Dynamic Data Rectification Using the Expectation Maximization Algorithm

Ashish Singhal and Dale E. Seborg
Dept. of Chemical Engineering, University of California, Santa Barbara, CA 93106

Although on-line measurements play a vital role in process control and monitoring process performance, they are corrupted by noise and occasional outliers (such as noise spikes). Thus, there is a need to rectify the data by removing outliers and reducing noise effects. Well-known techniques such as Kalman Filtering have been used effectively to filter noise measurements, but it is not designed to automatically remove outliers. A new methodology based on the Kalman filter rectifies noise as well as outliers in measurements. Filter equations were formulated in the form of probability distributions. Then the Expectation-Maximization algorithm was used to find the maximum-likelihood estimates of the true measurement values based on a state-space model, past data, and current observations. This approach was evaluated when the assumption of normally distributed outliers is not valid. The method can be used with any dynamic process model, as shown by integrating it with an extended Kalman Filter and by an augmented linear state-space model to account for unmeasured disturbances. It also can be used to provide diagnostic information about changes to the process or sensor failures.

Introduction

Filtering of data refers to finding an estimate of the true value of the measured variable based upon past measurements. An *optimal* estimator is a computational algorithm that processes measurements to generate an optimal estimate of the state of the process, based on some stated criterion of optimality. The estimator is based on process knowledge in the form of a dynamic model and assumed statistics of process noise and measurement errors (Gelb, 1974). The Kalman filter is the most widely used optimal filtering technique (Kalman, 1963). Although the Kalman filter produces optimal estimates of the true state of the system, it is not designed to remove gross errors or outliers from data. Other algorithms for removing gross errors, such as a velocity filter, remove outliers based on the change of the measurement from the previous sample. This approach does not take into account the variability of the process due to changes in the input variables. Tong and Crowe (1995, 1997) have used principal component analysis (PCA) to detect gross errors in measurement data for steady-state conditions, but it may not be suitable for dynamic processes where there are sufficient variations in variables. What is required, therefore, is a technique to define an outlier in some statistical sense for a dynamic process, and to rectify the data using an optimal estimator (Kalman filter), once the outlier has been detected.

It is possible to find optimal estimates of the true state if it is known *a priori* that the measurements contain outliers that have known (or assumed) statistical properties. If the measurement data are referred to as the incomplete data, and the set of measurement data with information about the occurrence of an outlier for each measurement as the complete data, the Expectation-Maximization (EM) algorithm (Dempster et al., 1977) can be used to estimate the missing information as well as the optimal estimates of the true state of the process using the Kalman filter.

Probabilistic Formulation of the Data-Rectification Problem

The equations for data rectification will now be formulated using probability distribution function (PDFs). An advantage of PDFs is that they are capable of capturing the full range of knowledge about a variable or parameter, from absolute

Correspondence concerning this article should be addressed to D. E. Seborg.

certainty to complete ignorance. Also, it is convenient to represent knowledge in the form of probabilities for use in the EM algorithm. The Gaussian probability distribution will be used for the development of the EM algorithm. For a continuous random variable x, with mean μ, and covariance matrix Σ, it will be represented as

$$p(x) = N(x;\, \mu, \Sigma). \tag{1}$$

If the probability distribution of process and measurement noise is Gaussian, then the minimum variance estimate of the true value of the measurement is also its optimal estimate (Gelb, 1974). The following subsection develops the equations for the rectification of data containing outliers. The application of the EM algorithm to the data-rectification problem is then described. The next subsection demonstrates the equivalence of the Kalman filter and EM estimates for the special case of no outliers being present.

Rectification of data containing random outliers

Assume that errors in $y(k)$ are additive in nature, that is,

$$z(t) = y(t) + \delta(t), \tag{2}$$

where

$z(k)$ = measured variable at instant k
$y(k)$ = true (noise free) value of the measured variable
$\delta(k)$ = error or noise vector with a probability distribution

$$p[\delta(t)] = N[\delta(t); 0, \Sigma] \tag{3}$$

Using past data and a dynamic model, the one-step-ahead prediction of $y(k)$ can be made, which has a Gaussian distribution with mean $y_f(k)$ and a covariance matrix $\Sigma_f(k)$. The measured variable $z(k)$ is also a Gaussian distribution with mean $y_f(k)$ and covariance matrix Σ.

If the measurements contain random outliers, than a single probability distribution described by Eq. 3 cannot account for the high variance of the noise spikes. This problem is overcome by assuming that measurement noise is sampled from two probability distributions, one having a small covariance representing regular noise, and the other having a large covariance representing noise spikes or outliers (Schick and Mitter, 1994; Thompson, 1996).

$$p(\delta(k)) = (1-\epsilon)\, N[\delta(k); 0, \Sigma] + \epsilon N[\delta(k); 0, b^2\Sigma], \tag{4}$$

where ϵ is the known prior probability of the occurrence of an additive outlier (noise spike or outlier), Σ is the sensor noise covariance, and b^2 is the assumed multiplying factor for the covariance of the outlier distribution such that $b \gg 1$. The measurements are assumed to be distributed as $N(z(k); y_f(k), \Sigma)$. Dropping the k dependence for convenience, the distribution $p(y|z)$ is given by Bayes' theorem:

$$\begin{aligned} p(y|z) &= \frac{p(y)\,p(z|y)}{p(z)} = \frac{p(y)\,p(z-y)}{p(z)} = \frac{p(y)\,p(\delta)}{p(z)} \\ &= \frac{N(y; y_f, \Sigma_f)\left[(1-\epsilon)\,N(\delta; 0, \Sigma) + \epsilon N(\delta; 0, b^2\Sigma)\right]}{N(z; y_f, \Sigma)} \\ &= \underbrace{\frac{(1-\epsilon)\,N(y; y_f, \Sigma_f)\,N(\delta; 0, \Sigma)}{N(z; y_f, \Sigma)}}_{\phi_1(y)} \\ &\quad + \underbrace{\frac{\epsilon N(y; y_f, \Sigma_f)\,N(\delta; 0, b^2\Sigma)}{N(z; y_f, \Sigma)}}_{\phi_2(y)}. \end{aligned} \tag{5}$$

The estimate of $y(k)$, which is denoted by $\hat{y}(k)$, is obtained by maximizing $p(y|z)$, and is referred to as the *rectified value* (Thompson, 1996). The problem is to determine the best estimates of $y(k)$, given corrupted observations $z(k)$.

If a probability density function $p(y)$ of a continuous random variable y can be represented as a finite mixture of basis density function $\phi_j(y)$, such that,

$$p(y) = \sum_j q_j \phi_j(y) \tag{6}$$

and there exist binary indicator variables θ_j such that one and only one $\theta_j = 1$, then the joint distribution $p(y, \theta)$ is given as (Thompson, 1996),

$$p(y, \Theta) = \prod_j \left[q_j \phi_j(y)\right]^{\theta_j} \tag{7}$$

$$\Theta = [\theta_1\ \theta_2 \cdots \theta_n]. \tag{8}$$

If $\zeta(k)$ is defined as a binary variable such that

$$\zeta(k) = \begin{cases} 0 & \text{if the error } \delta(k) \text{ is sampled from the distribution } \phi_1(y), \\ 1 & \text{if the error } \delta(k) \text{ is sampled from the distribution } \phi_2(y), \end{cases} \tag{9}$$

then, $\theta_1 \equiv \zeta$ and $\theta_2 \equiv 1 - \zeta$. Equation 5 is then written as

$$p(y, \zeta|z) = \left[\phi_1(y)\right]^{1-\zeta}\left[\phi_2(y)\right]^{\zeta}. \tag{10}$$

The indicator variable ζ can also be interpreted as the posterior probability of z being an outlier. Thompson (1996) has reported that under the assumed distributions for $p(y)$ and $p(\delta)$, $p(y|z)$ is a finite mixture of Gaussians that is maximized at the same y as its joint distribution, $p(y, \zeta|z)$.

Expectation-Maximization algorithm

The EM algorithm proposed by (Dempster et al., 1977) can be used to obtain maximum likelihood (ML) estimates of the desired information using incomplete information about the data. This requires that information be available in the form of continuous and differentiable probability distributions, which are maximized to obtain the ML estimates. For the present case of data rectification, the measurements z are called incomplete data, and the set $\{z, \zeta\}$ is called the complete data because it provides information whether the data point is an outlier or not. The EM algorithm finds the expected value of indicator variable (ζ), using the available data and their probability distribution. The reader may refer to Moon (1996) and Johnston and Kramer (1995, 1998) for other applications of the EM algorithm.

The EM algorithm is used to estimate $\hat{y}$ by maximizing Eq. 10. The algorithm consists of two steps: an *expectation* step, followed by a *maximization* step. In Eq. 10, y is the variable to be optimized upon the observed value z and the indicator ζ. The optimum value of y, for which $p(y, \zeta | z)$ is maximized in Eq. 10, is found by iteration. In the expectation step, the expected value of ζ is found using the current value $\hat{y}$. Then the expected value of ζ is used to calculate $\hat{y}$, which maximizes the log-likelihood function, $\log[p(y, \zeta | z)]$. The value $\hat{y}$ is found by taking the gradient of the log-likelihood function with respect to y, and then setting it equal to zero.

$$\nabla_y\{\log[p(y,\zeta|z)]\}$$

$$= \nabla_y\left\{(1-\zeta)\log\left[(1-\epsilon)\frac{N(y;y_f,\Sigma_f)\,N(\delta;0,\Sigma)}{N(z;y_f,\Sigma)}\right]\right\}$$

$$+ \nabla_y\left\{\zeta\log\left[\epsilon\frac{N(y;y_f,\Sigma_f)\,N(\delta;0,b^2\Sigma)}{N(z;y_f,\Sigma)}\right]\right\} = 0. \quad (11)$$

It can be easily verified that the solution of Eq. 11 produces a global maximum. Note that the term $N(z; y_f, \Sigma)$ is a constant at each time instant k. After simplifying the algebra, the solution of Eq. 11 is found as

$$\hat{y} = \left[\Sigma_f^{-1} + \left(1 - \zeta + \frac{\zeta}{b^2}\right)\Sigma^{-1}\right]^{-1} \times \left[\Sigma_f^{-1} y_f + \left(1 - \zeta + \frac{\zeta}{b^2}\right)\Sigma^{-1} z\right]. \quad (12)$$

The iterative EM procedure is now described as follows:

1. Start with $\hat{y} = y_f$.
2. *E-Step:* Calculated the expected value of ζ as

$$E[\zeta] = \bar{\zeta} = \frac{\phi_2(\hat{y})}{\phi_1(\hat{y}) + \phi_2(\hat{y})}. \quad (13)$$

Note that the expected value of a discrete variable can lie between two of its discrete values ($0 \le \bar{\zeta} \le 1$).

3. *M-Step*: Using this $\bar{\zeta}$, find

$$\hat{y} = \left[\Sigma_f^{-1} + \left(1 - \bar{\zeta} + \frac{\bar{\zeta}}{b^2}\right)\Sigma^{-1}\right]^{-1} \times \left[\Sigma_f^{-1} y_f + \left(1 - \bar{\zeta} + \frac{\bar{\zeta}}{b^2}\right)\Sigma^{-1} z\right]. \quad (14)$$

4. If $\hat{y}$ has converged to a value within a specified tolerance limit, then stop. Otherwise go to step 2.

The observation is labeled as an outlier if $\bar{\zeta}$ is greater than 0.5.

Remarks. The convergence of the EM algorithm is guaranteed when the probability distributions are Gaussian (Abraham and Chuang, 1993; Xu and Jordan, 1996).

Equivalence of Kalman filter and EM estimates

The minimum variance equations of the Kalman filter produce an estimate identical to the minimum variance Bayes estimate (Gelb, 1974), which can be obtained by minimizing the loss function

$$J = \int (\hat{y} - y)^T \Lambda (\hat{y} - y)\, p(y|z)\, dy \quad (15)$$

$$p(y|z) = \frac{N(y; y_f, \Sigma_f)\, N(y; z, \Sigma)}{N(z; y_f, \Sigma)}, \quad (16)$$

where Λ is an arbitrary positive semidefinite matrix, and $\hat{y}$ is an estimate of y. The numerator of Eq. 16 can be simplified by using the formula for multiplication of two normal distributions (Mardia et al., 1979; Thompson, 1996),

$$N(y; \mu_1, \Sigma_1)\, N(y; \mu_2, \Sigma_2) = N(\mu_2; \mu_1, \Sigma_1 + \Sigma_2)\, N(y; \mu, S) \quad (17)$$

$$S = \left[\Sigma_1^{-1} + \Sigma_2^{-1}\right]^{-1} \quad (18)$$

$$\mu = S\left[\Sigma_1^{-1}\mu_1 + \Sigma_2^{-1}\mu_2\right]. \quad (19)$$

Setting $\partial J/\partial y = 0$, and removing all constants, one can find the estimate $\hat{y}$, independent of Λ as,

$$\hat{y} = \int y N(y; \mu, S)\, dy = \mu = \left[\Sigma_1^{-1} + \Sigma_2^{-1}\right]^{-1}\left[\Sigma_1^{-1}\mu_1 + \Sigma_2^{-1}\mu_2\right] \equiv \left[\Sigma_f^{-1} + \Sigma^{-1}\right]^{-1}\left[\Sigma_f^{-1} y_f + \Sigma^{-1} z\right]. \quad (20)$$

The minimum variance Kalman filter estimate in Eq. 20 is equal to the ML estimate if the probability distributions for the prediction and noise are Gaussian (Gelb, 1974). This ML estimate is *exactly* equal to the EM estimate in Eq. 14 for the case when no outliers are present ($\bar{\zeta} = 0$). Also, Eq. 20 indicates that the estimate $\hat{y}$ is a weighted mean of the prediction y_f and the measurement z. Because the covariance of the prediction is greater than or equal to the covariance of noise (Gelb, 1974), the prediction is always weighted less than the measurement in Eq. 20. This is the reason why the Kalman filter is unable to rectify data containing outliers.

Discrete Kalman filter with expectation maximization

The standard Kalman filter equation with the EM steps are written for the linear state-space model described by

$$x(k+1) = Ax(k) + Bu(k) + Gw(k)$$
$$y(k) = Cx(k) + Du(k) + v(k), \tag{21}$$

where $x(k)$, $u(k)$, and $y(k)$ are the state, the input, and the output of the system at sampling instant k; $w(k)$ is white process noise; and $v(k)$ is measurement noise. Additional information about the model is given as:

$$\hat{x}(0) = x_0 \tag{22}$$
$$E\left[(x(0)-x_0)(x(0)-x_0)^T\right] = P_0 \tag{23}$$
$$E[w(k)] = E[v(k)] = 0 \tag{24}$$
$$E\left[w(k)w(k)^T\right] = Q \tag{25}$$
$$E\left[v(k)v(k)^T\right] = R \tag{26}$$
$$E\left[w(k)v(j)^T\right] = 0 \quad \forall k, j \tag{27}$$
$$E\left[(z(k)-\hat{y}(k))(z(k)-\hat{y}(k))^T\right] = \Sigma_f(k) \tag{28}$$
$$\Sigma_f(0) = CP_0C^T + R, \tag{29}$$

where $\hat{x}$ refers to the estimated quantity. The prediction and update equations are written as (Gelb, 1974)

$$x_f(k+1) = A\hat{x}(k) + Bu(k) \tag{30}$$
$$y_f(k+1) = Cx_f(k+1) + Du(k+1) \tag{31}$$
$$P_f(k+1) = AP(k)A^T + GQG^T \tag{32}$$
$$\Sigma_f(k+1) = CP_f(k+1)C^T + R \tag{33}$$
$$\hat{y}_{EM}(k+1) = EM\left(y_f(k+1), \Sigma_f(k+1), z(k+1), R\right) \tag{34}$$
$$K(k+1) = P_f(k+1)C^T\left[CP_f(k+1)C^T + R\right]^{-1} \tag{35}$$
$$\hat{x}(k+1) = \begin{cases} x_f(k+1) + K(k+1)\left[z(k+1) - y_f(k+1)\right] \\ \qquad \text{if } \hat{\zeta} < 0.5 \\ x_f(k+1) + K(k+1)\left[\hat{y}_{EM}(k+1) - y_f(k+1)\right] \\ \qquad \text{if } \hat{\zeta} \geq 0.5 \end{cases} \tag{36}$$
$$P(k+1) = \left[I - K(k+1)C\right]P_f(k+1), \tag{37}$$

where, the function EM $(y_f(k+1), \Sigma_f(k+1), z(k+1), R)$ returns the rectified value of the measurement $z(k+1)$ using the EM algorithm described earlier. To obtain better state estimates in the presence of outliers, the state update equation uses the rectified value $\hat{y}_{EM}(k+1)$ instead of the measurement $z(k+1)$ when $\bar{\zeta} \geq 0.5$ The rectification scheme is represented as Figure 1.

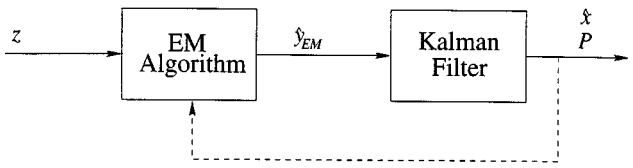

Figure 1. Kalman filter with expectation maximization.

Continuous-discrete extended Kalman filter with expectation maximization

Usually the model of the nonlinear plant is available in continuous-time form, and sampling of data is at discrete times. The nonlinear continuous-time plant model with discrete sampling is written as

$$\frac{dx}{dt} = f\left[x(t), u(t), w(t)\right]$$
$$y(k) = h_k\left(x(t_k), u(t_k)\right) + v(k)$$
$$t_k = k\Delta t, \tag{38}$$

where Δt is the sampling period. Assuming the initial conditions to be the same as Eqs. 22–29, the extended Kalman filter (EKF) equations (Gelb, 1974) with EM are written as

$$F = \left[\frac{\partial f\left(x(t), u(t_k), 0\right)}{\partial x(t)}\right]_{x(t)=\hat{x}(t)} \tag{39}$$
$$G = \left[\frac{\partial f\left(x(t_k), u(t_k), w(t)\right)}{\partial w(t)}\right]_{w(t)=0} \tag{40}$$
$$H = \left[\frac{\partial h\left(x(t), u(t_k), 0\right)}{\partial x(t)}\right]_{x(t_k)=x_f(k+1)} \tag{41}$$
$$\frac{d\hat{x}(t)}{dt} = f\left(\hat{x}(t), u(t_k), 0\right) \tag{42}$$
$$\frac{dP(t)}{dt} = FP(t) + P(t)F^T + GQG^T \tag{43}$$
$$y_f(k+1) = Hx_f(k+1) \tag{44}$$
$$\Sigma_f(k+1) = HP_f(k+1)H^T + R \tag{45}$$
$$\hat{y}_{EM}(k+1) = EM\left(y_f(k+1), \Sigma_f(k+1), z(k+1), R\right) \tag{46}$$
$$K(k+1) = P_f(k+1)H\left[HP_f(k+1)H^T + R\right]^{-1} \tag{47}$$

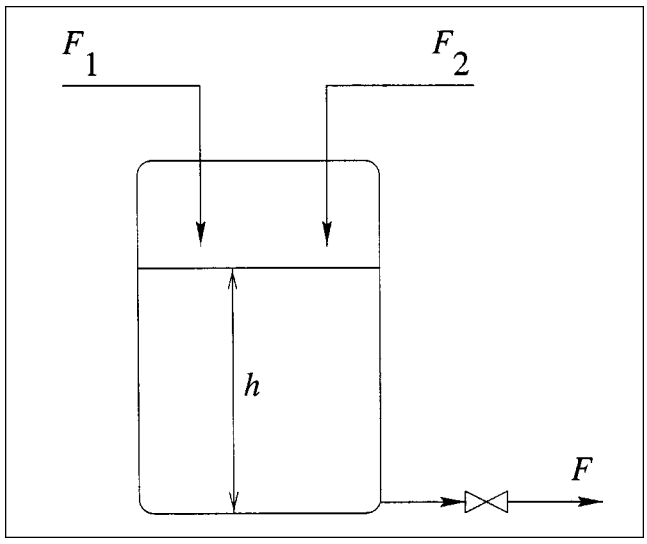

Figure 2. Liquid storage tank.

$$\hat{x}(k+1) = \begin{cases} x_f(k+1) + K(k+1)\left[z(k+1) - y_f(k+1)\right] \\ \qquad \text{if } \bar{\zeta} < 0.5 \\ x_f(k+1) + K(k+1)\left[\hat{y}_{EM}(k+1) - y_f(k+1)\right] \\ \qquad \text{if } \bar{\zeta} \geq 0.5 \end{cases} \tag{48}$$

$$P(k+1) = \left[I - K(k+1)H\right]P_f(k+1). \tag{49}$$

Equations 42 and 43 are integrated through time to estimate $x_f(k+1)$ and $P_f(k+1)$, respectively.

Simulation Examples and Results

Two physical examples are used to study the effectiveness of the EM algorithm in removal outlier and noise reduction. One example is a linear liquid storage system, and the other is a nonlinear continuous stirred-tank reactor (CSTR) system.

Example 1: Liquid storage tank system

Consider a tank of uniform cross-sectional area A, as shown in Figure 2. The tank is filled by two inlet streams F_1 and F_2, and the liquid leaves the bottom of the tank through a linear valve with a constant valve factor C_v. The flow out of the tank is directly proportional to the height h of liquid in the tank. The flow of stream F_2 is affected by upstream pressure fluctuations, and it is proportional to this upstream pressure. A dynamic material balance for the liquid in the tank gives

$$A\frac{dh}{dt} = F_1 + F_2 - C_v h. \tag{50}$$

The nominal values of process variables and constants are given in Table 1. For the preceding system, F_1 is the deterministic input and F_2 is a disturbance due to upstream pressure fluctuations. In particular, F_2 is a white-noise sequence of variance $10^{-3}[\text{m}^3/\text{min}]^2$. This system is discretized for a sampling period $\Delta t = 0.05$ min as

Table 1. Nominal Values and Constants for Example 1

$\bar{h}$	$\bar{F}_1$	$\bar{p}$	K	C_v	A
2 m	1 m^3/min	0.049 bar	20.504 $\text{m}^3/\text{bar}\cdot\text{min}$	1 m^2/min	1 m^2

$$x(k+1) = 0.951\,x(k) + 0.049\,u(k) + w(k) \tag{51}$$

$$y(k) = x(k) + v(k), \tag{52}$$

where x is the state variable h, u is the deterministic input F_1, $w(k)$ is process noise F_2, and y is the noisy measurement of h. The variance of measurement noise $v(k)$ is 10^{-2} cm^2. The deviation of each variable from its nominal value is denoted with a Δ prefix, for example Δh, and so on.

Outliers Having a Gaussian Distribution. The outliers in $v(k)$ are sampled from a normal distribution with zero mean and variance much larger than the noise variance (about 100 times greater). It is assumed that 10% of the measurements contain outliers. Therefore, for all simulations $\epsilon = 0.1$ and $b = 10$ are assumed as typical values. They could also be estimated from prior information about the sensor, for example, from maintenance records or mean time between failures.

The simulation results in Figure 3 indicate that the Kalman filter with EM is able to rectify each data point that contains

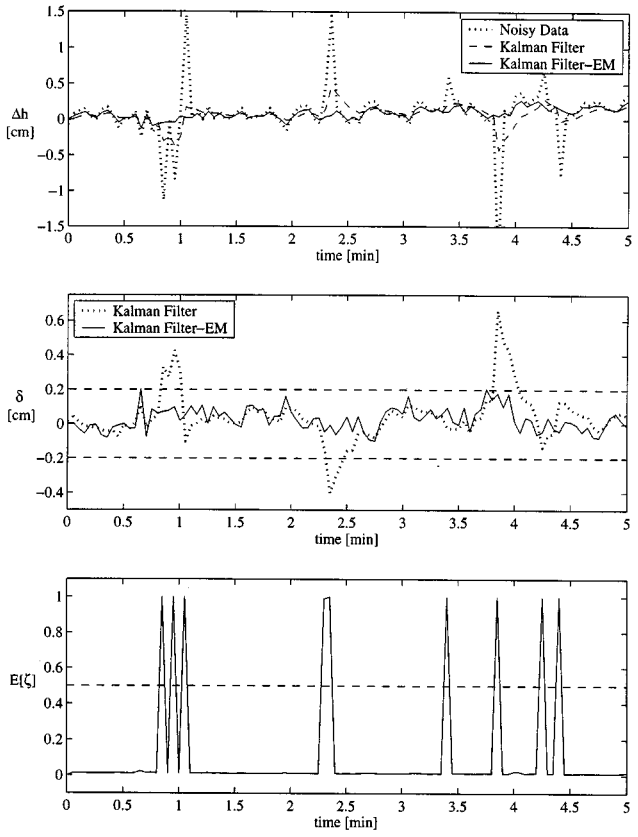

Figure 3. Simulation results for Example 1.

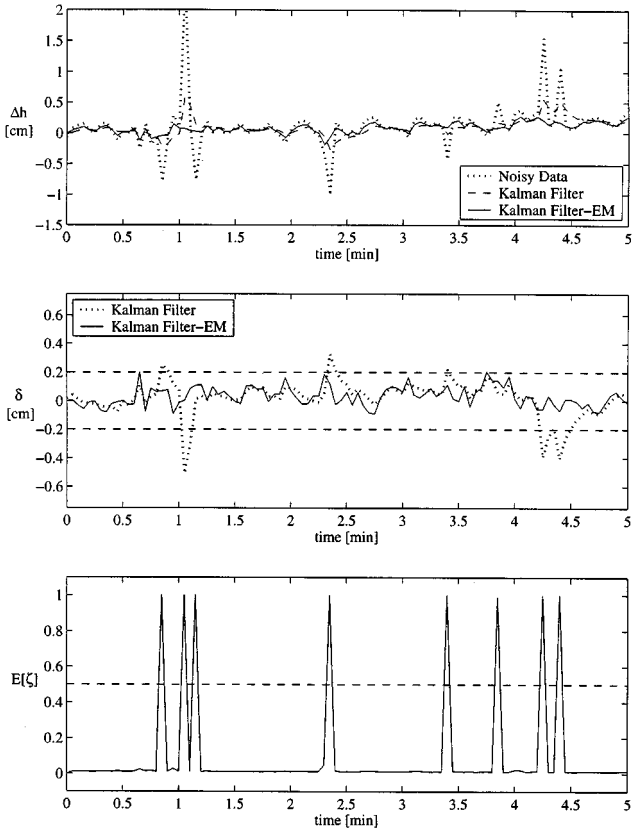

Figure 4. Simulation results for Example 1 with outliers having an exponential distribution.

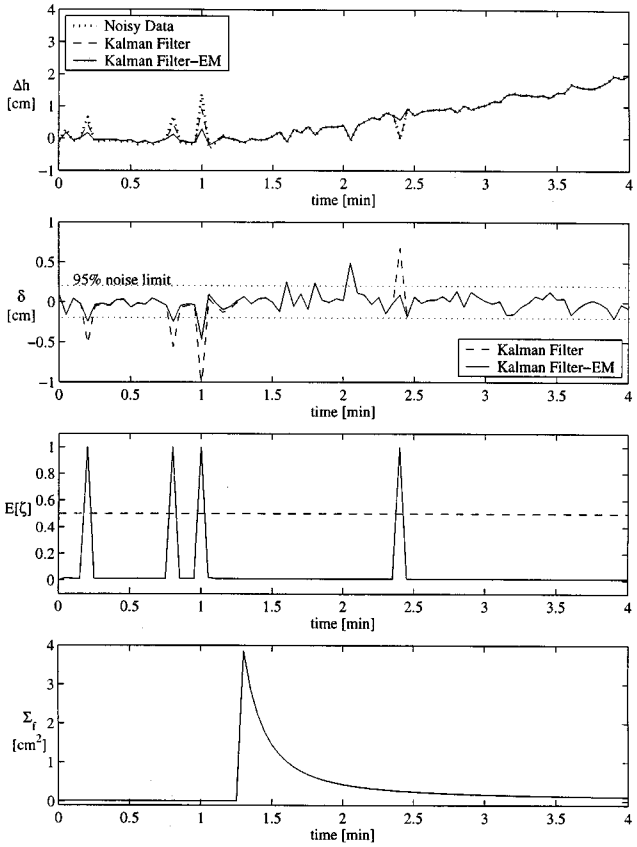

Figure 5. Simulation results for ramp disturbance at input for Example 1.

an outlier. Also rectification error, $\delta(k) = y(k) - \hat{y}(k)$, for the Kalman filter–EM is within the 95% noise limits, while the Kalman filter alone produces large rectification errors when outliers are present. The measurements that contain outliers are apparent from the expected value of the indicator variable ζ, as shown in Figure 3.

Outliers Having an Exponential Distribution. In practical applications the outliers can have a non-Gaussian probability distribution. The simulation for the case when the outliers are sampled from an exponential distribution with parameter $\lambda = 1$ is presented in Figure 4. Even though the assumption of normally distributed outliers is violated, the rectification using the Kalman filter–EM algorithm remains unaffected. The EM labels the data point as an outlier if it falls outside the range of regular noise; therefore, it does not matter whether the outlier comes from a Gaussian or some other distribution. All that is required is that the noise spike lies beyond the limits of regular noise. However, the estimate obtained in this case may not be the optimal estimate, because the noise distributions are not Gaussian (Gelb, 1974).

Ramp Disturbance. Here, the methodology of Juricek et al. (1998) is used to predict the future outputs, assuming a ramp disturbance with unknown slope is added to the process noise, *w*. For simplicity, the time at which the ramp starts is assumed to be known. A ramp of slope 0.05 m^3/min^2 is added to the process noise starting at $t = 1$ min with a sampling time $\Delta t = 0.05$ min. The disturbance introduced into the process is expressed mathematically as,

$$w(t) = \begin{cases} q(t) & 0 \le t < 1\,\text{min} \\ 0.05t + q(t) & t \ge 1\,\text{min} \end{cases} \qquad (53)$$

$$q(t) \equiv N(0, 10^{-3}).$$

Because $w(t)$ is an unmeasured disturbance, it is represented as a pseudodisturbance $d(k)$ in the state-space representation of the system as (Juricek et al., 1998),

$$x(k+1) = Ax(k) + Bu(k) + d(k). \qquad (54)$$

For a first-order system, $d(k)$ may be written as

$$d(k) = \begin{bmatrix} 1 & (k - k_0)\Delta t \end{bmatrix} \begin{bmatrix} d_0 \\ d_s \end{bmatrix}. \qquad (55)$$

The parameters of the disturbance, d_0 and d_s, are estimated using linear regression, as described by Juricek et al. (1998), for making future predictions. The simulation results are presented in Figure 5. The outliers can be detected by observing $E[\zeta]$ in Figure 5. The rectification error is large for outliers occurring at the beginning of the ramp, because the covari-

ance of the predictions is large, but decreases as more measurements are added. This again demonstrates that the methodology is quite general. The only requirements are an accurate dynamic model with a known forecast covariance, measurements with known variance, and assumed parameters ϵ and b for the EM algorithm.

Detection of Change using EM Algorithm

In the previous section a disturbance model was included in the existing state-space model estimate of the unknown ramp disturbance. It is possible to detect the change caused by a disturbance by observing the expected value of ζ, because ζ has a Bernoulli distribution with parameters p and n, where n is the window size of observations (Abraham and Chuang, 1993). For a single measurement, $p = \epsilon$, where ϵ is the known prior probability of occurrence of an outlier in Eq. 4.

By observing the expected value of ζ for a window of n samples, and counting the number of times that $E[\zeta] > 0.5$ in the current window, one can estimate the maximum number of outliers allowed at the $100(1-\alpha)\%$ confidence level for this window using the prior probability p of observing an outlier. As stated earlier, $p = \epsilon$ for a single output system. If r is the number of outliers in a window size of n, then

$$r_{100(1-\alpha)} = \max r$$

$$\text{such that } \sum_{k=0}^{r} \binom{n}{k} p^k (1-p)^{n-k} < 1\text{-}\alpha \qquad (56)$$

is the maximum number of outliers allowed at the $100(1-\alpha)\%$ confidence level. The time of detection of change is the instant when $r > r_{100(1-\alpha)}$ for that observation window. The window size n can be chosen greater than $\Theta(\epsilon^{-1})$, so that the expected number of outliers in the window is at least one.

In Figure 6 a ramp disturbance is introduced at time $t = 2$ min for Example 1. The EM algorithm and a model that does not incorporate a disturbance term is used to rectify the data. Due to the ramp disturbance, there is a significant mismatch between model predictions and measurement data. This leads to each observation being labeled as an outlier. The number of outliers in a window size of 10 samples increases to more than 3 (which is the 99% limit) after $t = 5.2$ min. This time can therefore be called the *detection time* t_c, for the change.

In Figure 7, a sensor failure at $t = 4$ min is introduced in the measurement of h, and the sensor stays constant at the last measurement. In this case, too, the model predictions deviate significantly from the measurements, and therefore all observations after $t = 4.5$ min are labeled as outliers. Thus the detection time $t_c = 4.5$ min.

Example 2: nonisothermal CSTR

This system consists of a stirred tank of cross-sectional area A, with cooling surface area A_C. A single feed stream containing a reactant A with flow rate of Q_F and concentration C_{AF} enters the tank as shown in Figure 8. The reactant A undergoes a first-order exothermic reaction A $\rightarrow$ products. The reactor contents are cooled by a coolant flowing through the jacket around the tank. Based on a relative gain array

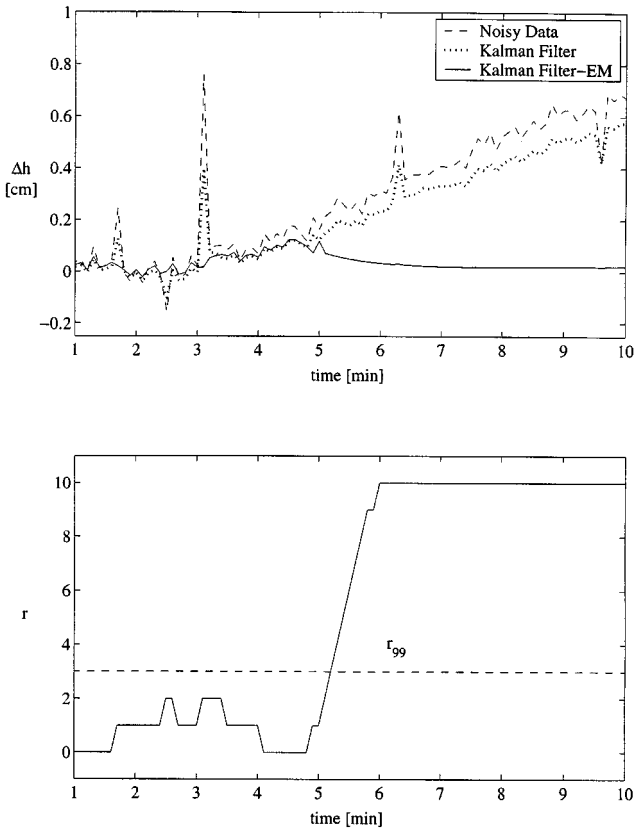

Figure 6. Detection of change of ramp disturbance at the input for Example 1.

analysis (RGA) of the linearized model, concentration C_A is controlled by adjusting the flow rate of the coolant, Q_C, while the temperature T is controlled by the inlet flow rate Q_F. The liquid level is controlled by changing the outlet flow Q. The dynamic model equations for the reactor system are written as

$$\frac{dC_A}{dt} = -k_0 e^{-E/RT} C_A + \frac{Q_F C_{AF} - C_v \sqrt{h}\, C_A}{Ah} \qquad (57)$$

$$\frac{dT}{dt} = \frac{k_0 e^{-E/RT} C_A (-\Delta H)}{\rho C_p} + \frac{Q_F (T_F - T)}{Ah} + \frac{UA_C (T_C - T)}{\rho C_p Ah} \qquad (58)$$

$$\frac{dT_C}{dt} = \frac{Q_C (T_{CF} - T_C)}{V_C} + \frac{UA_C (T - T_C)}{\rho_C C_{pC} V_C} \qquad (59)$$

$$\frac{dh}{dt} = \frac{Q_F - C_v \sqrt{h}}{A}, \qquad (60)$$

where k_0 is the Arrhenius constant; E is the activation energy for the reaction; R is the universal gas constant; C_v is the valve constant; ΔH is the heat of reaction; U is the overall heat transfer coefficient (assumed constant); and ρ, ρ_C and C_p, C_{pC} are densities and specific heats of reaction and

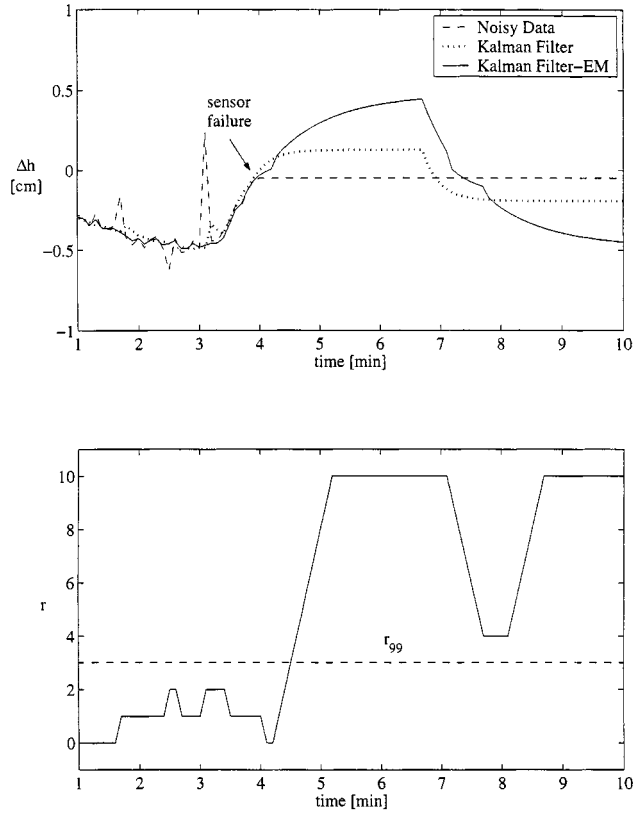

Figure 7. Detection of change for sensor failure for Example 1.

cooling liquid, respectively. The nominal operating values of process variables and constants are given in Table 2. The measurement noise covariance matrix is diag(5×10^{-8}, 5×10^{-1}, 5×10^{-1}, 5×10^{-3}) in appropriate units. This model is used for the extended Kalman filter–EM algorithm simulation.

Outliers Having a Gaussian Distribution. A sinusoidal excitation of amplitude 5 K is applied to the setpoint of T, and a

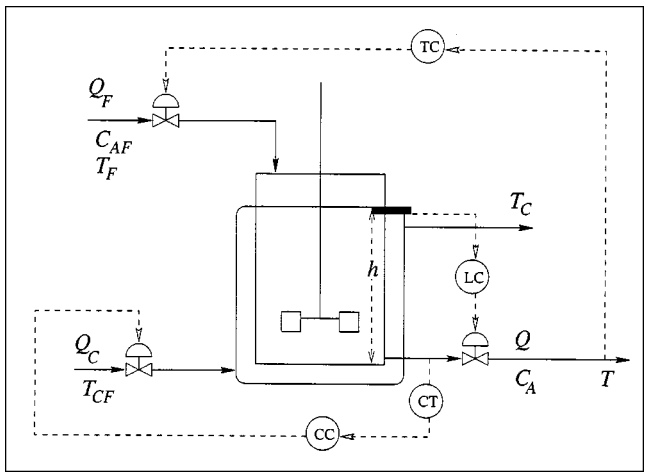

Figure 8. Nonisothermal CSTR.

Table 2. Constants and Nominal Operating Values for Example 2

Variable	Nominal Value	Variable	Nominal Value
$\bar{C}_A$	0.0372 mol/L	T_F	320 K
$\bar{T}$	402.35 K	UA_C	5×10^4 J/min·K
$\bar{T}_c$	345.44 K	E/R	8,750 K
$\bar{h}$	0.6 m	k_0	7.2×10^{10} min^{-1}
$\bar{C}_v$	166.67 L/min·m$^{1/2}$	A	0.167 m^2
$\bar{Q}_C$	15 L/min	ΔH	-5×10^4 J/mol
$\bar{Q}_F$	100 L/min	ρC_p	239.0 J/L·K
$\bar{C}_{AF}$	1.0 mol/L	$\rho_C C_{pC}$	4,175.0 J/L·K
$\bar{T}_{CF}$	300 K	V_C	10.0 L

simulation is performed where outliers are present in measurements that are sampled from a Gaussian distribution. Data rectification is performed using the EKF–EM approach, as described in the previous section, and the resulting error, δ, is presented in Figure 9. The extended Kalman filter alone is unable to rectify outliers in most cases, while the EKF–EM methodology produces satisfactory data rectification.

Outliers Having an Exponential Distribution. The data-rectification error for outliers present in measurements that were sampled from an exponential distribution is presented in Fig-

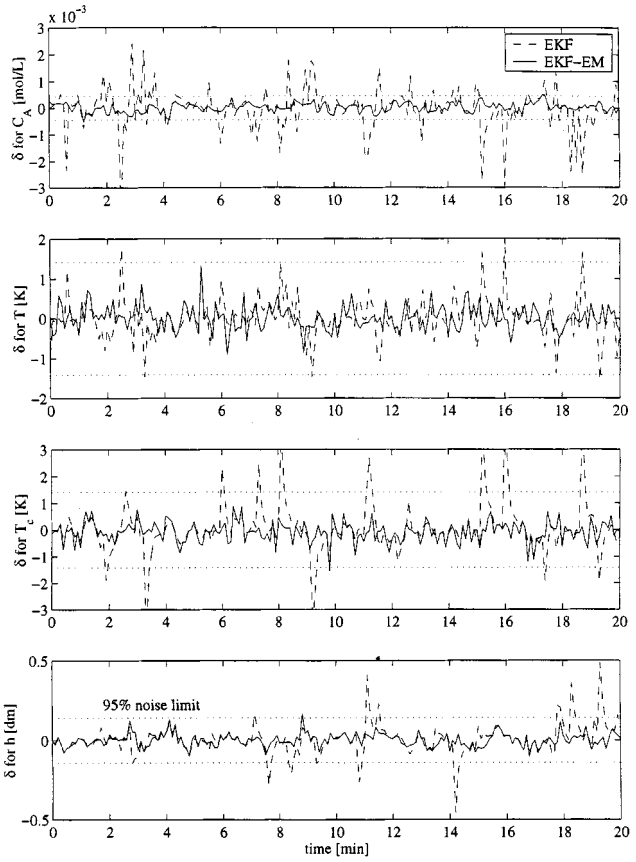

Figure 9. Rectification error for Example 2.

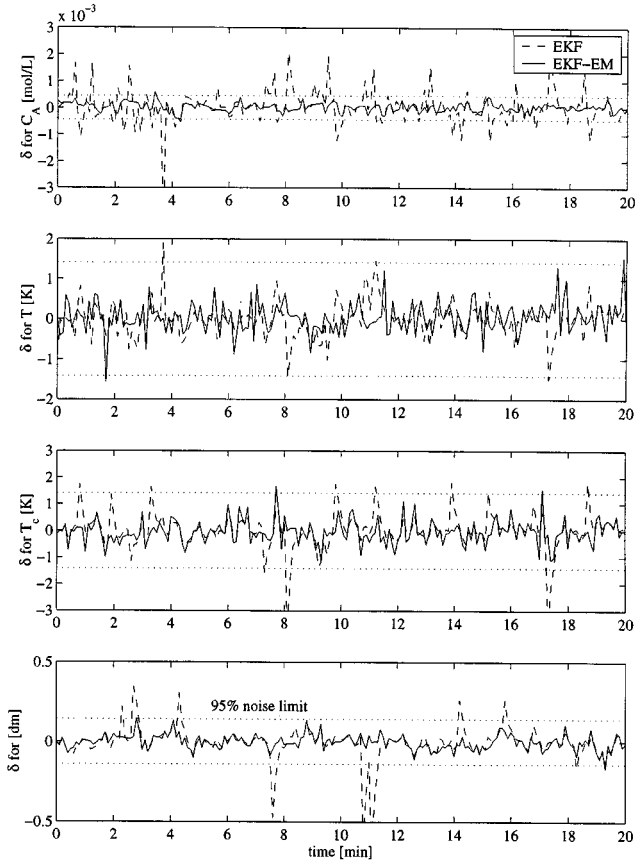

Figure 10. Rectification error with exponentially distributed outliers for Example 2.

ure 10. In this case as well, the EM approach produces satisfactory data rectification.

Detection of Change for Plant-Model Mismatch. In order to evaluate sensitivity to plant-model mismatch, a linear model for the nonlinear CSTR is derived by linearizing the nonlinear model. A sine wave with a large amplitude of 10 K in the setpoint of reactor temperature is introduced so that the responses of the linear and nonlinear models are significantly different. This mismatch can be detected by observing $\bar{\zeta}$ and counting the number of outliers labeled by the EM algorithm, as shown in Figure 11.

Suppose that there are four measurements, and that the probability of observing an outlier in each measurement is ϵ and is independent of occurrence of an outlier in another measurement. Let p be the probability such that $\bar{\zeta} > 0.5$. Then,

$$p = \binom{4}{1}\epsilon - \sum_{k=2}^{4}\binom{4}{k}\epsilon^{k}(1-\epsilon)^{4-k}. \qquad (61)$$

For $\epsilon = 0.1$, Eq. 61 gives $p = 0.34$. This value of p and $n = 10$ give $r_{99} = 6$.

The proposed methodology provides a statistical basis of identifying outliers using model predictions and the actual measurements. Also, the binary indicator variable ζ serves as a parameter that automatically identifies outliers and provides maximum-likelihood estimates of the true values of the

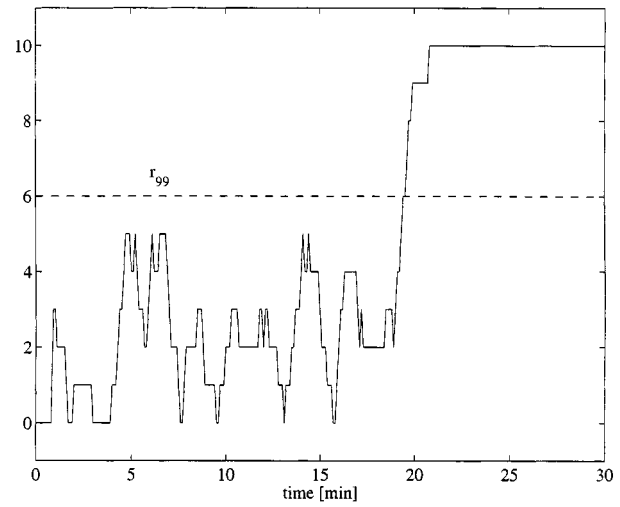

Figure 11. Detection of change from the number of outliers in window of 10 samples for Example 2.

measurements using the Kalman filter. Because this approach requires only model predictions, their covariance, and the measurement data, it is suitable for integration in any state estimation or filtering scheme. It has been demonstrated that it is possible to integrate the EM into the Kalman filter for linear systems as well as the extended Kalman filter for nonlinear systems.

Conclusions

A new dynamic data-rectification methodology based on the Kalman filter and Expectation-Maximization algorithm has been proposed. The new technique is capable of effectively removing outliers from measurement data, and is general enough to be used with any dynamic model of the process for data rectification. In addition to removal of outliers, this method provides information about possible changes (additive or nonadditive) to the process model or the measurement equation using a moving window of samples and observing the number of outliers estimated by the EM algorithm.

Acknowledgments

Financial support from the UCSB Process Control Consortium is gratefully acknowledged. The authors thank Dr. Michael Thompson at Procter & Gamble, Cincinnati, OH, for stimulating discussions and insight into the workings of the EM algorithm.

Literature Cited

Abraham, B., and A. Chuang, "Expectation-Maximization Algorithms and the Estimation of Time Series Models in the Presence of Outliers," *J. Time Ser. Anal.*, **14**, 221 (1993).

Dempster, A. P., N. M. Laird, and D. B. Rubin, "Maximum Likelihood from Incomplete Data Via the EM Algorithm," *J. Roy. Stat. Soc., Ser. B*, **39**, 1 (1977).

Gelb, A., *Applied Optimal Estimation*, MIT Press, Cambridge, MA (1974).

Johnston, L. P. M., and M. A. Kramer, "Maximum Likelihood Data Rectification: Steady State Systems," *AIChE J.*, **41**, 2415 (1995).

Johnston, L. P. M., and M. A. Kramer, ''Estimating State Probability Functions from Noisy and Corrupted Data,'' *AIChE J.*, **44**, 591 (1998).

Juricek, B. C., D. E. Seborg, and W. E. Larimore, ''Early Detection of Alarm Situations Using Model Predictions,'' *Proc. IFAC Workshop on On-Line Fault Detection and Supervision in the Chemical Process Industries*, Solaize, France (1998).

Kalman, R. E., ''New Methods in Weiner Filtering,'' *Proc. Symp. on Engineering Applications of Random Function Theory and Probability*, Wiley, New York (1963).

Mardia, K. V., J. T. Kent, and J. M. Bibby, *Multivariate Analysis*, Academic Press, London (1979).

Moon, T. K., ''The Expectation-Maximization Algorithm,'' *IEEE Signal Process. Mag.*, **11**, 47 (1996).

Schick, I. C., and S. K. Mitter, ''Robust Recursive Estimation in the Presence of Heavy-Tailed Observation Noise,'' *Ann. Stat.*, **22**, 1045 (1994).

Thompson, M. L., *Combining Prior Knowledge and Nonparametric Models of Chemical Processes*, PhD Thesis, MIT, Cambridge, MA (1996).

Tong, H., and C. M. Crowe, ''Detection of Gross Errors in Data Reconciliation by Principal Component Analysis,'' *AIChE J.*, **41**, 1712 (1995).

Tong, H., and C. M. Crowe, ''Detecting Persistent Gross Errors by Sequential Analysis of Principal Components,'' *AIChE J.*, **43**, 1242 (1997).

Xu, L., and M. I. Jordan, ''On the Convergence Properties of the EM Algorithm for Gaussian Mixtures,'' *Neural Comput.*, **8**, 129 (1996).

Manuscript received Nov. 24, 1999, and revision received Mar. 13, 2000.