A method of sensor fault detection and identification. *J Process Control*. 2005;15:321–339.
12. Steinder M, Sethi A. Probabilistic fault localization in communication systems using belief networks. *IEEE/ACM Trans Network*. 2004;12:809–822.
13. Volk U, Kniese D, Hahn R, Haber R, Schmitz U. Optimized multivariable predictive control of an industrial distillation column considering hard and soft constrains. *Control Eng Pract*. 2004;13:809–822.
14. Ahmed S, Huang B, Shah S. Validation of continuous-time models with delay. *Chem Eng Sci*. 2009;64:443–454.
15. Qin S, Li W. Detection and identification of faulty sensors in dynamic process. *AIChE J* 2001;47:1581–1593.
16. Geman S, Geman D. Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images. *IEEE Trans Pattern Anal Mach Intell*. 1984;6:721–741.
17. Korb K, Nicholson A. *Bayesian Artificial Intelligence*. Chapman & Hall/CRC, Roca Raton, USA, 2004.
18. The BUGS (Bayesian inference Using Gibbs Sampling) project website. Avaliable at: http://www.mrc-bsu.cam.ac.uk/bugs/.

Appendix

Estimation of Evidence Likelihood

Suppose that the likelihood of evidence $\varepsilon_{s,r}$ is about to be calculated. According to Eq. 4,

$$\begin{aligned}
p(\varepsilon_{s,r}|M,\mathcal{D}) &= \int_\Omega p(\varepsilon_{s,r}|\Theta,M,\mathcal{D})f(\Theta|M,\mathcal{D})d\Theta \\
&= \int_\Omega \theta_{s,r}\frac{c}{p(\mathcal{D}|M)}\prod_{i=1}^{S}\left[\prod_{j=1}^{R}\theta_{i,j}^{\eta_{i,j}+a_{i,j}-1}\cdot\left(\sum_{k=1}^{R}\Theta_{i,k}\right)^{\eta_i}\right]d\Theta \\
&= \frac{c}{p(\mathcal{D}|M)}\int_\Omega \theta_{s,r}\prod_{i=1}^{S}\left[\prod_{j=1}^{R}\theta_{i,j}^{\eta_{i,j}+a_{i,j}-1}\cdot\left(\sum_{k=1}^{R}\theta_{i,k}\right)^{\eta_i}\right]d\Theta \\
&= \frac{c}{\int_\Omega p(\mathcal{D}|\theta,M)f(\theta|M)} \\
&\quad \cdot\int_\Omega \theta_{s,r}\prod_{i=1}^{S}\left[\prod_{j=1}^{R}\theta_{i,j}^{\eta_{i,j}+a_{i,j}-1}\cdot\left(\sum_{k=1}^{R}\theta_{i,k}\right)^{\eta_i}\right]d\Theta \\
&= \left(\int_\Omega \theta_{s,r}\prod_{i=1}^{S}\left[\prod_{j=1}^{R}\theta_{i,j}^{\eta_{i,j}+a_{i,j}-1}\cdot\left(\sum_{k=1}^{R}\theta_{i,k}\right)^{\eta_i}\right]d\Theta\right)\Big/ \\
&\quad \left(\int_\Omega \prod_{i=1}^{S}\left[\prod_{j=1}^{R}\theta_{i,j}^{\eta_{i,j}+a_{i,j}-1}\cdot\left(\sum_{k=1}^{R}\theta_{i,k}\right)^{\eta_i}\right]d\Theta\right).
\end{aligned} \tag{A1}$$

The denominator of Eq. A1 is

$$\begin{aligned}
&\int_\Omega \prod_{i=1}^{S}\left[\prod_{j=1}^{R}\theta_{i,j}^{\eta_{i,j}+a_{i,j}-1}\cdot\left(\sum_{k=1}^{R}\theta_{i,k}\right)^{\eta_i}\right]d\Theta \\
&= \int_\Omega \prod_{i=1}^{S}\prod_{j=1}^{R}\theta_{i,j}^{\eta_{i,j}+a_{i,j}-1}\cdot\sum_{t_{i,1},\ldots,t_{i,R}}\binom{\eta_i}{t_{i,1},\ldots,t_{i,R}}\theta_{i,1}^{t_{i,1}}\ldots\theta_{i,R}^{t_{i,R}}d\Theta \\
&= \int_\Omega \prod_{i=1}^{S}\sum_{t_{i,1},\ldots,t_{i,R}}\binom{\eta_i}{t_{i,1},\ldots,t_{i,R}}\cdot\theta_{i,1}^{t_{i,1}+\eta_{i,1}+a_{i,1}-1}\ldots\theta_{i,R}^{t_{i,R}+\eta_{i,R}+a_{i,R}-1}d\Theta,
\end{aligned} \tag{A2}$$

where

$$\begin{aligned}
\binom{\eta_i}{t_{i,1},\ldots,t_{i,R}} &= \binom{t_{i,1}}{t_{i,1}}\binom{t_{i,1}+t_{i,2}}{t_{i,2}}\ldots\binom{t_{i,1}+t_{i,2}+\ldots+t_{i,R}}{t_{R_i}} \\
&= \frac{\eta_i!}{t_{i,1}!t_{i,2}!\ldots t_{i,R}!},
\end{aligned} \tag{A3}$$

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

Similarly, the numerator in Eq. A1 is

$$\int_\Omega \theta_{s,r}\prod_{i=1}^{S}\sum_{t_{i,1},\ldots,t_{i,R}}\binom{\eta_i}{t_{i,1},\ldots,t_{i,R}}\cdot\theta_{i,1}^{t_{i,1}+\eta_{i,1}+a_{i,1}-1}\ldots\theta_{i,R}^{t_{i,R}+\eta_{i,R}+a_{i,R}-1}d\Theta \tag{A4}$$

Denote all sets of $t_{i,1},\ldots,t_{i,R}$ as $\mathcal{T}_i$, combination of all $\mathcal{T}_i$ as $\mathcal{T}$, $\binom{\eta_i}{t_{i,1},\ldots,t_{i,R}}$ as $\mathbf{C}_{\eta_i}^{\mathcal{T}_i}$, and $\eta_{i,j} + a_{i,j} - 1$ as $q_{i,j}$. Then Eq. A2 is

$$\begin{aligned}\int_\Omega \prod_{i=1}^{S}\left[\sum_{\mathcal{T}_i}\mathbf{C}_{\eta_i}^{\mathcal{T}_i}\theta_{i,1}^{t_{i,1}+q_{i,1}}\ldots\theta_{i,R}^{t_{i,R}+q_{i,R}}\right]d\Theta \\ = \sum_{\mathcal{T}}\prod_{i=1}^{S}\mathbf{C}_{\eta_i}^{\mathcal{T}_i}\cdot\int_\Omega\prod_{i=1}^{S}\left[\theta_{i,1}^{t_{i,1}+q_{i,1}}\ldots\theta_{i,R}^{t_{i,R}+q_{i,R}}\right]d\Theta \\ = \sum_{\mathcal{T}}\frac{\prod_{i=1}^{S}\mathbf{C}_{\eta_i}^{\mathcal{T}_i}\prod_{j=1}^{R}\Gamma(t_{i,j}+q_{i,j}+1)}{\Gamma(N+A+1)},\end{aligned} \tag{A5}$$

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, Eq. A4 can be rewritten as

$$\begin{aligned}\int_\Omega \theta_{s,r}\prod_{i=1}^{S}\left[\sum_{\mathcal{T}_i}\mathbf{C}_{\eta_i}^{\mathcal{T}_i}\theta_{i,1}^{t_{i,1}+q_{i,1}}\ldots\theta_{i,R}^{t_{i,R}+q_{i,R}}\right]d\Theta \\ = \sum_{\mathcal{T}}\frac{\prod_{i=1}^{S}\mathbf{C}_{\eta_i}^{\mathcal{T}_i}\Gamma(t_{s,r}+q_{s,r}+2)\prod_{j\neq r}\Gamma(t_{i,j}+q_{i,j}+1)}{\Gamma(N+A+1)}\end{aligned} \tag{A6}$$

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

$$\begin{aligned}p(\varepsilon_{s,r}|M,\mathcal{D}) \\ = \frac{1}{N+A}\cdot\frac{\sum_{\mathcal{T}}\prod_{i=1}^{S}\mathbf{C}_{\eta_i}^{\mathcal{T}_i}(t_{s,r}+q_{s,r}+1)!\prod_{j\neq r}(t_{i,j}+q_{i,j})!}{\sum_{\mathcal{T}}\prod_{i=1}^{S}\mathbf{C}_{\eta_i}^{\mathcal{T}_i}\prod_{j=1}^{R}(t_{i,j}+q_{i,j})!}.\end{aligned} \tag{A7}$$

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 $t_{i,1},\ldots,t_{i,R}$ are independent for different rows i, and Eq. A7 can be simplified to

$$\begin{aligned}p(\varepsilon_{s,r}|M,\mathcal{D}) \\ = \frac{1}{N+A}\cdot\frac{\sum_{\mathcal{T}_s}\mathbf{C}_{\eta_s}^{\mathcal{T}_s}(t_{s,r}+q_{s,r}+1)!\prod_{j\neq r}(t_{s,j}+q_{s,j})!}{\sum_{\mathcal{T}_s}\mathbf{C}_{\eta_s}^{\mathcal{T}_s}\prod_{j=1}^{R}(t_{s,j}+q_{s,j})!} \\ \cdot\frac{\sum_{\mathcal{T}\backslash\mathcal{T}_s}\prod_{i\neq s}\mathbf{C}_{\eta_i}^{\mathcal{T}_i}\prod_{j=1}^{R}(t_{s,j}+q_{s,j})!}{\sum_{\mathcal{T}\backslash\mathcal{T}_s}\prod_{i\neq s}\mathbf{C}_{\eta_i}^{\mathcal{T}_i}\prod_{j=1}^{R}(t_{s,j}+q_{s,j})!} \\ = \frac{1}{N+A}\cdot\frac{\sum_{\mathcal{T}_s}\mathbf{C}_{\eta_s}^{\mathcal{T}_s}(t_{s,r}+q_{s,r}+1)!\prod_{j\neq r}(t_{s,j}+q_{s,j})!}{\sum_{\mathcal{T}_s}\mathbf{C}_{\eta_s}^{\mathcal{T}_s}\prod_{j=1}^{R}(t_{s,j}+q_{s,j})!},\end{aligned} \tag{A8}$$

where $\mathcal{T}\backslash\mathcal{T}_s$ 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 $t_{s,1},\ldots,t_{s,R}$ to evidences $\varepsilon_{s,1},\ldots,\varepsilon_{s,R}$ 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 later.

Lemma 1.

$$\sum_{\mathcal{T}_s}\binom{\eta_s}{t_1,\ldots,t_R}\prod_{j=1}^{R}(x_j+t_j)! = \prod_{k=1}^{R}x_k!\cdot\prod_{i=0}^{\eta_s-1}\left(\sum_{j=1}^{R}x_j+i+R\right) \tag{A9}$$

See appendix for details of the proof.

With Lemma 1, Eq. A8 can be further simplified. The denominator is

$$\begin{aligned}(N+A)\sum_{\mathcal{T}_s}\binom{\eta_s}{t_1,\ldots,t_R}\prod_{j=1}^{R}(t_j+\eta_{s,j}+a_{s,j}-1)! \\ = (N+A)\prod_{k=1}^{R}(\eta_{s,k}+a_{s,k}-1)!\cdot\prod_{i=0}^{\eta_s-1}\left(\sum_{j=1}^{R}[\eta_{s,j}+a_{s,j}]+i\right),\end{aligned} \tag{A10}$$

and the numerator is

$$\begin{aligned}\sum_{\mathcal{T}_s}\binom{\eta_s}{t_1,\ldots,t_R}(t_r+\eta_{s,r}+a_{s,r})!\cdot\prod_{j\neq r}(t_j+\eta_{s,j}+a_{s,j}-1)! \\ = (\eta_{s,r}+a_{s,r})!\prod_{k\neq r}(\eta_{s,k}+a_{s,k}-1)!\cdot\prod_{i=0}^{\eta_s-1}\left(\sum_{j=1}^{R}[\eta_{s,j}+a_{s,j}]+i+1\right)\end{aligned} \tag{A11}$$

By substituting Eqs. A10 and A11 in Eq. A8, the following result is obtained for the likelihood of evidence $\varepsilon_{s,r}$,

$$\begin{aligned}p(\varepsilon_{s,r}|M,\mathcal{D}) &= \frac{\eta_{s,r}+a_{s,r}}{N+A}\cdot\frac{\sum_{j=1}^{R}(\eta_{s,j}+a_{s,j})+\eta_s}{\sum_{j=1}^{R}(\eta_{s,j}+a_{s,j})} \\ &= \frac{\eta_{s,r}+a_{s,r}}{N+A}\cdot\left(1+\frac{\eta_s}{\sum_{j=1}^{R}(\eta_{s,j}+a_{s,j})}\right)\end{aligned} \tag{A12}$$

Proof of Lemma 1

$$\sum_{\mathcal{T}_s}\binom{\eta_s}{t_1,\ldots,t_R}\prod_{j=1}^{R}(x_j+t_j)! = \prod_{k=1}^{R}x_k!\prod_{i=0}^{n_s-1}\left(\sum_{j=1}^{R}x_j+i+R\right) \tag{A13}$$

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.

$$\text{RHS} = \prod_{j=1}^{2}x_j!\prod_{i=0}^{0}\left(\sum_{j=1}^{2}x_j+i+2\right) = x_1!x_2!(x_1+x_2+2) \tag{A14}$$

$$\text{LHS} = \sum_{\mathcal{T}} \binom{1}{t_1, t_2} \prod_{j=1}^{2} (x_j + t_j)! = (x_1 + 1)!x_2! + x_1!(x_2 + 1)!$$
$$= x_1!x_2!(x_1 + x_2 + 2) = \text{RHS} \tag{A15}$$

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

Induction on R

Note that Eq. A13 can be written as

$$\sum_{\mathcal{T}_s} \binom{\eta}{t_1, \ldots, t_R} \prod_{j=1}^{R} \prod_{r=1}^{t_j} (x_j + r)! = \prod_{i=0}^{n_s - 1} \left(\sum_{j=1}^{R} x_j + i + R \right). \tag{A16}$$

Thus,

$$\begin{aligned} &\prod_{i=0}^{n_s-1} \left(\sum_{j=1}^{R+1} x_j + i + R + 1 \right) \\ &= \prod_{i=0}^{n_s-1} \left(\sum_{j=1}^{R-1} (x_j + 1) + [(x_R + x_{R+1} + 1) + 1] + i \right) \\ &= \sum_{\mathcal{T}_s} \binom{\eta_s}{t_1, \ldots, t_{R-1}, T} \prod_{j=1}^{R-1} \prod_{r=1}^{t_j} (x_j + r) \\ &\quad \cdot (x_R + x_{R+1} + 1 + 1) \ldots (x_R + x_{R+1} + 1 + T) \\ &= \sum_{\mathcal{T}_s} \binom{\eta_s}{t_1, \ldots, t_{R-1}, T} \prod_{j=1}^{R-1} \prod_{r=1}^{t_j} (x_j + r) \\ &\quad \cdot (x_R + 1 + x_{R+1} + 1) \ldots (x_R + 1 + x_{R+1} + 1 + T - 1) \\ &= \sum_{\mathcal{T}_s} \binom{\eta_s}{t_1, \ldots, t_{R-1}, T} \prod_{j=1}^{R-1} \prod_{r=1}^{t_j} (x_j + r) \\ &\quad \cdot \binom{T}{t_R, t_{R+1}} \prod_{j=R}^{R+1} \prod_{r=1}^{t_j} (x_R + r) \end{aligned} \tag{A17}$$

In Eq. A17,

$$\binom{\eta_s}{t_1, \ldots, t_{R-1}, T} \cdot \binom{T}{t_R, t_{R+1}} = \frac{n_s}{t_1! \ldots t_{R-1}! T!} \cdot \frac{T!}{t_R! t_{R+1}!} = \binom{\eta_s}{t_1, \ldots, t_{R+1}}, \tag{A18}$$

so

$$\prod_{i=0}^{n_s-1} \left(\sum_{j=1}^{R+1} x_j + i + R + 1 \right) = \sum_{\mathcal{T}_s} \binom{\eta_s}{t_1, \ldots, t_{R+1}} \prod_{j=1}^{R+1} \prod_{r=1}^{t_j} (x_j + r). \tag{A19}$$

By multiplying Eq. A19 with $\prod_{j=1}^{R+1} x_i!$, we can get

$$\sum_{\mathcal{T}_s} \binom{\eta_s}{t_1, \ldots, t_{R+1}} \prod_{j=1}^{R+1} (x_j + t_j)! = \prod_{j=1}^{R+1} x_j! \prod_{i=0}^{n_s-1} \left(\sum_{j=1}^{R+1} x_j + i + R + 1 \right). \tag{A20}$$

Induction on $\mathbf{n_s}$

RHS of eq. A16 is

$$\begin{aligned} &\prod_{i=0}^{n_s} \left(\sum_{j=1}^{R} x_j + i + R \right) \\ &= \prod_{i=0}^{n_s-1} \left(\sum_{j=1}^{R} x_j + i + R \right) \cdot \left(\sum_{j=1}^{R} x_j + n_s + R \right) \\ &= \sum_{\mathcal{T}_s} \binom{\eta_s}{t_1, \ldots, t_R} \prod_{j=1}^{R} (x_j + t_j)! \cdot \left(\sum_{j=1}^{R} x_j + n_s + R \right) \\ &= \sum_{\mathcal{T}_s} \binom{\eta_s}{t_1, \ldots, t_R} \prod_{j=1}^{R} (x_j + t_j)! \sum_{k=1}^{R} (x_k + 1 + t_k) \\ &= \sum_{k=1}^{R} \sum_{\mathcal{T}_s} \binom{\eta_s}{t_1, \ldots, t_R} \prod_{j=1}^{R} (x_j + t_j)!(x_k + t_k + 1) \end{aligned} \tag{A21}$$

Following the multinomial theorem, which indicates that

$$\begin{aligned} &(x_1 + x_2 + \ldots + x_m)^{n+1} \\ &\quad = \sum_{\mathcal{T}_s} \binom{n}{t_1, \ldots, t_m} x_1^{t_1} \ldots x_m^{t_m} \cdot (x_1 + \ldots + x_m) \\ &\quad = \sum_{j=1}^{m} \sum_{\mathcal{T}_s} \binom{n}{t_1, \ldots, t_m} x_1^{t_1} \ldots x_m^{t_m} x_j \\ &\quad = \sum_{\mathcal{T}_s} \binom{n}{t_1, \ldots, t_m} x_1^{t_1} \ldots x_m^{t_m}, \end{aligned} \tag{A22}$$

Equation A21 can be written as

$$\begin{aligned} &\sum_{k=1}^{R} \sum_{\mathcal{T}_s} \binom{\eta_s}{t_1, \ldots, t_R} \prod_{j=1}^{R} (x_j + t_j)!(x_k + t_k + 1) \\ &\qquad = \sum_{\mathcal{T}_s} \binom{\eta_s}{t_1, \ldots, t_R} \prod_{j=1}^{R} (x_j + t_j)! \end{aligned} \tag{A23}$$

By multiplying Eq. A19 with $\prod_{k=1}^{R+1} x_k!$, we can get

$$\sum_{\mathcal{T}_s} \binom{\eta_s + 1}{t_1, \ldots, t_R} \prod_{j=1}^{R} (x_j + t_j)! = \prod_{k=1}^{R} x_k! \cdot \prod_{i=0}^{n_s} \left(\sum_{j=1}^{R+1} x_j + i + R + 1 \right), \tag{A24}$$

which completes the induction and the proof.

Manuscript received Sept. 20, 2008, revision received Feb. 19, 2009, and final revision received May 8, 2009.