Bayesian Method for Multirate Data Synthesis and Model Calibration

Xinguang Shao, Biao Huang, and Jong Min Lee
Dept. of Chemical and Materials Engineering, University of Alberta, Edmonton, AB, Canada T6G 2G6

Fangwei Xu and Aris Espejo
Syncrude Canada Ltd., Fort McMurray, AB, Canada T9H 3L1

DOI 10.1002/aic.12358
Published online August 23, 2010 in Wiley Online Library (wileyonlinelibrary.com).

Data-driven models are widely used in process industries for monitoring and control purposes. No matter what kind of models one chooses, model-plant mismatch always exists; it is, therefore, important to implement model update strategies using the latest observation information of the investigated process. In practice, multiple observation sources such as frequent but inaccurate or accurate but infrequent measurements coexist for a same quality variable. In this article, we show how the flexibility of the Bayesian approach can be exploited to account for multiple-source observations with different degrees of belief. A practical Bayesian fusion formulation with time-varying variances is proposed to deal with possible abnormal observations. A sequential Monte Carlo sampling based particle filter is used for simultaneously handling systematic and nonsystematic errors (i.e., bias and noise) in the presence of process constraints. The proposed method is illustrated through a simulation example and a data-driven soft sensor application in an oil sands froth treatment process. © 2010 American Institute of Chemical Engineers *AIChE J,* 57: 1514–1526, 2011
Keywords: model calibration, Bayesian fusion, particle filter, soft sensor, oil sands

Introduction

In today's competitive process industries, the pressure to improve the performance of processing facilities is intensive. Unavailability or inaccuracy of process information can have a significant impact on plant performance and thus result in substantial loss of revenues. Moreover, real-time process monitoring is becoming a prerequisite to sustain both plant safety and profitability. However, it is common that accurate real-time information on critical process variables is unavailable because of various causes such as sensor reading errors, sensor failures, and sensor unavailability.[1]

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

To obtain more accurate and reliable real-time process information, soft sensors (a.k.a. virtual sensor) have been investigated in many process industries (see Refs. 2–9 and references therein). The idea of soft sensors is to use a

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

process model that provides online estimates of difficult-to-measure quality variables (e.g., melt index, pH value, and concentration) from readily available process variables (e.g., temperature, pressure, and flowrate). To achieve a successful soft sensor application, the process model is the key. Generally, there are two well-known approaches for building a process model: first-principle approaches[10,11] and data-driven approaches.[9,12,13] A first-principle model is based on good understanding of the underlying fundamental principles such as mass and energy balances. A data-driven model is based on limited process knowledge and mainly relies on the historical data describing input (i.e., process variable) and output (i.e., quality variable) characteristics. Increased complexity of the process dynamics often prevents one from building accurate first-principle models. On the other hand, data-driven approaches have been extensively used for modeling of complex systems as the process related signals are rather easy to obtain from instruments and experiments.[9]

The main challenge of the data-driven modeling arises from large model-plant mismatch. As the available training data only describes a period of process historical behavior, the investigated process could have changed over time; therefore, large model-plant mismatch may still exist even though the model initially may be sufficient. A natural question then arises: how to use the latest observations of quality variables to update the model for better estimates.

Motivated by the above question, this article focuses on the development of a data-driven model update approach for soft sensor applications based on multiple-source quality variable observations. The proposed approach is built on a Bayesian framework,[14] which facilitates the inclusion of additional information in the form of prior knowledge and the synthesis of fast sampled but less accurate observations with more accurate but slow sampled observations to derive more accurate posterior distribution for the unknown state and parameters. To enhance the robustness, a practical Bayesian fusion formulation with time-varying variances is proposed and observation validity is taken into account. The Bayesian model calibration strategy is finally implemented by using a sequential Monte Carlo sampling approach[15] and applied to an industrial soft sensor design.

The remainder of this article is organized as follows: Section ''Date-Driven Models'' gives a literature review on data-driven modeling using different sampling rate of input/output data for soft sensor developments. Section ''Bayesian Calibration of the Date-Driven Model'' introduces model calibration strategies with Bayesian information synthesis. Section ''Bayesian Information Synthesis with Abnormal Observation Date'' introduces a robust Bayesian fusion formulation for handling abnormal observations. Section ''Sequential Monte Carlo Sampling Approach for Constrained Bayesian Estimation'' implements the Bayesian model calibration strategy as a sequential Monte Carlo sampling based constrained particle filter. Section ''An Illustrative Example'' presents a simulation example to show the characteristics and benefits of the proposed approach. An oil sands froth treatment process is introduced and data-driven soft sensor application results are illustrated in Section ''Industrial Application''. Last Section gives the conclusions.

Data-Driven Models

Both scientific and engineering communities have acquired extensive experience in developing and using data-driven modeling techniques. Despite a variety of model structures, two types of data-driven models are widely seen in the literature. One is dynamic model, and the other is static model.

Dynamic modeling based on fast-rate input/output

When on-line instruments are available for both input and output variables, a set of valid input and output data can be collected from the historical database. In this case, a fast-rate dynamic model can be generally identified for the investigated process.[16] The book by Ljung[17] is considered as a milestone in the field of dynamic identification theory. The identification methods described therein are commonly used for linear dynamic modeling, including autoregressive models (e.g., ARX, ARMAX), Output-Error (OE) models, Box-Jenkins (BJ) models, and state space models. Estimation techniques include prediction-error minimization schemes and various subspace methods. When linear models are not sufficient to capture system dynamics, one can resort to non-linear models such as non-linear ARX (NLARX)[18] and Hammerstein-Wiener models.[19]

Static modeling based on slow-rate input/output

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

For static data-driven modeling, linear regression methods (e.g., ordinary least squares, OLS,[20]) are commonly used. However, OLS may suffer from numerical problems when a data set is collinear, which is not uncommon in chemical processes. Principal component regression (PCR) and partial least squares (PLS) address the collinearity by projecting the original process variables onto a low dimensional space of orthogonal latent variables. PCR and PLS techniques are well reviewed in Ref. 21–23 and references therein. For the nonlinear case, nonlinear regression methods, such as artificial neural network,[24] support vector machine,[6] and fuzzy logic,[25] could be used.

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

Ignoring process dynamics in a static model is one of the causes of model inaccuracy.[26] To improve this, dynamic modeling using fast-rate input and slow-rate output has received considerable attention in both academic and industrial communities.

A special case widely investigated is known as dual-rate system identification, where the output sampling time is slower than the input sampling time. Early contributions can be found in Lu and Fisher,[27,28] in which a polynomial transformation technique and a least squares algorithm are presented to produce fast-rate output based on the measurements of fast-rate input and slow-rate output. The main disadvantage of their algorithm is that additional parameters are introduced. Li et al.[29] and Wang et al.[16] use a so-called lifting technique to extract the original fast single-rate system by identifying a higher dimension lifted model. However, this technique becomes impractical when the output sampling rate is very slow and irregular. Ding and Chen[30] propose to use an auxiliary finite impulse response (FIR) model to predict the noise-free fast-rate output, and then identify a single-rate dynamic model based on the fast-rate input and the estimated output. Raghavan et al.[31] use an Expectation-Maximum (EM) based approach to interpolate fast-rate output, and then apply a single-rate dynamic identification method; both regular and irregular sampled slow-rate output data can be treated in this approach. However, implementation of the EM algorithm can be expensive for practical applications, and the solution may converge to a local optimum. Zhu et al.[26] propose to use an OE method to identify a dynamic model directly from the fast-rate input and slow-rate output by minimizing the summation of the squared error between the model output and the measurement at the slow rate; the method has the potential to deal with irregular output, and the authors demonstrated their work through an industrial case study, but it requires a good initial model to avoid the local optimum. Mo et al.[32] propose to use a FIR model as an initial model, and then apply a fast single-rate OE model for the dynamic identification. Lu et al.[33] developed a multirate dynamic inferential model based on multiway PLS approach, and demonstrated its efficacy through the Tennessee Eastman process. Lakshminarayanan et al.[34] developed a method called Data Selection and Regression (DSAR) for identifying irregularly sampled systems and applied it to soft sensor development on a two-reactor train system.

Bayesian Calibration of the Data-Driven Model

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

$$y_{k+1} = \hat{f}(\phi_k, \theta) + \varepsilon_{k+1}, \tag{1}$$

where $\phi_k = [y_k,\cdots,y_{k-n_y},u_k,\cdots,u_{k-n_u}]^T$ is a regressor vector consisting of output and input. n_y and n_u are the model order parameters, which can be determined by minimizing Akaike information criterion[35]; $\hat{f}(\cdot)$ is a selected model structure describing a linear or nonlinear relationship between the input and output variables; θ is the model parameter estimated from the training data; and ε_{k+1} is the output residual. Note that for a static model, the regressor only contains one input term.

In reality, no model is perfect. The model-plant mismatch could be significant in a data-driven model, and mainly arises in two stages. One is in the modeling stage, such as misuse of model structure or insufficiency of training data; the other is in the application stage, such as the drift of operating conditions or the degradation of equipment efficiency. To obtain a better estimate of the true quality information, it is important to synthesize all the available observations, and then use them to update the existing model with the consideration of uncertainty.

Strategies for model updating roughly fall into two categories: model refinement and model calibration.[36] Model refinement involves the change of model structure, for example, using a nonlinear model to replace a linear model, which is desirable for fundamentally improving the predictive capability; however, the practical feasibility of refinement is often restricted by available knowledge and computing resource. In contrast, model calibration utilizes mathematical means to match model predictions with reliable observations, which is a cheaper way for practical applications.

Model calibration

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

$$\hat{y}_{k+1} = \hat{f}(\phi_k, \theta_k), \tag{2}$$

where θ_k represents time-varying model parameters.

However, in many situations, calibrating model parameters is still unable to compensate model-plant mismatch, for example, due to the use of incorrect model structure. Then, the following bias correction form could be used,[37,38]

$$\hat{y}_{k+1} = \hat{f}(\phi_k, \theta_k) + \gamma_k, \tag{3}$$

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

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

$$\hat{y}_{k+1} = \rho_k \hat{f}(\phi_k, \theta_k) + \gamma_k, \tag{4}$$

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

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

Bayesian information synthesis

There are a few data fusion (or information synthesis) approaches to resolve the above question[39–41] of which the Bayesian inference based approach is the most unified one. Figure 1 shows three most popular strategies for Bayesian information synthesis. Figure 1(a) shows the state vector fusion method, also known as the distributed approach, where a group of Bayesian filters are used to obtain individual observation based estimates, and then fused together (e.g., linear combination) to obtain an improved joint estimate. It is a favorable choice for processes with numerous observation sources because of computation cost as well as the parallel implementation and fault-tolerance issues.[42]

However, this approach requires consistent Bayesian filters, and inappropriate combination of individual estimate can deteriorate the final result.[43] Figure 1(b) shows the measurement fusion approach, also known as the centralized approach, in which all the observations are directly fused to obtain synthesized process information, and then a single Bayesian filter is used to obtain the final estimate. Figure 1(c) shows a hybrid use of the distributed and centralized approaches, resulting in a more complicated sequential fusion scheme. It yields the same result as centralized one when the number of observation sources equals to two. In this article, the centralized approach is selected as it is the best way to synthesize observations in the sense that no information will be lost during the fusion procedure[41] and the number of observation sources for the problem investigated in this article is not large.

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

$$
\begin{aligned}
x_{k+1} &= \begin{bmatrix} 0 & \cdots & 0 & 0 & & & & & & \\ I & & & 0 & & & & & & \\ & \ddots & & \vdots & & & & & & \\ & & I & 0 & & & & & & \\ & & & & 0 & \cdots & 0 & 0 & & \\ & & & & I & & & 0 & & \\ & & & & & \ddots & & \vdots & & \\ & & & & & & I & 0 & & \end{bmatrix} x_k + \begin{bmatrix} \rho_k \hat{f}(x_k, u_k, \theta_k) + \gamma_k \\ 0 \\ \vdots \\ 0 \\ u_k \\ 0 \\ \vdots \\ 0 \end{bmatrix} + \begin{bmatrix} I \\ 0 \\ \vdots \\ 0 \\ 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix} \omega_k^x, \\
\theta_{k+1} &= \theta_k + \omega_k^\theta, \\
\rho_{k+1} &= \rho_k + \omega_k^\rho, \\
\gamma_{k+1} &= \gamma_k + \omega_k^\gamma, \\
y_{T_k^n k}^n &= H x_{T_k^n k} + v_{T_k^n k}^n \\
&= [1 \quad 0 \quad \cdots \quad 0] x_{T_k^n k} + v_{T_k^n k}^n, n = 1, \cdots, N,
\end{aligned}
\tag{5}
$$

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

With the calibration parameter vector denoted as $\Theta_k = [\theta_k, \rho_k, \gamma_k]^T$, Equation 5 can also be represented by a probabilistic graph as shown in Figure 2, where all the unknown nodes are considered as random variables. (Note that the arc between x_{k-1} and x_k is left out if $\hat{f}(\cdot)$ is a static model.)

The objective of Bayesian information synthesis is to construct the a posteriori distribution, $p(x_k, \Theta_k | D_k)$, of the state (or unknown true quality variable), x_k, and the calibration parameter, Θ_k, based on available multiple-source noisy observations, $D_k = \{Y_1, \cdots, Y_k\}$, where $Y_k = \{y_k^1, \cdots, y_k^N\}$.

In principle, the required posterior distribution can be obtained by recursively following two steps: prediction and update.

Prediction. At time $k-1$, all the evidence up to time $k-1$ has been taken into account, and the posterior distribution $p(x_{k-1}, \Theta_{k-1} | D_{k-1})$ has been estimated. Then, the prior distribution at time k can be obtained as[44]:

$$
\begin{aligned}
&p(x_k, \Theta_k | D_{k-1}) \\
&= \int p(x_k, \Theta_k | x_{k-1}, \Theta_{k-1}) p(x_{k-1}, \Theta_{k-1} | D_{k-1}) dx_{k-1} d\Theta_{k-1}.
\end{aligned}
\tag{6}
$$

Here the probabilistic models $p(x_k | x_{k-1}, \Theta_{k-1})$ and $p(\Theta_k | \Theta_{k-1})$ are defined by the system equations and the associated statistics of ω_k^x, ω_k^θ, ω_k^ρ, and ω_k^γ. A graphical interpretation is shown in Figure 3.

Update. At time k, the latest observation $Y_k = \{y_k^1, \cdots, y_k^N\}$ is available (see Figure 4), then the posterior distribution can be obtained via Bayes' rule,

$$
\begin{aligned}
p(x_k, \Theta_k | D_k) &= \frac{p(Y_k | x_k, \Theta_k) p(x_k, \Theta_k | D_{k-1})}{p(Y_k | D_{k-1})} \\
&= \frac{p(y_k^1 | x_k, \Theta_k) p(y_k^2 | x_k, \Theta_k) \cdots p(y_k^N | x_k, \Theta_k) p(x_k, \Theta_k | D_{k-1})}{p(y_k^1, y_k^2, \cdots, y_k^N | D_{k-1})} \\
&\propto p(x_k, \Theta_k | D_{k-1}) \prod_{n=1}^{N} p(y_k^n | x_k, \Theta_k),
\end{aligned}
\tag{7}
$$

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

$$
p(Y_k | D_{k-1}) = \int p(Y_k | x_k, \Theta_k) p(x_k, \Theta_k | D_{k-1}) dx_k d\Theta_k. \tag{8}
$$

Once the posterior distribution is obtained, it can be used for point state inference, such as mode, mean, or median estimate.

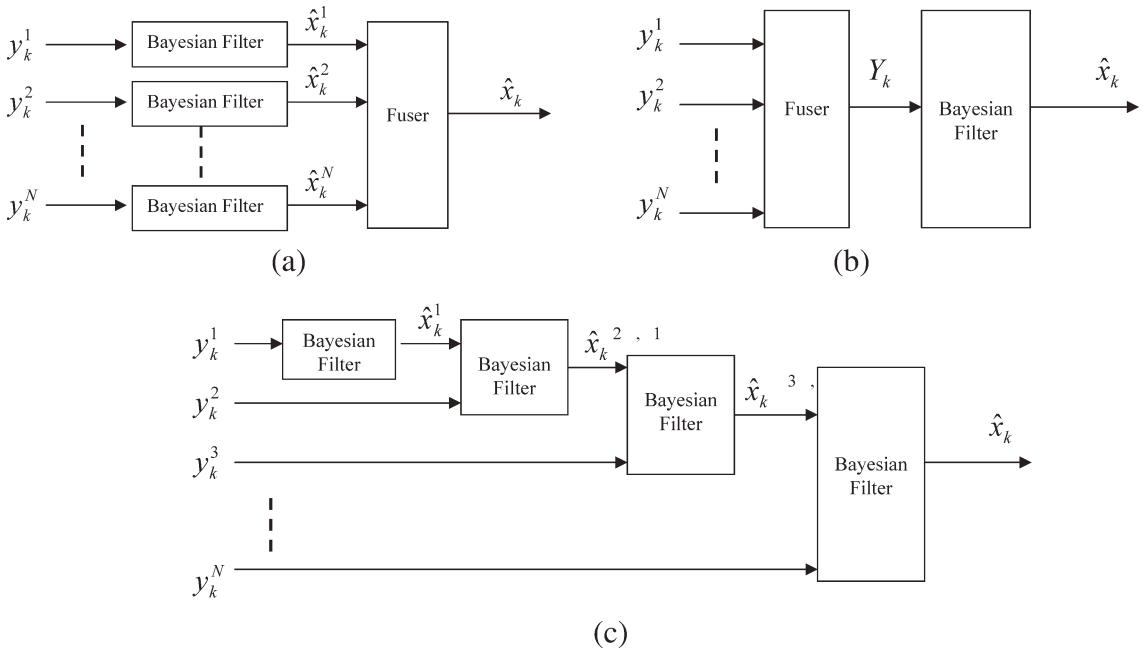

Figure 1. Bayesian filter based data fusion strategies: (a) distributed approach, (b) centralized approach, (c) hybrid (sequential) approach.

Bayesian Information Synthesis with Abnormal Observation Data

In reality, no sensor (or observation source) can provide precise measurements continuously. Because of sensor malfunction, transmission error, or human data entry error, one may obtain ''unexpected'' values for a measured variable. Such abnormal data can be propagated through the fusion procedure and cause a divergent estimate. To achieve a robust estimate in the presence of abnormal data, this section describes a variance adaptation scheme for Bayesian information fusion.

It is well known that the observation noise variance is important for information fusion, as it directly determines the relative weight assigned to the observation source.[45] However, in the real world, the variance of the true observation noise is rarely known; it is generally pre-estimated and kept unchanged during the application. This will yield the same weight to an observation source regardless of its measurement quality. To circumvent this, we assume the noise is subject to a Gaussian distribution with a time-varying variance, namely,

$$p(y_k^n|x_k,\Theta_k) \sim \mathcal{N}(Hx_k,\sigma_n^2(k)), \tag{9}$$

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

Thus, the problem is how to define the normality or abnormality. Hua and Wu[46] suggests that the distance between the n^{th} sensor's observation with respect to the rest sensors can be used to quantify the abnormality. Their method requires at least three observation sources and assumes that the majority of the sources provide correct and consistent measurements.

In this work, motivated by the approach widely used by practicing engineers in the actual operations, a variance adaptation scheme is developed. For an individual sensor, we partition its measurements into three categories: valid, possibly valid, and invalid. (See Figure 5 for an illustration.)

A validity state, λ_k^n, is introduced to indicate the observation validity (i.e., normality) of the n^{th} sensor at time k. $\lambda_k^n = 1$ indicates that the observation data is valid (i.e., normal), and $\lambda_k^n = 0$ indicates that the observation data is

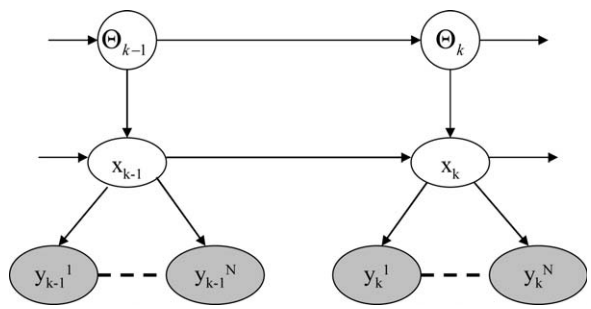

Figure 2. Graphical representation of Eq. 5; grey nodes represent known variables.

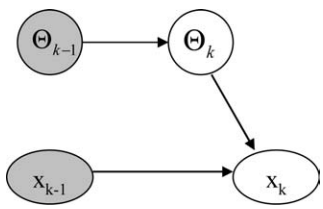

Figure 3. Prediction step for Bayesian inference.

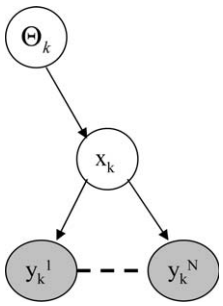

Figure 4. Update validation step for Bayesian inference.

invalid (i.e., abnormal). Then, the time-varying noise variance $\sigma_n^2(k)$ is defined as

$$\sigma_n^2(k) = \begin{cases} \sigma_n^2, & \text{if } y_k^n \in [\alpha_1^n, \alpha_2^n], \text{ i.e., valid,} \\ \frac{1}{p(\lambda_k^n=1)}\sigma_n^2, & \text{if } y_k^n \in [\beta_1^n, \alpha_1^n) \text{ or } y_k^n \in (\alpha_2^n, \beta_2^n], \text{ i.e., possibly valid,} \\ \infty, & \text{if } y_k^n \in (-\infty, \beta_1^n) \text{ or } y_k^n \in (\beta_2^n, +\infty), \text{ i.e., invalid,} \end{cases} \tag{10}$$

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

In Eq. 10, the probability function $p(\lambda_k^n = 1|y_k^n) \in [0,1]$ is a user specified function, which can have different formulations. One option is

$$p(\lambda_k^n = 1|y_k^n) = \begin{cases} \frac{(\beta_1^n-\alpha_1^n)^2-(y_k^n-\alpha_1^n)^2}{(\beta_1^n-\alpha_1^n)^2}, & \text{if } y_k^n \in (\beta_1^n, \alpha_1^n), \\ \frac{(\beta_2^n-\alpha_2^n)^2-(y_k^n-\alpha_2^n)^2}{(\beta_2^n-\alpha_2^n)^2}, & \text{if } y_k^n \in (\alpha_2^n, \beta_2^n). \end{cases} \tag{11}$$

The rationale for Figure 5 and Eq. 11 is based on common industrial practice: (i) specification range for a quality variable does not change substantially during a continuous operation, although input variables can have different operating points; (ii) when a measurement is unusually large or small, the measurement is regarded as abnormal and discarded.

Substituting Eqs. 9 and 10 into Eq. 7, one can obtain the posterior distribution as

$$p(x_k,\Theta_k|D_k) \propto p(x_k,\Theta_k|D_{k-1}) e^{-\left\{\frac{(Hx_k-y_k^1)^2}{2\sigma_1^2}(\lambda_k^1=1)+\cdots+\frac{(Hx_k-y_k^N)^2}{2\sigma_N^2}(\lambda_k^N=1)\right\}}. \tag{12}$$

In Eq. 12, the contribution of an individual sensor to the estimate is decreased (i.e., increasing the variance), if its measurement has low probability to be valid. The influence of a particular sensor will be negligible, as the variance goes to infinity, meaning that its measurement is invalid.

Sequential Monte Carlo Sampling Approach for Constrained Bayesian Estimation

Analytical solutions for Eq. 12 are unavailable in general except for special cases such as unconstrained linear systems with Gaussian noise. However, in practical applications, constrained nonlinear processes are commonly encountered.[47] To implement the Bayesian model calibration strategy described in the previous section, sequential Monte Carlo sampling based particle filter (PF) is applied.

Unlike most other Bayesian filters (e.g., extended Kalman filter, unscented Kalman filter), a particle filter does not rely on linearization technique or Gaussian assumptions. It approximates any a posteriori distribution by a set of particles, $\{x_k^{(i)}, \Theta_k^{(i)}\}$, and their associated weights, $w_k^{(i)} \geq 0$, in a discrete summation form:

$$\hat{p}(x_k, \Theta_k|D_k) = \sum_{i=1}^{M} w_k^{(i)} \delta\left(\begin{bmatrix} x_k \\ \Theta_k \end{bmatrix} - \begin{bmatrix} x_k^{(i)} \\ \Theta_k^{(i)} \end{bmatrix} \right), \tag{13}$$

where $\delta(\cdot)$ is the Dirac delta function; M is the number of particles; w_k^i is the normalized weight associated with the i^{th} particle at time k, and $\sum_{i=1}^{M} w_k^i = 1$.

The ideal case for Monte Carlo sampling is to generate particles directly from the true posterior pdf $p(x_k, \Theta_k|D_k)$, which is generally unavailable. Thus, an easy-to-implement distribution, the so called importance sampling distribution denoted by $q(x_k, \Theta_k|D_k)$, has to be defined before sampling. It is shown that PF is asymptotically unbiased if the support region for the posterior distribution is a subset of the importance sampling distribution.[15]

In practice, the distribution defined by the system equation is often used for importance sampling, i.e., $q(x_k, \Theta_k|D_k) = p(x_k, \Theta_k|x_{k-1}, \Theta_{k-1})$. With this choice, one can derive that the unnormalized importance weight, $\tilde{w}_k^{(i)}$, as[48]

$$\begin{aligned} \tilde{w}_k^{(i)} &\propto w_{k-1}^{(i)} p(Y_k|x_k^{(i)}, \Theta_k^{(i)}) \\ &\propto w_{k-1}^{(i)} e^{-\sum_{n=1}^{N}\left\{\frac{(Hx_k-y_k^n)^2}{2\sigma_n^2}(\lambda_k^n=1)\right\}}, \end{aligned} \tag{14}$$

and the normalized weight is calculated as

$$w_k^{(i)} = \frac{\tilde{w}_k^{(i)}}{\sum_{i=1}^{M} \tilde{w}_k^{(i)}}. \tag{15}$$

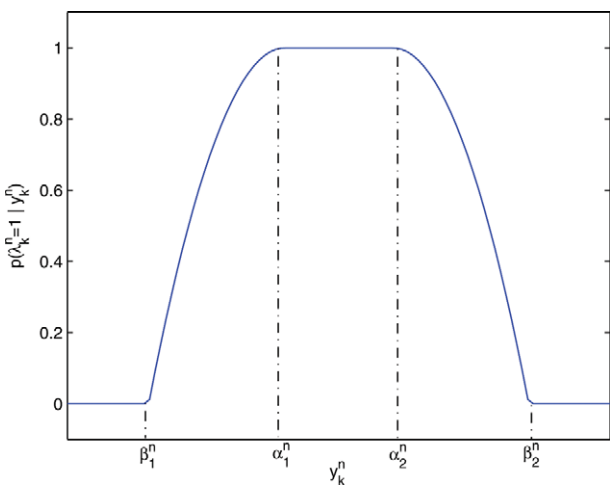

Figure 5. Observation validity given a sensor reading.

[Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

The nature of sample based representation of PF facilitates incorporating constraints into the estimation procedure. Lang et al.[49] discuss how to accept/reject the particles in the PF algorithm based on constraint knowledge. As a slight modification from their work, a constrained likelihood function is defined as:

$$L_c(x_k^{(i)}, y_k^{(i)}) = \begin{cases} 1, & if\ \{x_k^{(i)}, y_k^{(i)}\} \in C_k, \\ 0, & if\ \{x_k^{(i)}, y_k^{(i)}\} \notin C_k, \end{cases} \quad i = 1, \cdots, M, \quad (16)$$

where C_k represents a constraint region at time k. Then, Eq. 14 is modified as

$$\tilde{w}_k^{(i)} \propto w_{k-1}^{(i)} p(Y_k | x_k^{(i)}, \Theta_k^{(i)}) \cdot L_c(x_k^{(i)}, y_k^{(i)}). \quad (17)$$

This modification enables the algorithm to discard all the particles violating state/output constraints because no weighting values are assigned to them.

To avoid particles degeneracy problem,[50] an importance sampling step is usually followed by a resampling procedure to place the particles in regions with high probability. For different resampling strategies, readers are referred to Boloc et al.[51] and references therein.

A constrained PF-based model calibration algorithm is summarized as follows[47]:

Step a. initialization: generate initial particles $\{x_0^{(i)}, \Theta_0^{(i)}\}_{i=1}^M$ from $p(x_0)$, and set $k = 1$;

Step b. importance sampling: generate predicted particles, $\{x_k^{(i),-}, \Theta_k^{(i),-}\}_{i=1}^M$, based on importance sampling distribution;

Step c. weighting: calculate constrained likelihood and importance weights according to Eqs. 14 to 17;

Step d. resampling: generate resampled particles, $\{x_k^{(i)}, \Theta_k^{(i)}\}_{i=1}^M$, based on weighting information and resampling strategy; reset the weights uniformly as $w_k^{(i)} = \frac{1}{M}$;

Step e. output calculate $\hat{x}_k = \sum_{i=1}^M w_k^{(i)} \cdot x_k^{(i)}$ and $\hat{\Theta}_k = \sum_{i=1}^M w_k^{(i)} \cdot \Theta_k^{(i)}$ for the output, set $k = k + 1$ and go back to step b.

An Illustrative Example

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

$$x_k = 0.9 \cdot x_{k-1} - 0.5 \cdot x_{k-2} \cdot (1 + x_{k-1}^2) + u_{k-1} + 0.5 \cdot u_{k-2} + d_{k-1} \quad (18)$$

where $u_{(\cdot)}$ is the input with a sampling time of 1 min; $d_{(\cdot)}$ is the unknown process disturbance (or modeling mismatch term); $x_{(\cdot)}$ denotes the process quality variable (or model output) which has two approaches to measure its values. The first approach has fast sampling rate (1 min) but with low accuracy (controlled by the measurement noise, see Eq. 21), whereas the second one has slow sampling rate (4 h) but with high accuracy.

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

$$u_k = \frac{0.1}{1 - 0.978 q^{-1}} \cdot e_k \quad (19)$$

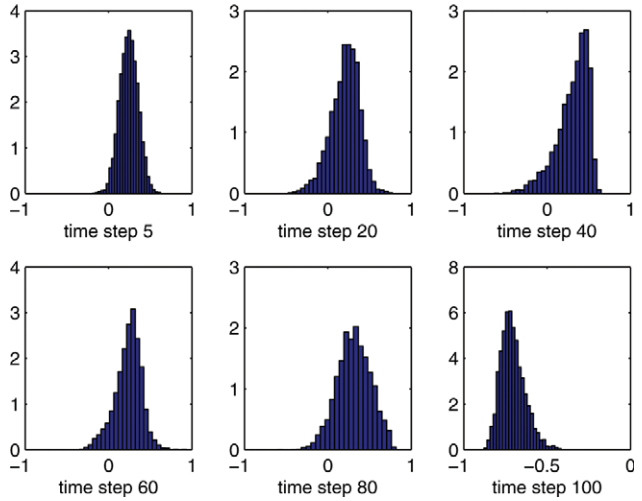

Figure 6. Evolution of the simulated output for the numeric example.

[Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

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

And the unmodeled disturbance term d is designed as

$$d_k = 0.5 \cdot \cos\left(\frac{k}{10\pi}\right) + 0.2 \cdot n_k \quad (20)$$

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

Algorithm characteristics

Non-Gaussianity. Figure 6 shows the evolution of the true unknown output x_k for the above simulation example in which we can see that the distribution for the output is non-Gaussian due to process nonlinearity and the unmodeled disturbance. Traditional Gaussian filters are not suitable to estimate the posterior distribution of x_k, and Monte Carlo sampling based approach is, therefore, selected.

Multirate Observation Fusion. Figure 7 shows a scatter plot of 30 days measurements from two different observation sources in which the slow sampled source is designed with higher accuracy (i.e., smaller uncertainty), whereas the fast sampled source is designed with lower accuracy (i.e., larger uncertainty). Traditional filtering approaches usually use only one of the two observation sources for posterior estimation, namely either $p(x_k|Y_k^1)$ or $p(x_k|Y_k^2)$, due to the implementation difficulties of multirate data. Monte Carlo sampling based approach allows one to use both observation sources for posterior estimation, namely $p(x_k|Y_k^1, Y_k^2)$. Figure 8 shows that fused observation (at time step k) can yield less uncertainty information (i.e., smaller variance) and, therefore, is more likely to produce a better posterior estimate.

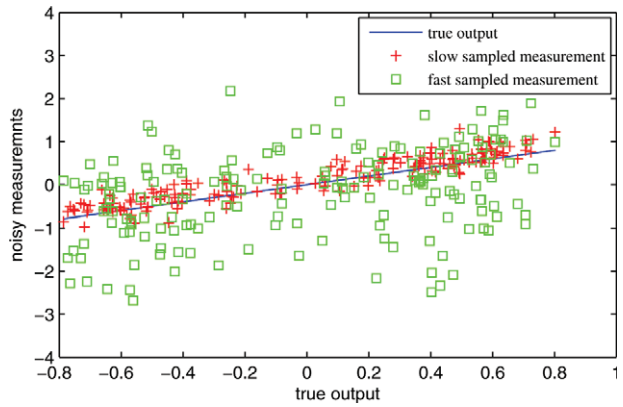

Figure 7. Comparison of two different measuring approaches.

[Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

Robust to Abnormal Readings. Because of the measurement uncertainties, abnormal readings are inevitable in practice, especially for those sensors with large uncertainties. Figure 9 shows the benefit of using time-varying variance to control the influence of a particular measurement (e.g., a possible abnormal reading $y_k^1 = -1.8$). Figure 9(a) shows the fused observation is unable to support the true distribution well when using a prefixed constant variance for each observation source; whereas Figure 9(b) shows an improved fusion result by adjusting each variance according to Eq. 10 with parameters set as $\alpha_1^1 = \alpha_1^2 = -1$, $\alpha_2^1 = \alpha_2^2 = 1$, $\beta_1^1 = \beta_1^2 = -2$, $\beta_2^1 = \beta_2^2 = 2$.

Constraint Handling. Another benefit of using Monte Carlo sampling approach is that it can easily incorporate lower and upper bound constraints of uncertain variables, which is helpful to confine the distribution shape of the related variables and improve the estimation performance. Further information of Bayesian constrained estimation can be found in Shao et al.[47] and reference therein.

Model calibration

Given the data-driven model as Eq. 18 excluding the unmodeled term d, Figure 10 shows the model prediction

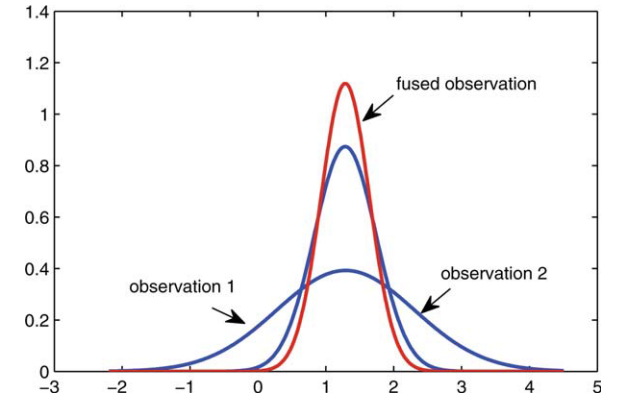

Figure 8. Illustration of observation fusion.

[Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

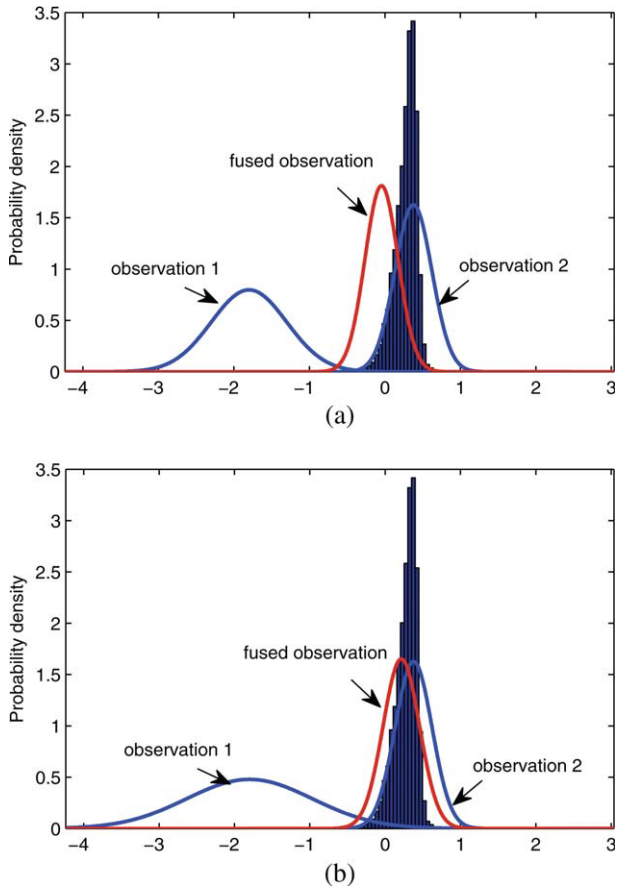

Figure 9. Observation fusion with one possible abnormal reading ($y_k^1 = -1.8$).

(a) Poor fusion result with prefixed constant measurement noise variances ($\sigma_1^2 = 0.5^2$, $\sigma_2^2 = 0.2^2$); (b) improved fusion result with time-varying variances calculated based on Eq. 10 $\sigma_1^2(k) = 0.83^2$, $\sigma_2^2(k) = 0.2^2$). [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

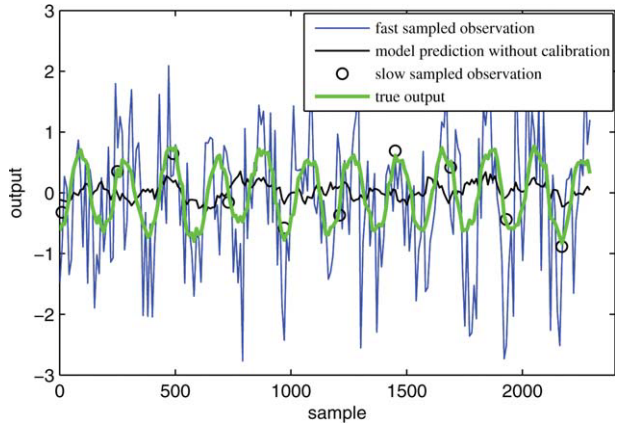

Figure 10. Comparisons of existing measurements and model prediction with the true output.

[Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

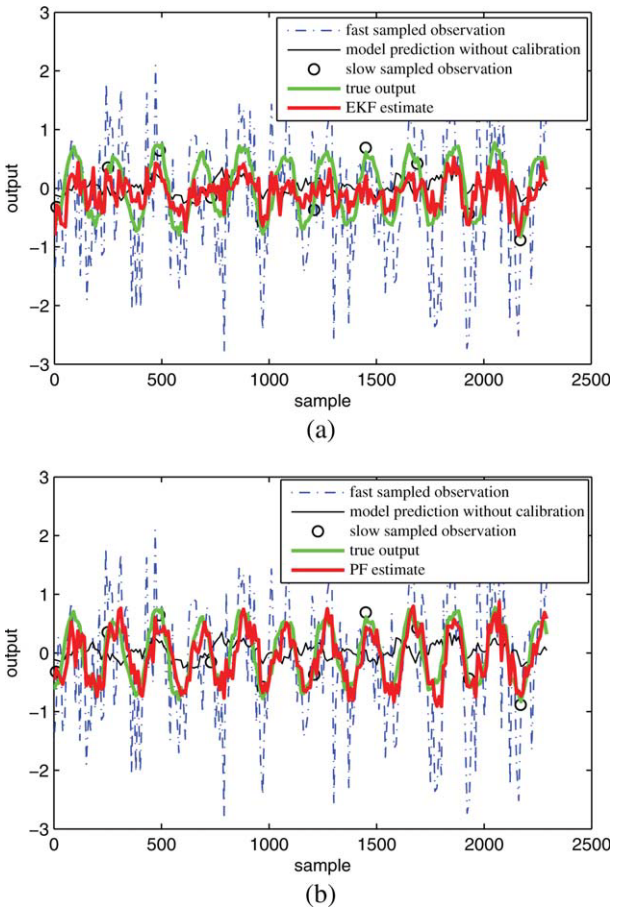

Figure 11. Estimate results with different model calibration approaches; (a) EKF-based approach, and (b) PF based approach.

[Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

(without calibration), the fast rate measurements, the slow rate measurements, and the true output. From the figure, it is observed that there is a large mismatch between the true output and the model prediction. Noted that this kind of comparison is only possible in simulation.

To compensate the model-plant mismatch as much as possible, the proposed Bayesian model calibration strategy is applied to the following reconstructed system:

$$
\begin{aligned}
x^a_{x+1} &= \begin{bmatrix} 0 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} x^a_k + \begin{bmatrix} \rho_k \hat{f}(x^a_k, u_k, \theta_k) + \gamma_k \\ 0 \\ u_k \end{bmatrix} + \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} \omega^{x^a}_k . \\
\rho_{k+1} &= \rho_k + \omega^{\rho}_k, \\
\gamma_{k+1} &= \gamma_k + \omega^{\gamma}_k, \\
y^1_k &= [1 \;\; 0 \;\; 0] x^a_k + v^1_k, \\
y^2_{Tk} &= [1 \;\; 0 \;\; 0] x^a_{Tk} + v^2_{Tk},
\end{aligned}
\tag{21}
$$

where $x^a_k = [x_k, x_{k-1}, u_{k-1}]^T$ is the augmented state variable; y^1_k is the fast sampled measurement with large uncertainty $v^1_k \sim N(-0.1, 1^2)$; y^2_{Tk} is the slow sampled measurement with small uncertainty $v^2_{Tk} \sim N(0.1, 0.2^2)$, and $T = 240$ in this example. $\omega^x_k \sim N(0, 0.5^2)$, $\omega^{\rho}_k \sim N(0, 0.1^2)$, $\omega^{\gamma}_k \sim N(0, 0.1^2)$; parameters for sensor validation range are chosen as $\alpha^1_1 = \alpha^2_1 = -1$, $\alpha^1_2 = \alpha^2_2 = 1$, $\beta^1_1 = \beta^2_1 = -2$, $\beta^1_2 = \beta^2_2 = 2$; lower and upper bound constraints for output are set as $[-2,2]$; 100 particles are used for Monte Carlo sampling.

Figure 11 shows the estimate results using different model calibration approaches. From the comparison, we can see that particle filter based Bayesian calibration approach gives better estimate than the multirate EKF based approach.[52] In fact, it is also much easier to implement the proposed algorithm, as it is applicable to nonlinear functions without the need of linearization.

Industrial Application

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

Background

Oil sands are mixtures of quartz, clay, water, bitumen, and accessory minerals. Athabasca oil sands in Northern Alberta, Canada, is one of the largest oil sands reserve in the world and, currently, produces over one million barrels of oil per day. In the process of producing oil from oil sands, the main task is to separate bitumen from other components. The separation is performed through a chain of industrial clarifying units, among which the Inclined Plates Settler (IPS) units are one of the main components of the froth treatment process.

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

The occurrence of the gravity movement in an IPS unit leads to the hydrocarbon-rich phase (light phase) float up and be collected by the outlet boxes to be discharged as overflow product. The denser water and solid-rich phase (heavy phase) settles down the plate and is collected in the hopper to be discharged as underflow tailings. Figure 12 gives a schematic representation of an IPS unit.

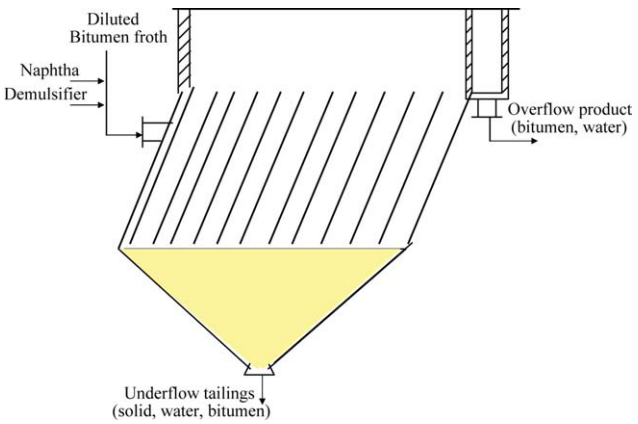

Figure 12. Schematic diagram for an Inclined Plates Settler (IPS) unit.

[Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

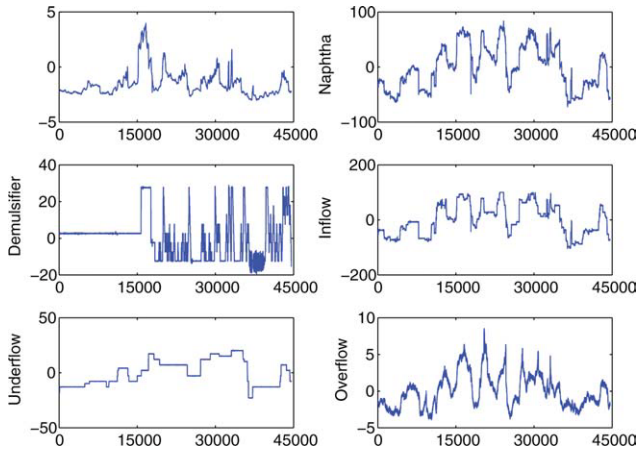

Figure 13. Input and output data for NLARX modeling of the investigated IPS unit.

[Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

The water percentage (a.k.a. water content) in the overflow product is a particularly important variable as it reflects the bitumen froth quality and process performance. In practice, both laboratory analysis (e.g., using Karl Fischer titration[53]) and hardware instrument (e.g., water-cut meter) are available in the overflow stream. Although the laboratory analysis provides more accurate measurements, the sampling rate, which is 2 h in this case, is too slow to serve for monitoring and control purposes. Water-cut meter readings are fast sampled but not accurate enough. Because of the considerable variability in sands, water, clay, and bitumen content, the water-cut meter occasionally needs to be removed for maintenance, which leads to the unavailability of online water content information. This poses challenge to control the Naphtha and Demulsifier additions. Therefore, there is an economic necessity to develop a soft sensor to obtain more accurate and reliable real-time water content information.

Model estimation

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

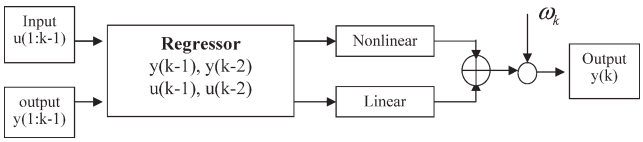

Figure 14. A general structure of nonlinear ARX model.

Figure 13 shows the collected raw input and output data. For proprietary reason, the actual operating ranges have been modified. After data preprocessing, a second-order NLARX is identified to represent the process dynamics.

Figure 14 shows the structure of the NLARX model, which describes nonlinear dynamics using a parallel combination of nonlinear and linear blocks. A Sigmoid network is used for the nonlinear part, and the estimated model is described as

$$\hat{y}_{k+1} = \hat{f}(\phi_k, \theta) = (\phi_k - r) \cdot P \cdot L + (1 + \exp(-(\phi_k - r) \cdot Q \cdot b - c))^{-1} \cdot a + d, \quad (22)$$

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

Training and validation results are both shown in Figure 15. From the figure, we can see that the identified NLARX model is able to capture the process dynamics and provide fairly reasonable estimate. A testing result on a set of fresh data is shown in Figure 16. From the figure, we can see that there is certain amount of mismatch between the lab data and the NLARX estimate as expected.

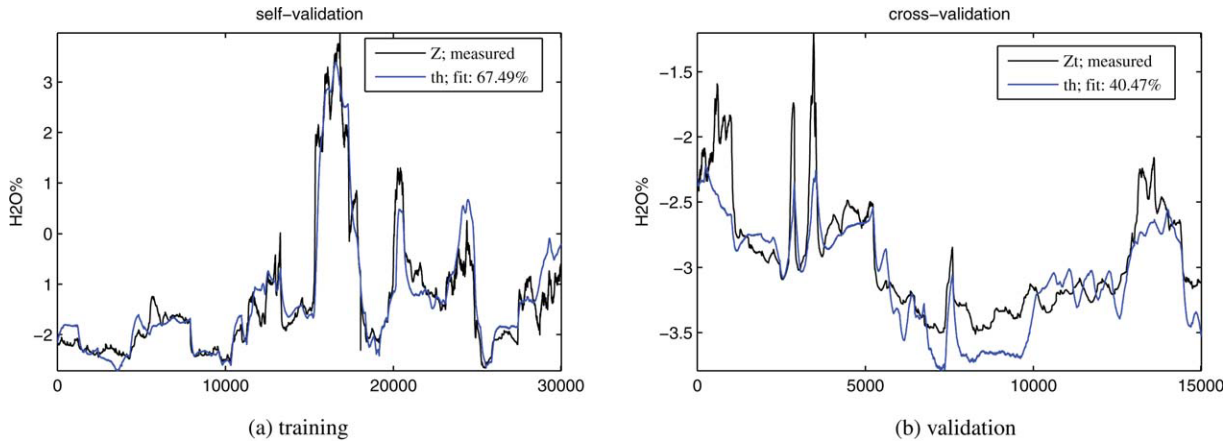

Figure 15. Model simulation and validation results for the investigated IPS unit.

[Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

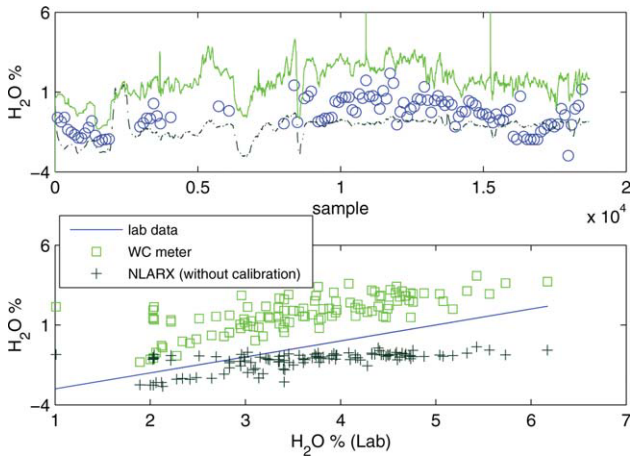

Figure 16. NLARX estimate without model calibration for new data set collected in July 2009.

[Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

Bayesian calibration

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

$$
\begin{aligned}
x_{x+1} &= \begin{bmatrix} 0 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} x_k + \begin{bmatrix} \rho_k \hat{f}(x_k, u_k, \theta_k) + \gamma_k \\ 0 \\ u_k \end{bmatrix} + \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} \omega_k^x, \\
\theta_{k+1} &= \theta_k + \omega_k^\theta, \\
\rho_{k+1} &= \rho_k + \omega_k^\rho, \\
\gamma_{k+1} &= \gamma_k + \omega_k^\gamma, \\
y_k^1 &= [1 \;\; 0 \;\; 0] x_k + v_k^1, \\
y_{T_k^2 k}^2 &= [1 \;\; 0 \;\; 0] x_{T_k^2 k} + v_{T_k^2 k}^2,
\end{aligned}
\tag{23}
$$

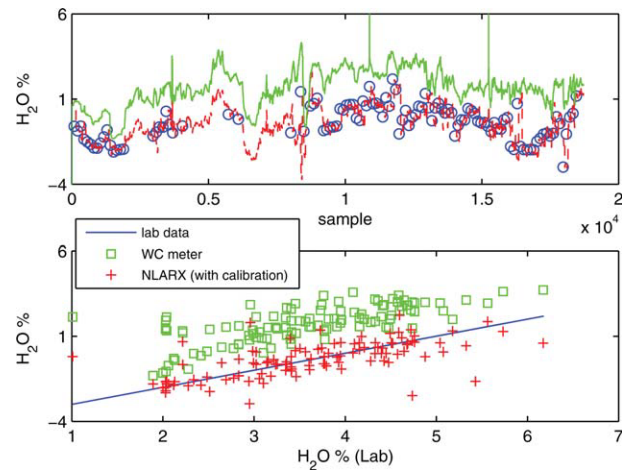

Figure 17. NLARX estimate with model calibration for new data set collected in July 2009.

[Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

Table 1. Performance Comparison for Water-Content Estimate

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

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

During the estimation, non-negativity (based on actual operating range) constraint is imposed on state variable x_k; 5% perturbations are added to depict the process and model uncertainties; observation noises are chosen as $v_k^1 \sim \mathcal{N}(0, 1^2)$ and $v_{T_k^2 k}^2 \sim \mathcal{N}(0, 005^2)$; parameters (based on virtual operating range) for sensor validation are set as $\alpha_1^1 = \alpha_1^2 = -3$, $\alpha_2^1 = \alpha_2^2 = 1$, $\beta_1^1 = \beta_1^2 = -4$, $\beta_2^1 = 4$, $\beta_2^2 = 6$; 100 particles are used for sequential Monte Carlo filtering.

Figure 17 shows the soft sensor result after model calibration. From the figure, we can see that the overall estimation performance has been improved significantly. As the modified posterior distribution has taken abnormal observations into account, the estimates are not affected by the sudden abnormal changes in water-cut meter readings.

Table 1 presents the comparisons of soft sensor estimates and water-cut meter readings with the lab data as the reference in terms of accuracy (i.e., mean absolute error, MAE), variability (i.e., standard deviation, STD), and overall performance (i.e., rooted mean square error, RMSE). We can clearly see that the soft sensor with model calibration provides the best prediction of the water content. This soft sensor has now been put on-line. Comparing to traditional measuring techniques (i.e., hardware sensor and lab analyzer), the developed soft sensor requires much less maintenance effort, thanks to the inclusion of model calibration strategy. Given the obtained benefits, more Bayesian soft sensors are planed for additional processes.

Conclusions

This article presents a practical approach for data-driven model calibration using multiple-source observations. The approach is built within a Bayesian framework to synthesize fast sampled but low accurate observations with high accurate but slow sampled observations to obtain more accurate process information. To enhance the robustness in the presence of abnormal data, a robust Bayesian fusion formulation with time-varying observation noise variance is proposed. A sequential Monte Carlo sampling based particle filter is then applied to carry out the Bayesian model calibration strategy. Compared to other approaches, the nature of sample based representation of PF facilitates the handling of constrained non-linear and non-Gaussian estimation problems. The proposed approach is used for a data-driven soft sensor development, which has been successfully demonstrated for water-content monitoring in an oil sands plant.

Acknowledgments

Financial support from the Natural Sciences and Engineering Research Council of Canada and Syncrude Canada Ltd. are gratefully acknowledged.

Literature Cited

1. Fortuna L1. *Soft sensors for monitoring and control of industrial processes*. Springer, London, UK, 2007.
2. Rao M, Corbin J, Wang Q. Soft sensors for quality prediction in batch chemical pulping processes. *In Proceedings of the IEEE International Symposium on Intelligent Control*. Chicago, IL, 1993.
3. Qin SJ, Yue H, Dunia R. Self-validating inferential sensors with application to air emission monitoring. *Ind Eng Chem Res*. 1997;36: 1675–1685.
4. Jos de Assis A, Filho R. Soft sensors development for on-line bioreactor state estimation. *Comput Chem Eng*. 2000;24:1099–1103.
5. Chen LZ, Nguang SK, Li XM, Chen XD. Soft sensors for on-line biomass measurements. *Bioprocess Biosyst Eng*. 2004;26:191–195.
6. Yan W, Shao H, Wang X. Soft sensing modeling based on support vector machine and Bayesian model selection. *Comput Chem Eng*. 2004;28:1489–1498.
7. Fortuna L, Graziani S, Xibilia M. Virtual instruments in refineries. *IEEE Instrum Meas Mag*. 2005;8:26–34.
8. Khatibisepehr S, Huang B. Dealing with irregular data in soft sensors: Bayesian method and comparative study. *Ind Eng Chem Res*. 2008;47:8713–8723.
9. Kadleca P, Gabrys B, Strandtb S. Data-driven soft sensors in the process industry. *Comput Chem Eng*. 2009;33:795–814.
10. Grantham SD, Ungar LH. A first principles approach to automated troubleshooting of chemical plants. *Comput Chem Eng*. 1990;14: 783–798.
11. Friedman YZ, Neto EA, Porfirio CR. First principles distillation inference models for product quality prediction. *Hydrocarbon Process*. 2002;81:53–58.
12. MacGregor J. Data-based latent variable methods for process analysis, monitoring and control. *Comput Aided Chem Eng*. 2004; 18:87–98.
13. Kano M, Nakagawa Y. Data-based process monitoring, process control, and quality improvement: recent developments and applications in steel industry. *Comput Chem Eng*. 2008;32:12–24.
14. Huang B. Bayesian methods for control loop monitoring and diagnosis. *J Process Control*. 2008;10:829–838.
15. Doucet A, Godsill S. On sequential simulation-based methods for Bayesian filtering. Technical Report, 1998.
16. Wang J, Chen T, Huang B. Multirate sampled-data systems: computing fast-rate models. *J Process Control*. 2004;4:79–88.
17. Ljung L. In: Kailath T editor. *System Identification: Theory for the User, 2nd ed*. Englewood Cliffs: Prentice Hall, 1999.
18. Chen R, Tsay R. Nonlinear additive ARX models. *J Am Stat Assoc*. 1993;88:955–967.
19. Bai E. An optimal two-stage identification algorithm for Hammerstein-Wiener nonlinear systems. *Automatica*. 1998;34:333–338.
20. Björck A. *Numerical Methods for Least Squares Problems*. SIAM, Philadelphia, PA, 1996.
21. Nelson P, Taylor P, MacGregor J. Missing data methods in PCA and PLS: score calculations with incomplete observations. *Chemom Intell Lab Syst*. 1996;35:45–65.
22. Dayal B, MacGregor J. Improved PLS algorithms. *J Chemom*. 1997;11:73–85.
23. Kresta J, Marlin T, MacGregor J. Development of inferential process models using PLS. *Comput Chem Eng*. 1994;18:597–611.
24. Bishop CM. *Neural networks for pattern recognition*. USA: Oxford University Press, 1995.
25. Nagai E, Arruda L. Soft sensor based on fuzzy model identification. 16th IFAC World Congres., Prague, Czech Republic, 2005.
26. Zhu Y, Telkamp H, Wang J, Fu Q. System identification using slow and irregular output samples. *Journal Process Control*. 2009; 19:58–67.
27. Lu W, Fisher D. Output estimation with multi-rate sampling. *Int J Control*. 1988;48:149–160.
28. Lu W, Fisher D. Least-squares output estimation with multirate sampling. *IEEE Trans Auto Cont*. 1989;34:669–672.
29. Li D, Shah SL, Chen T. Identification of fast-rate models from multirate data. *Int J Control*. 2001;74:680–689.
30. Ding F, Chen T. Identification of dual-rate systems based on finite impulse response model. *Int J Adaptive Control Signal Process*. 2004;18:589–598.
31. Raghavan H, Tangirala AK, Gopaluni RB, Shah SL. Identification of chemical processes with irregular output sampling. *Control Eng. Pract*. 2006;14:467–480.
32. Mo S, Chen X, Zhao J, Qian J, Shao ZA. A Two-stage method for identification of dual-rate systems with fast input and very slow output. *Ind Eng Chem Res*. 2009;48:1980–1988.
33. Lu N, Yang Y, Guao F, Wang W. Multirate dynamic inferential modeling for multivariable processes. *Chem Eng Sci*. 2004;59: 855–864.
34. Tun M, Lakshminarayanan S, Emoto G. Data selection and regression method and its application to softsensing using multirate industrial data. *J Chem Eng Japan*. 2008;41:374–383.
35. Akaike H. A new look at the statistical model identification. *IEEE Trans Auto Cont*. 1974;19:716–723.
36. Xiong Y, Chen W, Tsui K-L, Apley DW. A better understanding of model updating strategies in validating engineering models. *Comput Methods Appl Mech Eng*. 2009;198:1327–1337.
37. Singh A. Modeling and Model Updating in the Real-Time Optimization of Gasoline Blending, M.Sc. thesis, University of Toronto, 1997.
38. Mu S, Zeng Y, Liu R, Wu P, Su H, Chu J. Online dual updating with recursive PLS model and its application in predicting crystal size of purified terephthalic acid (PTA) process. *J Process Control*. 2006;16:557–566.
39. Kewley D. Notes on the use of Dempster-Shafer and Fuzzy Reasoning to fuse identity attribute data. Technical Report, 1992.
40. Braun J. Dempster-Shafer theory and Bayesian reasoning in multisensor data fusion. *Sensor Fusion: Architectures, Algorithms and Applications IV; Proceedings of SPIE 4051*. Orlando, FL, 2000.
41. Koks D, Challa S. An introduction to Bayesian and Dempster-Shafer data fusion. Technical Report, 2003.
42. Saha RK, Chang KC. An efficient algorithm for multisensor track fusion. *IEEE Trans Aerospace Electron Syst*. 1998;43:200–210.
43. Gan Q, Harris CJ. Comparison of two measurement fusion methods for Kalman-filter-based multisensor data fusion. *IEEE Trans Aerospace Electron Syst*. 2001;37:273–279.
44. Gordon N, Salmond D, Smith A. Novel approach to nonlinear/non-Gaussian Bayesian state estimation. *IEE Proc F Radar Signal Process*. 1993;140:107–113.
45. Punska O. Bayesian Approaches to Multi-Sensor Data Fusion, M.Sc. thesis, Department of Engineering, University of Cambridge, 1999.
46. Hua G, Wu Y. Measurement integration under inconsistency for robust tracking. Proceedings of the 2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. New York, 2006.
47. Shao X, Huang B, Lee JM. Constrained Bayesian State Estimation - a Comparative Study and a New Particle Filter Based Approach. *J Process Control*. 2010;20:143–157.
48. Arulampalam MS, Maskell S, Gordon N, Clapp T. A tutorial on particle filters for online nonlinear/non-gaussian Bayesian tracking. *IEEE Trans Signal Process*. 2002;50:174–188.
49. Lang L, Chen W, Bakshi B, Goel P, Ungarala S. Bayesian estimation via sequential Monte Carlo sampling-Constrained dynamic systems. *Aotomatica*. 2007;43:1615–1622.
50. van der Merwe R, Doucet A, de Freitas N, Wan E. The unscented particle filter; Technical Report, 2000.
51. Boloc M, Djuric P, Hong S. Resampling algorithms for particle filters: a computational complexity perspective. *EURASIP J Applied Signal Process*. 2004;15:2267–2277.
52. Gudi R, Shah S, Gray M. Adaptive multirate state and parameter estimation strategies with application to a bioreactor. *AIChE*. 1995; 41:2451.
53. Scholz E. *Karl Fischer Titration*. Springer, Berlin, 1984.

Manuscript received Jan. 19, 2010, and revision received Jun. 12, 2010.