8th International IFAC Symposium on
Dynamics and Control of Process Systems
Preprints Vol.1, June 6-8, 2007, Cancún, Mexico

BAYESIAN METHODS FOR CONTROL LOOP MONITORING AND DIAGNOSIS

Biao Huang *,[1]

* *Department of Chemical and Materials Engineering, University of Alberta, Edmonton, AB T6G 2G6, Canada*

Abstract: There exist many algorithms for control performance monitoring. There are also many algorithms available for process monitoring. There are, however, few methods available for synthesis of various monitoring technologies to form a diagnosing system for optimal decision making. This paper is concerned with establishing and demonstrating a novel probabilistic diagnostic framework for control loop monitoring. The new framework possesses a number of desired properties including, for example, probabilistic diagnosing procedure, flexibility in synthesizing different monitoring technologies, robustness in the presence of missing data or missing variables, ease of expansion or shrinking of the diagnosing system, ability to incorporate *a priori* process knowledge, and capability for decision making. As the backbone of the proposed framework, the emerging Bayesian methods are introduced and shown to be the appropriate tools. Several representative control loop diagnostic problems are formulated under the Bayesian framework and their solutions are demonstrated through examples. The experiences and challenges learned from industrial applications of Bayesian methods are summarized and some of future research directions are discussed.

Keywords: Bayesian methods, performance assessment, diagnosis, control monitoring, process monitoring, Bayesian network

1. INTRODUCTION

Established in later 1980s by [Harris, 1989], the research on control loop monitoring has been and remains to be one of the most active research areas in process control community. A number of algorithms have been developed and many successful applications have been reported; see reviews by [Qin, 1998; Huang and Shah, 1999; Harris et al., 1999; Jelali, 2006]. It is estimated that several hundreds of papers have published in this or related direction [Horch and Dumont, 2003]. On the practical side, Eastman Kodak recently reported regular loop monitoring on over 14000 PID loops. Some commercial control performance assessment software including multivariate performance assessment has also been available in the market.

While the significant progress is being made in control performance monitoring, other related developments also go hand in hand including, sensor monitoring [Jelali, 2006], actuator monitoring [Horch, 2000], oscillation detection [Thornhill and Hagglund, 1997; Horch, 2000], model validation [Huang et al., 2003], model predictive control monitoring [Kesavan and Lee, 1997; Schafer and Cinar, 2004], nonlinearity detection [Choudhury et al., 2004] etc. Most of monitoring methods target specific problem in a control loop and suc-

[1] Corresponding author; Tel: +1(780)492-9016, Fax: +1(780)492-2881, E-mail: biao.huang@ualberta.ca

cessful case studies have been reported. However, the synthesis of these monitoring technologies has rarely been addressed. The common practice was that one monitoring algorithm was developed for specific problem and then tested with the targeted problem. Obviously, different problems can result in similar symptoms and may affect more than one monitoring algorithms. There is a need to consider various monitoring problems simultaneously in a systematic manner.

There are a number of challenging issues in synthesizing monitoring problems: 1) While the source of the problem may be unique (e.g. change of disturbance dynamics), its symptoms can be similar to that resulted from different problem sources (e.g. change of the plant model). The model validation algorithm that is designed to detect plant-model mismatch can not immune from disturbance model change. Thus, while each monitoring algorithm may work well when only the targeted problem occurs, relying on a single monitoring algorithm can be misleading in general. For example, change of the disturbance model may be diagnosed as change of the plant model if only the plant model validation algorithm is applied. 2) To resolve this issue, one needs to investigate how problem sources can affect each others' monitoring algorithm. Probabilistic dependence of each monitoring algorithm on all problem sources has to be investigated in order to achieve this objective. 3) All processes operate, to a certain degree, in an uncertain world. The occurrence of a problem, its symptom, and its interconnection with other problems/symptoms all have some uncertainties. A solution has to be built upon a probabilistic framework. Thus, a joint probability distribution among all problem sources as well as all symptoms detected need to be established. As elaborated shortly, the computation of the probabilities and statistical inferences grows exponentially with the number of problem sources and observed symptoms. 4) Most of the existing monitoring methods are data based. While they have advantages of simplicity, it is obvious that certain *a priori* knowledge of process is not only helpful but also necessary when multiple problem sources are considered. For example, a process flow chart indicating components interrelationship may be available and should be considered in making a meaningful diagnosis and decision. It is of a considerable challenge to integrate data based algorithms with *a priori* knowledge.

In view of problems raised above, a novel probabilistic framework to synthesize various monitoring technologies for control loop and instrument monitoring/diagnosis is developed in this paper. A Bayesian model also known as graphical model [Murphy, 2002], elaborated shortly, is a suitable framework for solving such problems.

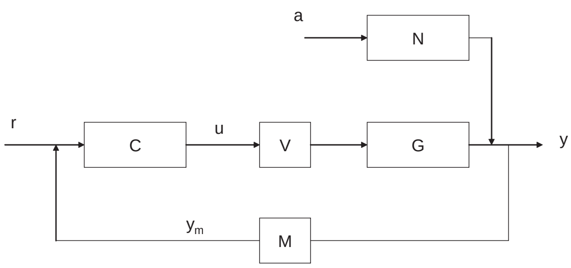

Fig. 1. Block diagram of control loop

2. PROBLEM STATEMENT

An example of the processes to be consdered is shown in Fig.1, where C, V, G, N, M represent the controller, actuator, plant, disturbance, and sensor, respectively. For illustration, we make the following assumptions:

- The control law C needs not to be known and may be a PID or an optimization based constrained control law such as model predictive control.
- Noise model N is varying among a set of models.
- The nominal plant model G is unknown except for the case of model validation.
- Both actuator V and sensor M are subject to fault (or problem). For example, the actuator may be subject to stiction and sensor may be subject to bias error.

Control loop performance is the primary interest of this work and it can vary due to various reasons such as change in the plant, in the controller, and/or in disturbance dynamics, or simply due to faults in the sensors or actuators.

We further assume that certain monitors are available. Every monitor is, however, subject to disturbances and thus false alarms, and each monitor can be sensitive to other faults that are not supposed for this monitor to look after. For example, the model monitor may be sensitive to valve nonlinearity although it is designed to monitor the plant model only.

3. ESTABLISHING A NOVEL FRAMEWORK FOR CONTROL LOOP DIAGNOSIS

As has been discussed in the last section, a typical control loop consists of at least four components, sensor, actuator, controller and plant, each subject to possible performance degradation or fault. Any problem in one of these four components can affect control loop performance. Each of them has its monitoring algorithms to monitor the problems and these algorithms may all be affected by one or more of the four components. Imagine a simplest network of eight nodes, representing four components and four monitors, and their relations are described by conditional probabilities. To completely determine the relation among all nodes, one needs to know the joint probabilities of *eight* random variables. With increased components to be considered and monitors to be added, the complexity of the network can quickly go beyond computational possibility. The emerging Bayesian garphic model (or simply called Bayesian model in the sequel), which exploits the independence of random variable network, is the choice of the method that sheds a light on the problem solution.

The building block of the Bayesian model is a network of nodes connected by conditional probabilities. These nodes are random variables, which can be continuous, discrete or even binary. Consider the simplest binary random variables. If there are n binary random variables, the complete distribution is specified by $2^n - 1$ joint probabilities. In the illustrative Fig. 2, there are 4 binary nodes, each node having two possible outcomes. For example, node A may take the value A or $\bar{A}$. To completely determine the distribution of the 4 binary variables, one needs to determine joint probability $P(A, B, C, D)$ that has 16 outcomes. By taking account that sum of all probabilities must equal to 1, one needs to calculate 15 probabilities. However, as illustrated in Fig. 2, by exploring the relationship of each node, only 7 probabilities need to be determined, a considerable reduction from 15. The saving comes from exploiting the conditional independence between certain variables (i.e. no arcs between certain variables). The structure of the graphic relationship is also the example of incorporating the *a priori* process knowledge so that the conditional dependence/independence between certain variables is fully utilized. With the increase of the nodes, the saving of computations is exponential, making it possible to apply Bayesian inference theory in practice.

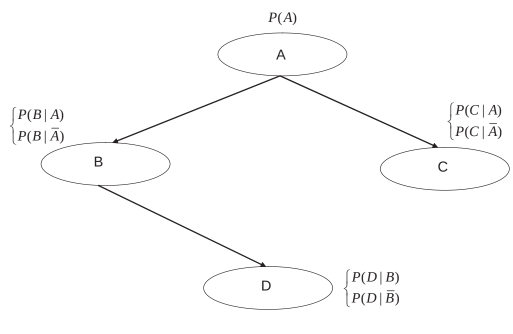

Fig. 2. An example of Bayesian graphic model

If a Bayesian model like the one shown in Fig. 2 is available, one can make a variety of inferences. For example, if we have the observations of B, C, D, we would like to make an inference to determine whether A=A or A=$\bar{A}$. The decision process can be written under Bayesian formulation as $P(A|BCD)$ that can be calculated, according to Bayes theorem

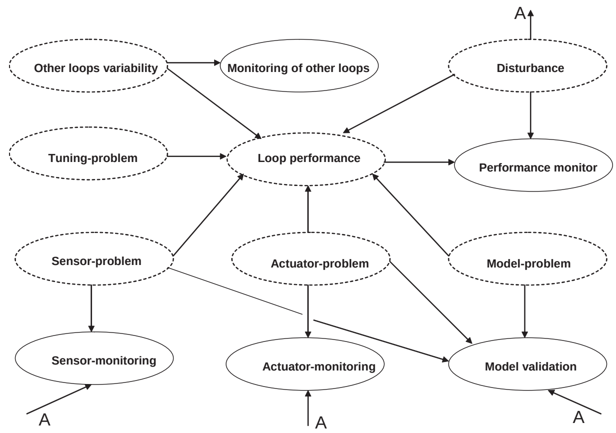

Fig. 3. A control loop monitoring and diagnosis architecture

$$P(A|BCD) = \frac{P(ABCD)}{P(BCD)} = \frac{P(ABCD)}{\sum_A P(ABCD)}$$

Using the structured chain rule of the Bayesian inference [Korb and Nicholson, 2003] according to relationship of four nodes in Fig. 2, the joint probability can be calculated as

$$P(ABCD) = P(A)P(C|A)P(D|B)P(B|A)$$

The seven probabilities specified in Fig. 2 are sufficient to calculate $P(A|BCD)$ to make an inference about the state of A. To show the flexibility of the Bayesian modeling approach, we consider the inference of C given observation D where in this case C is hidden node (node that can not be observed directly). Both A and B are also unobserved (or missing measurements). According to Bayes theorem, it can be derived that

$$P(C|D) = \frac{\sum_{AB} P(ABCD)}{\sum_{ABC} P(ABCD)}$$

That is to say, even though C is correlated with B and D, and in fact directly depends on A, one can still make an inference about C from the evidence in D in the absence of A and B.

A general monitoring and diagnosis architecture for the four components discussed so far is shown in Fig.3. In this figure, the solid nodes are the evidence (observation) nodes while the dashed nodes are hidden nodes (variables to be inferred). Each of the four components has its own (direct) monitor except for the controller (tuning) component. The control performance monitor is the monitor of the whole control loop performance rather than the control tuning itself, which is therefore affected by all four components. All other monitors may be affected by one or more other components too. Disturbance affects all monitors as expected, thus introducing false alarms. In addition, variability from other interacting loops can also have an effect on the loop of concern. There are monitors for other control loops as well.

The scope of this paper is to establish and demonstrate a novel diagnostic framework for solving problems stated above, namely a probabilistic framework to synthesize various existing monitors for control loop monitoring and diagnosis. The development of individual monitors has been well addressed in the literature and will not be considered here. While the main objective is to build a new framework for control loop diagnosis, specifically we will consider several representative problems of common interests. They include control loop performance problem, process model mismatch problem, actuator problem, sensor problem, and model predictive control performance problem. The task of Bayesian inference is to isolate problem source(s) through probabilistic synthesis of all or partial readings of the monitors and then make optimal diagnosis and decisions.

4. BAYESIAN METHODS FOR CONTROL LOOP MONITORING AND DIAGNOSIS

4.1 Direct Bayesian inference: sensor problem diagnosis

To start Bayesian approach for diagnosis, consider a simple yet illustrative sensor fault diagnosis Bayesian model shown in Fig.4. The process is subject to the change of gain, change of input signal, sensor bias, and change of variance in the measurement disturbance (varying variance). The sensor reading is modeled by

$$y = Ku + f + e$$

where y is the sensor reading; process gain K takes two values: 1 and 0 corresponding to normal and abnormal operation, respectively; input u takes three different values -1, 0, and 1; sensor bias f takes two values: 0 and 1 corresponding to bias and non-bias, respectively; noise e has the following distribution

$$e \sim N(0, \sigma^2) \qquad (1)$$

where variance σ^2 also takes two values: 1 and 2 representing normal and abnormal sensor noise, respectively. The graph of the Bayesian model shown in Fig.4 is built using Netica^{TM}. The node through which the arc originates is called the parent node and the node where it terminates is called the child node. The node without parents is also called root nodes.

The diagnosis process is triggered by sensor readings, which are then synthesized with prior (unconditional probability) of each root nodes, together with conditional probability distributions of each child nodes of Fig.4. The prior of each root node can be determined from performance of the equipments (e.g. tendency to fault) or simply from

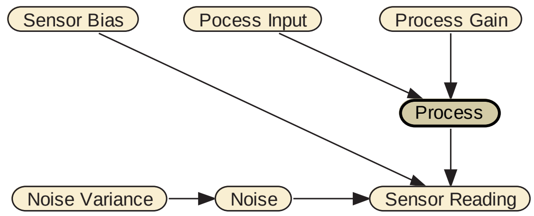

Fig. 4. Sensor monitor structure

historical data. The conditional probability distribution (the sensor reading node) is built according to eqn.(1). The intermediate node, Process node, is a function node, meaning that it is a function of its parent nodes only, and completely determined by its parents. Following the structured chain rule [Korb and Nicholson, 2003], the joint distribution among remaining six random variables, namely sensor bias, process input, process gain, noise variance, noise, and sensor reading, can be established. Through Bayes theorem, probabilistic inferences can be made. Many software packages can be used to perform the inference calculations.

An instance of probabilistic diagnosis is considered here. For a given time instance, there are two readings (evidences) available, namely the input reading and sensor output reading. They are $u = -1$ and $y = 0.7$, respectively. The following posterior (probabilities conditioned on the measurements) can be calculated: $P(Sensor = Bias|E) = 97\%$, $P(ProcessGain = Abnormal|E) = 98.7\%$, $P(NoiseVariance = normal|E) = 74.1\%$, where E represents all evidences (measurements) available. Thus, one can conclude with high confidence (no less than 97% probability) that 1) the sensor has bias and 2) Process Gain is abnormal. One can also conclude that noise variance is likely (with 74.1% probability) to be normal. This example, although relatively simple, indicates the power of Bayesian methods in synthesizing uncertain variables. The distinguished feature of the Bayesian methods relative to other existing diagnosis methods is the explicit quantification of the probability of faults.

4.2 Probabilistic synthesis and missing data handling: valve monitoring and diagnosis

In this section, we will show how information that appears to have some redundancy on one hand or remotely related on the other hand can be synthesized using Bayesian methods to produce a meaningful and reliable diagnosis. It is also shown how Bayesian methods can handle missing data or missing variables in a very natural manner. Valve diagnosis problem will be used as the example.

Actuator fault detection and isolation (FDI) problem has been well studied in the literature. It is well known that sensor nonlinearity can affect control performance and may cause oscillations of control loops. Many algorithms have been developed over the last few years for the detection of valve nonlinearity [Horch, 2000]. Valve monitoring has also been shown to be challenging owing to the nonlinear nature of the problem; thus valve monitors are often subject to numerous false alarms. In this section, it will be shown how the Bayesian methods can be utilized to improve valve monitoring/diagnosis. Any monitor that is available in the literature may be adopted as the valve monitor for this study.

Since valve nonlinearity affects control performance, it should be reflected in certain performance measures (signatures) of control loops. In this section, it will be explored how control performance signatures can be used to improve performance of valve diagnosis although they may appear remotely related. The signatures of choice include H_2 and H_∞ norms of the closed-loop process, which can be calculated following time series analysis of the closed-loop routine operating data [Huang and Shah, 1999].

The process considered is a 2×2 Wood-Berry column model with two interacting loops controlled by two PI controllers, respectively. The process model is

$$G = \begin{bmatrix} \frac{12.8e^{-s}}{6.7s+1} & \frac{-18.9e^{-2.9s}}{21s+1} \\ \frac{6.6e^{-7s}}{10.9s+1} & \frac{-19.4e^{-2.9s}}{14.4s+1} \end{bmatrix}$$

Both loops are subject to backlash valve nonlinearity. There are two types of disturbances affecting the loops; one of them is integrating disturbance (random walk) and the other one non-integrating. Which one is acting on the process is not known during the monitoring process. The two disturbance models are given respectively by

$$N_1 = \begin{bmatrix} \frac{7.6e^{-8.1}}{14.9s+1} & \frac{0.22e^{-7.7s}}{22.8s+1} \\ \frac{9.8e^{-3.4s}}{13.2s+1} & \frac{0.14e^{-9.2s}}{12.1s+1} \end{bmatrix}$$

$$N_2 = \begin{bmatrix} \frac{7.6e^{-8.1s}}{s} & \frac{0.22e^{-7.7s}}{22.8s+1} \\ \frac{9.8e^{-3.4s}}{13.2s+1} & \frac{0.14e^{-9.2s}}{s} \end{bmatrix}$$

For this 2×2 process under PID control, three values of each control performance signature may be calculated, two for the two individual outputs and one for the overall performance in multivariate sense. They are calculated according to the following procedure: First, two univariate time series models are estimated based on two output data sets, $y_1(t)$, $y_2(t)$, respectively, and two H_2's and two H_∞'s are calculated according to the estimated models. This procedure therefore results in two H_2 measures, $H2y1$ and $H2y2$; two H_∞ measures, $Hinfy1$ and $Hinfy2$. Meanwhile, a multivariate time series model can also be obtained from two output time series and MIMO H_2 and H_∞ can also be calculated, namely one MIMO H_2 measure, $H2MIMO$, and one MIMO H_∞ measure, $HinfMIMO$. There seems some redundancy between MIMO and SISO performance measures. However, Bayesian methods are tools naturally dealing with redundancy and "squeeze" the essential information from the redundancy. The Bayesian model is built in Fig.5.

The valve monitors chosen have the following performance:

Actual Status	P(Detected)	P(Not detected)
Backlash	0.70	0.30
No Backlash	0.50	0.50

It is obvious the two valve monitors have a high false alarm rate, 50%, indicating a need for improvement before it can have meaningful applications. The prior for valve 1 to have backlash is about 30% while that for valve 2 is 25%. The priors are known from the knowledge of the valve such as its track record of reliability or its age of usage. It may also be estimated from historical data from the same brand of valves.

In this simulation, there is backlash in valve 1 and no backlash in valve 2. An instance of diagnosis is considered where valve 1 monitor indicates backlash (which triggers the diagnosis), and valve 2 monitor indicates no backlash. Other signatures are not used yet. Due to the high false alarm rate, the conclusion of valve 1 being backlash can not be made. This can be seen with Bayesian inference, by which the calculated posterior of valve 1 being backlash is only 36.5% (owing to small prior of backlash and large false alarm rate) and the posterior of valve 2 being no backlash is 83.6%. Therefore, with only the valve monitors as evidences, the diagnosis conclusion of valve 1 being backlash can not be made. When all signatures are used however, the posterior can be updated. The additional evidences are $H2MIMO = 8.6$, $HinfMIMO = 103.1$, $H2y1 = 2.2$, $Hinfy1 = 37$, $H2y2 = 4.6$, $Hinfy2 = 5.3$. With these additional evidences, the posterior of valve 1 being backlash is now about 100% (even though it has small prior of backlash) and the posterior of valve 2 being no backlash 100%, indicating definitely a problem in valve 1 and definitely no problem in valve 2.

The Bayesian inference is also robust in the presence of missing data or missing variables. When three of the six signatures, $HinfMIMO$, $H2y1$, and $H2y2$, are no longer available (missing), the new calculated posterior of valve 1 being backlash is 81% and valve 2 being no backlash is 89.3%. The fact that the Bayesian inference can handle missing data and missing variables very naturally has a

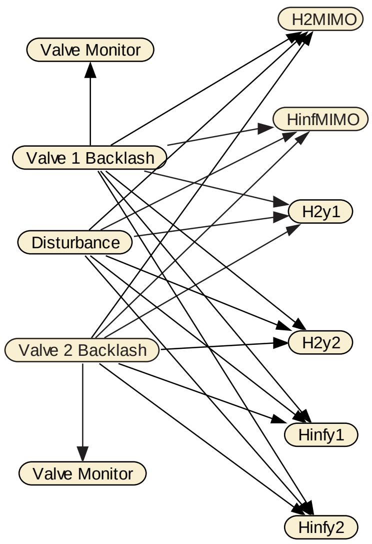

Fig. 5. Bayesian model for valve backlash diagnosis.

great practical significance. Most of monitoring or diagnosis algorithms fail if certain variables are no longer available or certain data are missing. *The Bayesian methods exploit all available information to achieve a best inference result.*

4.3 Synthesizing monitors system: control loop performance diagnosis

In this section, we consider synthesis of various monitors to perform control loop performance diagnosis using Bayesian methods. The research on control loop performance monitoring has been and remains to be one of the most active research areas in process control community. Many methods for control performance monitoring have been established [Harris, 1989; Huang and Shah, 1999; McNabb and Qin, 2005]. One common feature of the existing methods is that they monitor performance of the overall control loop, which can be influenced by all components of the loops. In this sense, they are monitors of loop performance rather than just the controller itself.

Consider a control loop performance diagnosis platform that consists of the valve monitor, model monitor (or model validation), sensor monitor, and control performance monitor. Because the process models include plant model and disturbance model, two model monitors are needed in order to distinguish which of the two models has changed. Two model validation algorithms are adopted here. The local approach with output error (OE) formulation only detects plant model change, independent of disturbance model changes [Huang, 2000]. The local approach with prediction error model (PEM) formulation detects changes of both plant model and disturbance model [Huang et al., 2003]. The control performance monitor can be index-based monitor with minimum variance control benchmark, LQG benchmark [Huang and Shah, 1999] or other suitable benchmarks [Qin, 1998; Harris et al., 1999; Jelali, 2006]. Valve monitor can be any of existing monitors available in the literature. All of the monitors use mean centered data except for the bias monitor; thus they are insensitive to sensor bias. To detect sensor bias, the original data are used in the model validation algorithm. In this case, sensor bias error becomes an additive fault in the model validation problem. Thus the sensor bias monitor is in fact the model validation algorithm for detecting additive fault [Basseville and Nikiforov, 1993]. All monitors but the OE based model validation algorithm are sensitive to the change of disturbance models. The Bayesian model of control loop diagnosis is shown in Fig.6.

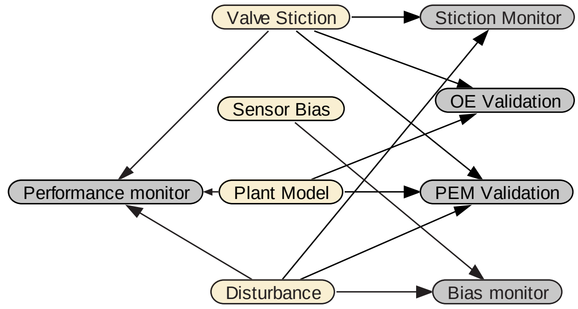

Fig. 6. Bayesian framework for control loop performance diagnosis

An instance of control loop diagnosis is considered here. The outcome of the performance monitor is descretized into only three values, namely optimal, good, and bad. The evidences in this instance are: Performance monitor indicates 'bad' performance that actually triggers this diagnosis procedure; no stiction is detected; data fail both OE model validation test and PEM model validation test; bias is not detected. The posterior through Bayesian inference indicates that the poor control loop performance is most likely induced by plant model change with probability 88.9%. The conditional probabilities for other problems are less than 30%. The conditional probability of having non-integrating disturbance is 70.5%.

4.4 Dynamic Bayesian methods

Our discussions up to now are limited to the static Bayesian model and inference. The evolution of the variables with time has not been considered, i.e. the diagnosis is conducted instance by instance without considering possible temporal relations. For example, the fact that, a healthy equipment has more tendency to have normal operation than an aged equipment in the next sampling instance if it is normal at the current time, has not been considered. The dynamic relation can be utilized to improve the reliability of the diagnosis.

Let $o_1, o_2, \ldots, o_n$ be dynamic observations, and $h_1, h_2, \ldots, h_n$ be the corresponding hidden nodes over the same time window. Define $O_n := \{o_1, o_2, \ldots, o_n\}$ and $H_n := \{h_1, h_2, \ldots, h_n\}$. A time evolution state space model can be written as

$$p(h_n|h_{n-1}) = f(h_{n-1}, n)$$
$$p(o_n|h_n) = g(h_n, n)$$

where $f(.)$ and $g(.)$ are probability distribution functions.

As an example, a special case of this class of models is the Hidden Markov Model [Rabiner, 1989] (HMM) where the model can be simplified to

$$p(h_n|h_{n-1}) = A$$
$$p(o_n|h_n) = \Pi$$

where A and Π are both a probability matrix.

Now denote $p(h_n, O_n)$ as joint probability function of h_n and O_n. Following the approach of [Smyth, 1994], the following recursive algorithm for dynamic Bayesian inference can be derived:

$$p(h_n, O(n)) = \int_{h_{n-1}} p(h_n, h_{n-1}, O_n) dh_{n-1}$$
$$= \int_{h_{n-1}} p(h_n, h_{n-1}, o_n, O_{n-1}) dh_{n-1}$$
$$= \int_{h_{n-1}} p(o_n|h_n, h_{n-1}, O_{n-1}) p(h_n, h_{n-1}, O_{n-1}) dh_{n-1}$$
$$= \int_{h_{n-1}} p(o_n|h_n) p(h_n|h_{n-1}, O_{n-1}) p(h_{n-1}, O_{n-1}) dh_{n-1}$$
$$= p(o_n|h_n) \int_{h_{n-1}} p(h_n|h_{n-1}) p(h_{n-1}, O_{n-1}) dh_{n-1} \quad (2)$$

According to Bayes law,

$$p(h_n, O_n) = p(h_n|O_n) p(O_n)$$

Substituting this into (2) yields

$$p(h_n|O_n) p(O_n) =$$
$$p(o_n|h_n) \int_{h_{n-1}} p(h_n|h_{n-1}) p(h_{n-1}|O_{n-1}) p(O_{n-1}) dh_{n-1}$$

Further simplification yields

$$p(h_n|O_n) =$$
$$\frac{p(O_{n-1})}{p(O_n)} p(o_n|h_n) \int_{h_{n-1}} p(h_n|h_{n-1}) p(h_{n-1}|O_{n-1}) dh_{n-1}$$
$$= k p(o_n|h_n) \int_{h_{n-1}} p(h_n|h_{n-1}) p(h_{n-1}|O_{n-1}) dh_{n-1} \quad (3)$$

where $k = \frac{p(O_{n-1})}{p(O_n)}$ is a normalizing factor to ensure sum of probabilities to be 1. This derivation gives a recursive algorithm for dynamic Bayesian inference of h_n given evidence $o_1, o_2, \ldots, o_n$; namely, given conditional probability at time $n-1$, $p(h_{n-1}|O_{n-1})$, the conditional probability at next time n, $p(h_n|O_n)$, can be calculated recursively.

The remaining question is how $p(h_n|h_{n-1})$ and $p(o_n|h_n)$ can be specified in practice. For the hidden Markov model [Rabiner, 1989], $p(h_n|h_{n-1})$ is defined as state transition probability, and $p(o_n|h_n)$ is known as emission probability. In state space notation they represent state transition and observation respectively. The observation o_n is typically continuous-valued and may be written as

$$o_n = \phi(h_n) + e_n$$

where $\phi()$ is a nonlinear function and e_n follows certain probability distribution function $p(e_n)$. Then

$$p(o_n - \phi(h_n)|h_n) = p(e_n)$$

from which, $p(o_n|h_n)$ can be readily derived.

The state transition probability $p(h_n|h_{n-1})$ may be derived from *a priori* information such as the reliability of an instrument or from historical data. For example, a sensor takes two states, normal and abnormal. Its mean time between failure (MTBF) is a reliability data and may be derived as [Smyth, 1994]

$$MTBF = \frac{T}{1 - a_{11}}$$

where T is the sampling time and a_{11} is the state transition probability $p(h_{t+1} = normal|h_t = normal)$. Here h_t represents the state of the sensor.

To illustrate, let's revisit the sensor diagnosis example. If the sensor diagnosis Bayesian model is rolled over along time dimension, a dynamic Bayesian model can be formed as shown in Fig.7, where we only show the dynamic Bayesian model at two time instances, t and $t+1$. It can be expanded to include other time instances easily. The intra-connections within each time instance are the same as that of the static Byesian sensor model, while the inter-connections between consecutive instances represent dynamic state transitions. The dynamic Bayesian inference derived above can be used to make sequential sensor diagnosis.

As an example, let the process gain status be of main interest for the diagnosis. With both sensor reading (y=0.4) and input reading (u=-1) at time instance t, the Bayesian inference indicates that the conditional probability for the process gain to be abnormal is 65.9%. The predicted gain abnormality at time $t+1$ and $t+2$ is 72.7% and

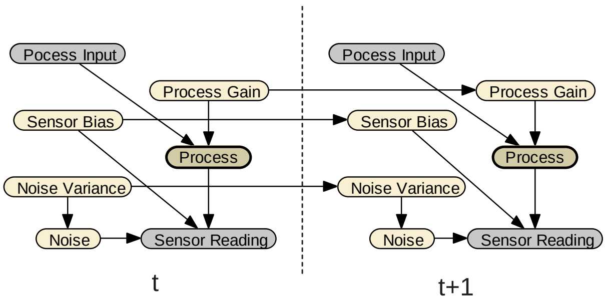

Fig. 7. Dynamic Bayesian model for sensor fault diagnosis.

78.2%, respectively. Evolving time instance from t to $t+1$, new input reading ($u_{t+1} = -1$) and sensor readings ($y_{t+1} = 0.6$) are available in addition to that at time t. With evidences in the two time instances, the dynamic Bayesian inference shows that the probability for the gain to be abnormal at $t + 1$ is 99.1%, and the predicted abnormal status at future time $t + 2$ is 99.3%, and the smoothed probability of abnormality of the gain at t (looking back from time $t + 1$) is now 51.2%. Clearly, the dynamic Bayesian models provide us with more reliable diagnosis by using all available observations.

5. BEYOND BASIC CONTROL LOOPS: BAYESIAN METHODS FOR MODEL PREDICTIVE CONTROL ANALYSIS

In this section, how Bayesian methods can be used for MPC performance analysis and decision making will be addressed, with a focus on the analysis of variance changes through tuning and constrain limit changes through constraint adjustments, and their impact on MPC performance. In industrial MPC applications, constraints on some controlled variables (CV) and manipulated variables (MV) are often set conservatively [Singh and Seto, 2002], leading to possible loss of profits. It is of practical interest to know how the change of the constraint limits or variance reductions impacts MPC performance.

For illustration, consider an MPC application with two MVs and two CVs. One of the CVs y_1 is a quality variable that determines the profit. Due to the variability, all CVs have to be back off certain distance from the constraints to avoid constraint violations. In general, the maximum profit of quality variables lie in the constraints, but the quality variables are often not able to operate near their optimum either due to their own variability or due to constraints on MVs and other CVs. There are three possible methods to push quality CVs closer to their optimum, namely reduce variability, relax constraints on some MVs, and relax constraints on some CVs.

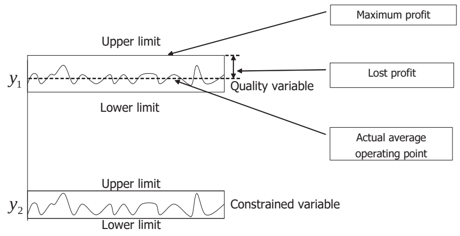

Fig. 8. Base case operation

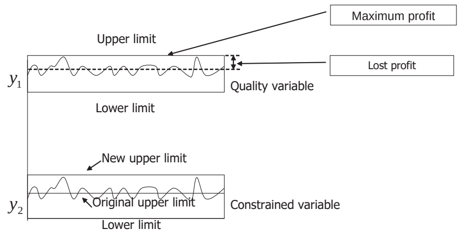

Fig. 9. Benefit by relaxing constraint limit

Because of disturbances, there is variability on both y_1 and y_2. Assuming that the maximum profit point of y_1 is located on its upper limit as shown in Fig.8, it is clear that the actual (average) operating point (dash line) is not at its optimum, leading to lost profit. The lost profit in this case may be attributed to the non-quality CV, y_2, which has been operated near its constraint. Any further upward move of y_1 may make y_2 violate its constraint due to the interaction. Therefore, operation of each CV is determined by the mean values and variances of all CVs and MVs that are interacting. It is assumed that, for CVs, a reasonable percentage of constraint violation, for example, 5%, is allowed such that 95% of operation falls within the constraints for a normal process operation. To move the operating point of y_1 closer to its optimum, the solutions may be to relax the constrain limits of y_2 as indicated by Fig.9 or to reduce the variability as indicated by Fig.10.

The analysis methods proposed here will serve for two objectives. If some CVs and MVs can be changed their variances or the constraint limits, these methods will provide a guideline on how and which variances or constraint limits should be changed in order to achieve the desired profit. If, however, the variances or constraints can not be changed, these studies will inform operation personnel which variances or constraints the profit depends on the most and one has to be careful in setting the variance targets or constraint limits for these variables. There is tendency to set conservative constraint limits in practice. This type of analysis will help reduce the conservativeness for profit sensitive variables.

In general, some of CVs and some of MVs may be allowed to relax their constraints to certain percentage. The optimal operation of the process

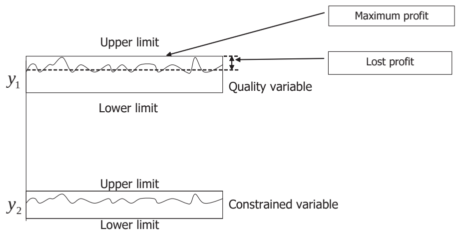

Fig. 10. Benefit by reducing variability

after the relaxation of constraints without reducing the variability can be formulated as

$$\begin{array}{ll} Max_{u^o} & \textit{Profit Function} \\ \underline{st} & \textit{P(CV constraint violation)} < 5\% \\ & \textit{MVs are all inside constraints} \\ & \textit{Steady state input-output relation} \end{array} \quad (4)$$

where u^o is the optimal operating point of MVs; the probability distribution function (e.g. Gaussian) is determined from the routine operating data in terms of the shape (variance/covariance for Gaussian).

Similarly, process variability may be changed by tuning the MPC. As an example, we consider the tuning through adjustment of weights on CVs and MVs. Upon each set of tuning of the MPC weighting parameters, an optimal control problem at the dynamic control level is solved and new distribution of input/output in terms of the shape (variance/covariance for Gaussian) is determined in a similar way as [Xu et al., 2007]. With the shape of the new distribution function, the optimal operating points owing to the change of weighting are again solved from eqn.(4).

Let n CVs and m MVs be allowed to adjust their wights or relax their constraints. Optimizations will be carried out off-line to find the optimal operating points for each combination of changes. Let q be the number of quality CVs, a Bayesian model can be created with n CVs and m MVs as parent nodes and q CVs as the child node. The probability distribution function of q quality CVs operating at optimal operating points represents the conditional probability distribution of the quality CVs. The expected profit of changing the weighting or the constraints can be inferred from the Bayesian model. On the other hand, if certain profit is desired, the maximum aposterior estimate of changes in the constraints or in the weights can be inferred too. Some industrial application results will be presented in [Agarwal et al., 2007] in the same conference.

6. CONCLUSION AND PROBLEMS FOR FUTURE RESEARCH

A novel framework for synthesizing control loop monitoring and diagnostic problems is developed in this paper. It is shown that the emerging Bayesian methods are the appropriate solution for control loop monitoring and diagnosis. Not only do the Bayesian methods quantify the probability for diagnosis solutions, but also they are flexible in the form of models, structures, and data including the missing variables. Several process and control diagnosis problems including control loop performance diagnosis, valve problem diagnosis, sensor problem diagnosis, dynamic inference diagnosis, and model predictive control performance analysis have been formulated under the Bayesian framework and their solutions are illustrated through examples.

We have also applied Bayesian methods to several other problems with actual industrial applications background, including soft sensors and process modeling. The following are some of experiences gained from these applications and problems that need to be addressed:

(I) Bayesian methods are applicable to a wide range of practical problems. Not only can the Bayesian methods be used to infer unknowns but also they can be used for decision making to optimize certain objective functions.

(II) Since Bayesian methods solve probability distribution problem, the results can be prioritized accordingly to facilitate decision making. Other statistics such as confidence intervals are inherited in the solution.

(III) Bayesian methods can handle many types of models including quantitative and qualitative models, linear and nonlinear models, first principle and data-driven models, Gaussian and non Gaussian distributed models, etc.

(IV) Bayesian methods handle missing data or missing/hidden variables in a very natural way. The celebrated Kalman filter is a Bayesian method and the state is so called hidden variable.

(V) Although exploiting independency of variables such as Bayesian network does reduce computation burden considerably, the computation load and need of large memory are the bottleneck when dealing with large scale problems. Efficient computation methods and minimization of memory requirement are the active areas of research. A good reference can be found in [Pernestal, 2007].

(VI) Determination of prior is another challenging problem and has been historically a controversial issue too. Appropriate setup of the prior can improve the inference while incorrect choice of prior may harm the inference. Certainly, the prior gets less important when more training data are used. Without any information, however, the prior can be simply set as uniform. In fact, conventional data driven approaches (such as likelihood estimation approach) are naturally related to the Bayesian approach when the prior is set to be uniform [Pernestal, 2007]. Thus debate of conventional data-driven vs Bayesian approach will be going on, which together with determination of prior is a direction of active research. The good references can be found in [Kass and Wasserman, 1996; Jaynes, 2001; Pernestal, 2007].

(VII) The representation of the Bayesian graphic model is not unique. The determination of independency of variables is nontrivial, particularly for large scale problems. Structure learning of graphic models directly from training data is another problem deserved research. Some references of this topic can be found in [Cooper and Herskovits, 1992; Pernestal, 2007].

(VIII) The conditional probability distributions may be determined from first-principle derivations or through historical data training. Both can result in some errors or uncertainties. Sensitivity of Bayesian inference to these uncertainties is yet to be studied.

REFERENCES

N. Agarwal, B. Huang, and E.C. Tamayo. Bayesian approach for constraint analysis of mpc performance and industrial application. In *DYCOPS 8, accepted*, Cancun, Mexico, 2007.

M. Basseville and I.V. Nikiforov. *Detection of Abrupt Changes*. Prentice Hall, Englewood Cliffs, New Jersey, 1993.

M. Choudhury, S.L. Shah, and N.F. Thornhill. Diagnosis of poor control-loop performance using higher-order statistics. *Automatica*, 40(10): 1719–1728, 2004.

G.F. Cooper and E. Herskovits. A bayesian method for the induction of probabilitic networks from data. *Machine Learning*, 9:309–347, 1992.

T.J. Harris. Assessment of closed loop performance. *Can. J. Chem. Eng.*, 67:856–861, 1989.

T.J. Harris, C.T. Seppala, and L.D. Desborough. A review of performance monitoring and assessment techniques for univariate and multivariate control systems. *J. of Process Control*, 9:1–17, 1999.

A. Horch. *Condition monitoring of control loops.* PhD thesis, Royal Institute of Technology, Sweden, 2000.

A. Horch and G. Dumont. Guest editorial. *International Journal of Adaptive Control and Signal Processing: special issue for control loop performance assessment*, 17:523–525, 2003.

B. Huang. Multivariable model validation in the presence of time-variant disturbance dynamics. *Chemical Engineering Science*, 55:4583–4595, 2000.

B. Huang and S.L. Shah. *Control Loop Performance Assessment: Theory and Applications.* Springer Verlag, London, 1999.

B. Huang, A. Malhotra, and E.C. Tamayo. Model predictive control relevant identification and validation. *Chem. Eng. Sci.*, 58:2389–2401, 2003.

B.T. Jaynes. *Probability Theory - the logic of science.* Cambridge University Precess, 2001.

M. Jelali. An overview of control performance assessment technology and industrial applications. *Control Engineering Practice*, 14:441–466, 2006.

R.E. Kass and L. Wasserman. The selection of prior distributions by formal rules. *Journal of the American Statistical Association*, 91:1343–1370, 1996.

P. Kesavan and J.H. Lee. Diagnostic tools for multivariable model-based control systems. *Ind. Eng. Chem. Res.*, 36:2725–2738, 1997.

K.B. Korb and A.E. Nicholson. *Bayesian artificial intelligence.* Chapman & Hall/CRC UK, 2003.

C.A. McNabb and S.J. Qin. Projection based mimo control performance monitoring: Ii - measured disturbances and setpoint changes. *Journal of Process Control*, 15(1):89–102, 2005.

K.P. Murphy. *Dynamic Bayesian Networks: Representation, Inference and Learning.* PhD thesis, University of California, Berkeley, 2002.

A. Pernestal. *A Bayesian approach to fault isolation with application to diesel engine diagnosis.* PhD thesis, KTH School of Electrical Engineering, Sweden, 2007.

S.J. Qin. Control performance monitoring - a review and assessment. *Computers and Chemical Engineering*, 23:173–186, 1998.

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

J. Schafer and A. Cinar. Multivariable mpc system performance assessment, monitoring, and diagnosis. *Journal of Process Control*, 14:113–129, 2004.

P. Singh and K. Seto. Analyzing apc performance. *Chemical Engineering Progress*, pages 60–66, 2002.

P. Smyth. Hidden markov models for fault detection in dynamic systems. *Pattern Recognition*, 279:149–164, 1994.

N.F. Thornhill and T. Hagglund. Detection and diagnosis of oscillation in control loops. *Control Engineering Practice*, 5(10):1343, 1997.

F. Xu, B. Huang, and A. Akande. Performance assessment of model predictive control for variability and constraint tuning. *Ind. Eng. Chem. Res.*, 46:1208–1219, 2007.