Detecting Abnormal Process Trends by Wavelet-Domain Hidden Markov Models

Wei Sun, Ahmet Palazoğlu
Dept. of Chemical Engineering and Materials Science, University of California, Davis, Davis, CA 95616

Jose A. Romagnoli
Dept. of Chemical Engineering, The University of Sydney, NSW 2006, Australia

A novel method for detection of abnormal conditions during plant operation uses wavelet-domain hidden Markov models (HMMs) as a powerful tool for statistical modeling of wavelet coefficients. By capturing the interdependence of wavelet coefficients of a measured process variable, a classification strategy is developed that can detect abnormal conditions and classify the process behavior on-line. The method is extended to include multiple measured variables in detection and classification. Two case studies illustrate the potential of this method.

Introduction

With the widespread availability of distributed control systems (DCS), continuous monitoring of chemical process operations is greatly facilitated. Plant operators are asked to manage the operation in such a way as to ensure optimal production levels, while attending occasional alarm situations that may result from equipment malfunctions. Timely identification of such abnormal situations may prove to be critical when there is a potential for a safety hazard that may affect not only the plant and its personnel, but also the surrounding communities. Yet, most operators rely on personal expertise for such a task, and, in some cases, the events may exceed the capabilities of any human operator, thus leaving the plant vulnerable to costly shutdowns and, in the worst case, to possibly fatal accidents (Nimmo, 1995). Hence, it is imperative that human expertise is complemented by computerized support systems that are certainly feasible with today's technological advances. Such support systems consist of various data analysis and interpretation strategies that can provide guidance to the plant personnel for handling abnormal situations. A key component of such a system is process *trend analysis* that monitors important process variables to detect underlying changes that may herald hazardous situations.

Trend analysis critically depends on how a process signal is represented initially to facilitate subsequent classification task(s). This representation should capture, in a compact manner, not only the frequency content of the signal, but also its evolution along the time scale to accurately pin down transitions between distinct process behavior. A relatively recent breakthrough in the signal processing literature, namely the wavelet transform (Daubechies, 1992; Mallat, 1989), enabled the expression of a signal through a set of coefficients that carry localized time-frequency information particular to the signal in question. This precipitated a number of studies in process applications (Motard and Joseph, 1994) and provided a significant impetus in process trending where a compact, information-rich signal representation is a prerequisite (Bakshi and Stephanopoulos, 1994). Such a framework also enjoyed success in fault detection and trend analysis studies (Daiguji et al., 1997; Vedam and Venkatasubramanian, 1997; Bakshi, 1998).

In all process-related applications of wavelets to date, no one has considered the statistical relationships among the wavelet coefficients, and treated them as, essentially, independent and uncorrelated variables. One exception is the work by Bakshi (1998) in which he considers the correlation among wavelet coefficients at each scale. It has been shown recently by Crouse et al. (1998) that such relationships (at and across scales) can be exploited very effectively in signal estimation, thus paving the way for the development of trending and fault detection strategies for process systems. In this article, we propose a new trend analysis method that takes advantage of the wavelet-domain hidden Markov trees (HMTs) for constructing statistical models of wavelets. In that respect, we also consolidate our earlier studies on hidden

Correspondence concerning this article should be addressed to A. Palazoğlu.

Markov models for process trend classification within a rigorous framework (Wong et al., 1998, 2001).

In the sequel, we will first briefly review the fundamentals of the wavelet transform and hidden Markov models to set the stage for the wavelet-domain hidden Markov models (HMMs). We will then discuss our trend detection and classification strategy. Case studies will highlight the salient features of the method and allow us to articulate future directions in this area.

Wavelet Transform

The merits of the wavelet transform can be best understood by contrasting it with the well-known Fourier transform. The Fourier transform of a signal yields a set of Fourier coefficients that quantify the power of the signal at the underlying frequency components of the signal. On the other hand, the wavelet transform of the signal produces a set of wavelet coefficients that not only quantify the power of the signal at the salient frequency resolutions, but also capture the time intervals associated with them. Such a time-frequency resolution is crucial in detecting trends and discontinuities in process signals.

A process signal $f(t)$ can be represented through the discrete wavelet transformation (DWT) as follows (Daubechies, 1992; Crouse et al., 1998)

$$f(t) = \sum_K c_K \phi_{J_0,K}(t) + \sum_{J=-\infty}^{J_0} \sum_K d_{J,K} \Psi_{J,K}(t) \qquad (1)$$

with $c_K \equiv \int f(t)\phi^*_{J_0,K}(t)dt$, and $d_{J,K} \equiv \int f(t)\Psi^*_{J,K}(t)dt$. Here, we define the wavelet function $\Psi(t)$ and the scaling function $\phi(t)$ as

$$\Psi_{J,K}(t) \equiv 2^{-J/2}\Psi(2^{-J}t - K)$$
$$\phi_{J_0,K}(t) \equiv 2^{-J/2}\phi(2^{-J_0}t - K), \quad J,K \in Z \qquad (2)$$

In this representation, integer J indexes the *scale* or resolution of analysis, that is, smaller J corresponds to a higher resolution, and J_0 indicates the coarsest scale or the lowest resolution. K indicates the spatial (time) location of the analysis. For a wavelet $\Psi(t)$ centered at time zero and frequency f_0, *the wavelet coefficient* $d_{J,K}$ measures the signal content around time $2^J K$ and frequency $2^{-J}f_0$. The *scaling coefficient* c_K measures the local mean around time $2^{J_0}K$.

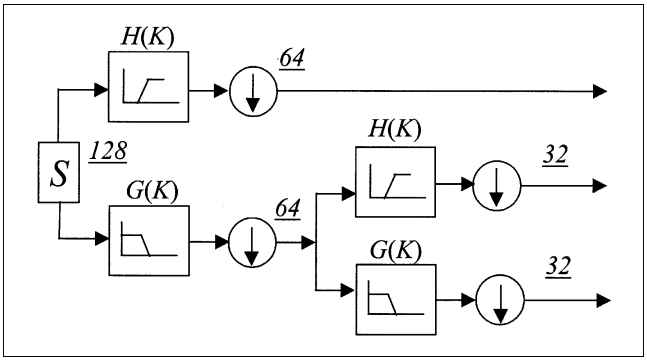

Figure 1. Wavelet analysis by quadrature mirror filters.

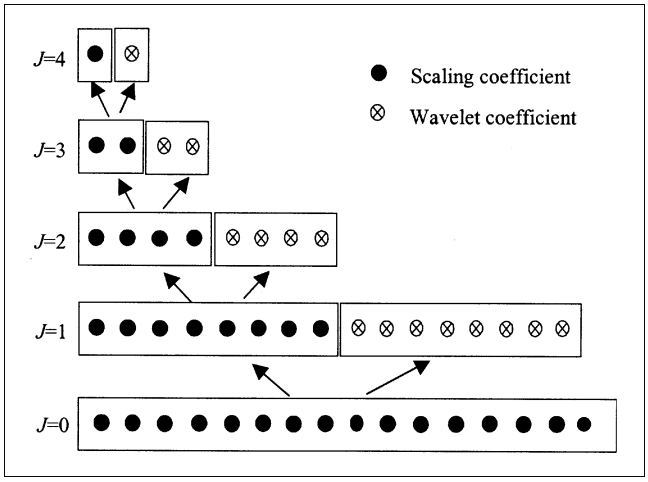

Figure 2. Wavelet decomposition through MSD.

Mallat (1989) proposed an algorithm, referred to as Multiresolution Signal Decomposition (MSD), to efficiently perform discrete wavelet transformation (DWT). Its basic idea is to use a low-pass filter and a high pass filter to decompose a dyadic-length discrete signal (time-series) into low frequency and high frequency components, respectively. As shown in Figure 1, for a signal S consisting of 128 points, one also performs a down-sampling operation to reduce the number of points in each scale by half.

One can show that the relationship between high-pass and low-pass FIR filters and the corresponding wavelet and scaling functions can be expressed as

$$\phi(t) = \sqrt{2} \sum_K H[K]\phi(2t - K)$$
$$\Psi(t) = \sqrt{2} \sum_K G[K]\phi(2t - K) \qquad (3)$$

This approach greatly facilitates the calculation of wavelet and scaling coefficients as typically implemented in the Wavelet Toolbox (MATLAB, 1997). One can associate the scaling coefficients with the signal approximation, and the wavelet coefficients as the signal detail. This decomposition process can be described in Figure 2, for a signal with 16 (2^4) points.

In DWT, the scaling coefficients are decomposed iteratively at each scale (Figure 2) and the dependency between adjacent scales is obvious. Furthermore, for orthogonal wavelet decomposition, wavelet coefficients should be uncorrelated between scales. However, for most signals in practical applications, there is a residual dependency after decomposition, even though the dependency of the wavelet coefficients may be local. In other words, for scaling and wavelet coefficients, there exists a dependency within and across the scales. This is consistent with the clustering and persistence properties of the wavelet coefficients (Crouse et al., 1998) that state that, for a large (small) wavelet coefficient, the adjacent coef-

ficients are also likely to be large (small) and that such values propagate across scales.

Having established the dependency of the wavelet coefficients, one still has to find the appropriate framework for modeling their probability density functions. The simplest choice, a Gaussian model, is not appropriate since the wavelet decomposition tends to produce large number of small coefficients and small number of large coefficients. This is the very property that one takes advantage of in data compression and denoising. Alternatively, the marginal probability of each coefficient can be better represented by a mixture density. Instead of assigning a statistical model to wavelet coefficients, Crouse et al. (1998) suggest assigning a set of states to each coefficient and then associating a probability density function with each state, $f(w|S)$. Intuitively, one can choose a two-state model in which the coefficients can belong either to a high-variance state $[f(w|S=1)]$ or to a low-variance state $[f(w|S=2)]$, resulting in a two-state zero-mean Gaussian mixture model.

Remark

To enhance the fidelity of the fit (Sorenson and Alspach, 1971), more complex mixture models (even with nonzero means) can also be used, naturally at the expense of increased computational burden. This framework also allows the use of non-Gaussian densities (Redner and Walker, 1994).

The last element of the statistical model is the definition of the dependencies among the wavelet coefficients. Given the persistence and clustering properties alluded to earlier, it appears logical to assume Markovian dependencies between the adjacent state variables (not the wavelet coefficients). Such a structure gives rise to the hidden Markov trees (HMTs). Note that this statistical model, as suggested by Crouse et al. (1998), can also be used to describe dependencies among the scaling coefficients (albeit with nonzero means). For this latter case, Gaussian mixture models can reasonably explain the salient distributions of the scaling coefficients. In the following section, we will focus on wavelet coefficient modeling for simplicity, but the examples will demonstrate the use of HMT for scaling coefficients.

Hidden Markov Models

HMM is a doubly stochastic model, which not only can capture the deterministic dependencies between wavelet coefficients, but also take into account the random factors involved in the measurements. The underlying backbone is a Markov process, where the future state is independent of the previous state if the current state is known for a commonly used first-order system. This Markov process is unobservable. One can only observe the occurrences, which are assumed to be decided by the underlying state variable. Typically, HMM has a chain structure, which corresponds to the time evolution of the observed variable (Figure 3). S_k is the hidden state at time t, decided by the transition probability matrix A; b_{sk} is the probability of observation O_k corresponding to state S_k. Such models have been successfully applied in speech recognition (Rabiner and Juang, 1993) and fault detection (Smyth, 1994). We have also used the HMMs in trend detection and classification (Wong et al., 1998).

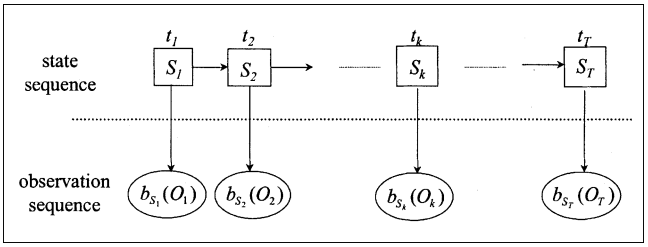

Figure 3. Conventional HMM chain structure.

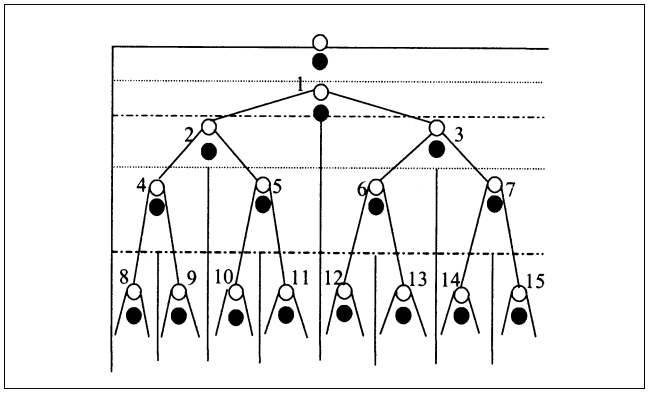

Figure 4. Tree structure HMM in wavelet domain (Crouse et al., 1998).

Crouse et al. (1998) extended the chain structure model into a tree structure model. Here, the underlying backbone becomes a tree structure, and the Markov property exists from the root to the leaves through the branches (Figure 4). This binary tree structure model is the ideal tool for the modeling of coefficients in the wavelet domain.

The HMT model is specified via the parameters, μ_i^m, σ_i^m and *initial*, $\pi_i^m = p(S_i = m)$ and *transition* probabilities, $a_i^{m,n} = p(S_i = m|S_{\rho(i)} = n)$. The subscript $\rho(i)$ refers to the parent node, hence, $S_{\rho(i)}$ is the parent state. Consequently, the HMT model is defined via the following parameters:

(1) $\pi_1^m = p(S_1 = m)$ is the probability mass function (pmf) for the first node.

(2) $a_i^{mn} = p(S_i = m|S_{\rho(i)} = n)$ is the conditional probability that S_i is in state m given its parent $S_{\rho(i)}$ is in state n.

(3) μ_i^m and σ_i^m are the conditional mean and the standard deviation, respectively, of the wavelet coefficient w_i at the ith node, given S_i is in state m, with $f(w_i|S_i = m)$.

The *training* problem determines the set of model parameters given above for an observed set of wavelet coefficients. In other words, we first obtain the wavelet coefficients for the time-series data that we are interested in and, then, the model parameters that best explain the observed data are found by using the maximum likelihood principle. Crouse et al. (1998) proposed an *expectation maximization* (EM) approach that jointly estimates the model parameters and the hidden state probabilities. This is essentially an upward and downward EM method, which is extended from the Baum-Welch method developed for the chain structure HMM. Details of the algorithm can be found elsewhere (Crouse et al., 1998; Sun, 2001).

Remark

For a limited amount of training data, to avoid overfitting, one can achieve a robust training result by assuming an identical distribution for a certain number of nodes, referred to as *tying*. Tying can be applied within and across the scales and increases the number of training data for a certain distribution in the model by simplifying the model structure. Simply put, tying indicates that a certain number of nodes share the same statistical distribution, the same number of states, and the same distribution parameters. In signal denoising, one is interested in the shrinkage of noisy components, and the noise components are assumed to be identically distributed, therefore, tying can significantly help capture such statistical features. In trend analysis, however, the signal trend plays a more important role in characterizing the process failure, and, thus, tying may distort the trend characteristics and will not be employed here.

Trend Analysis via Wavelet-Domain HMTs

In the application of wavelet-domain HMMs to the trend analysis problem, the key is to develop HMMs for all relevant process operating conditions. For example, process data collected during normal operation will help us construct the reference HMM, and data obtained under faulty/abnormal conditions will give rise to other sets of HMMs labeled accordingly. Thus, the trend analysis problem can be stated as follows: *given a set of known models of the process operating conditions, determine the likelihood of a new set of observations*.

Figure 5 depicts the strategy that will be followed in this application. In the training phase, time-series data collected under various conditions are used to develop models using the EM algorithm mentioned before. The monitoring phase, then, considers the on-line signal and determines the model that best explains it, thus, classifying the event associated with the model.

The process data are represented in the wavelet domain in the form of scaling and wavelet coefficients. Ideally, modeling all these coefficients can glean all the information regarding the observed process operating condition, but will result in a large model (tree structure), and increase the computational effort in the training phase. It must be noted that, for different operating conditions, the scaling coefficients (approximation) and the wavelet coefficients (detail) play different roles. Thus, for a specific trend analysis application, a different set of coefficients may be chosen, leading to a trade-off between classification accuracy and computational cost. Undoubtedly, such a decision can be made *a priori* based on the nature of the fault.

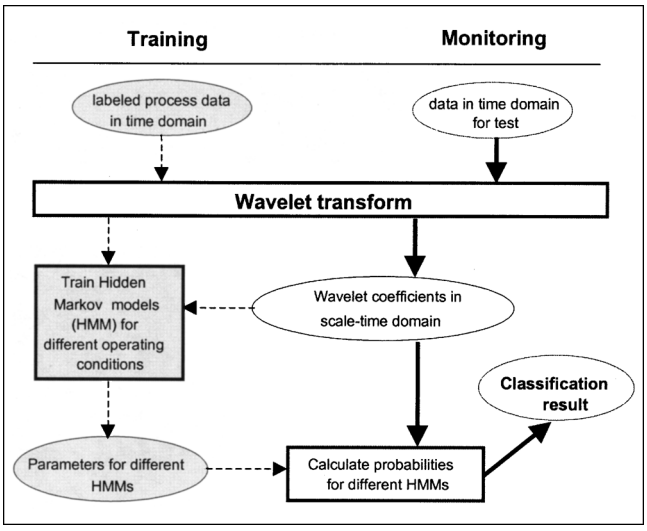

Figure 5. Trend analysis strategy via wavelet domain HMTs.

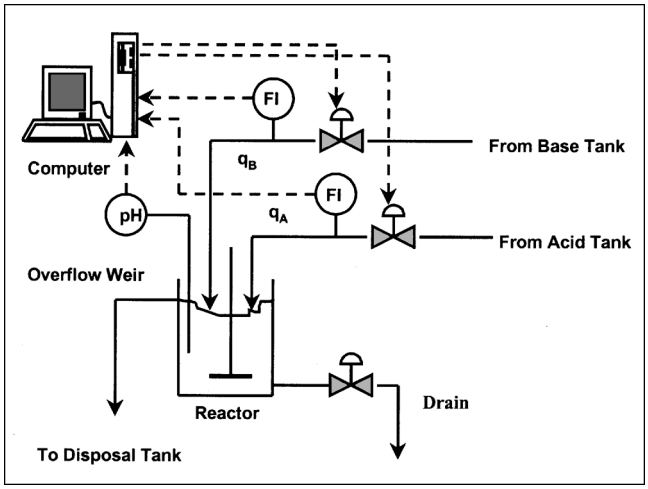

Figure 6. pH process.

Case Study I

We will first study the simulation of a pH neutralization process (Figure 6). The process has been previously studied by Galán et al. (2000) and the reader is referred to that article for details of the model. An acid stream (HCl solution) and an alkaline stream (NaOH and $NaHCO_3$ solution) are fed to a 2.5-L constant volume, well-mixed tank, where the pH is measured through a sensor located directly in the tank. The pH value is maintained at 7.0 by a PI controller ($K_c = 0.8$, $\tau_I = 55$ s).

When an unexpected event occurs, the controller may not maintain the pH value within the allowed range of operation and its performance degrades, thus resulting in an abnormal (faulty) operating condition. In this case study, we shall consider four distinct situations other than the normal operating condition as follows:

- *Abnormal Condition I* (**AI**). The pH value shows a sustained deviation of more than ± 0.5 (region 2 in Figure 7), which could result from a large and sudden change in either the flow or the pH value of the acid stream as a result of changes in the upstream process.
- *Intermediate* (**I**). The pH value indicates deviations (region 3 in Figure 7), which could result from the same source as Abnormal I, but the deviation remains within ± 0.5. This region may act as a warning (buffer zone) for an imminent change in normal operating conditions.
- *Abnormal II* (**AII**). The pH value exhibits high amplitude, high frequency oscillations (region 4 in Figure 7), which

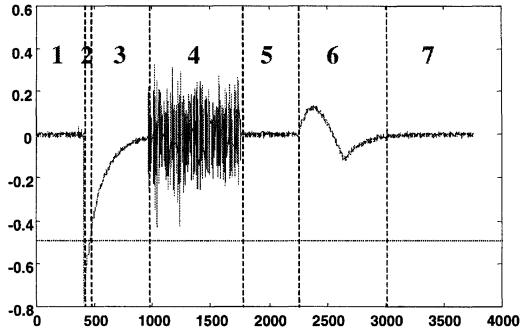

Figure 7. Signal trend of the pH process for testing.

could be the outcome of a sensor failure or other equipment malfunction, such as pump cavitation.

• *Abnormal III* (**AIII**). The pH value increases slowly and reaches a maximum point, then comes back to 7.0 slowly (region 6 in Figure 7), which could be the result of a temporary sensor drift.

Any other operation around pH = 7 is considered as normal (**N**).

A moving window (Wong et al., 1998) is used to analyze the data, since only a limited number of data points can be considered at one time for the wavelet analysis. Typically, if a small window size is chosen, one may capture process changes quickly, but the window may not contain enough information to sufficiently reflect the current process operating condition, thus generating ambiguous classifications. Large window sizes can consider more information, which is helpful to recognize the process trend, but may lead to large time delays for the detection and classification of trends. In this case study, we used an adaptive window size. We shall use a small window size for rapidly changing data and large window sizes for long lasting phenomena, based on the spectral analysis of the signal. The window is moved every sampling time. 16 separate simulation runs are used, 15 for training purposes and the last one (depicted in Figure 7) for testing the methodology. There are 1,200 data points in each simulation set for training and 3,755 data points in the simulation for testing. The data was sampled every 45 s.

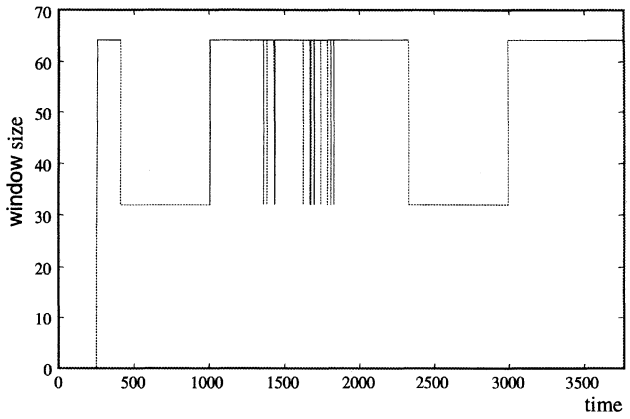

Figure 8. Adaptive selection of window sizes during testing.

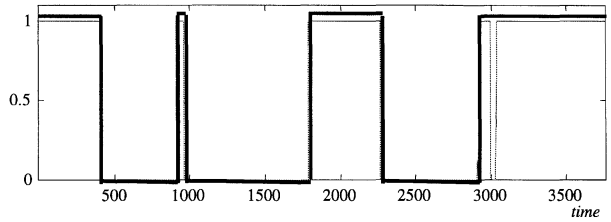

Figure 9a. Classification of trends using HMT for single variable. Solid lines represent the real probabilities and dashed lines the calculated probabilities.

Probabilities of normal (**N**) operating condition.

In this example, we used the Haar wavelet with a single scaling tree, since the Haar wavelet is considered to be effective for edge detection (Crouse et al., 1998). We modeled the scaling coefficients using a two-component ($M = 2$) HMT model with nonzero mixture means. The models were trained using multiple observations (without tying). We did not model the wavelet coefficients since the approximate signal contains more distinguishing features among the studied abnormalities, as the primary goal in this study is to detect if the pH deviation is beyond the tolerable limit ± 0.5. In other words, the decision depends more on the information provided by the scaling coefficients than the wavelet coefficient.

To keep the analysis simple, we considered two choices in this work, the 32 point-length window and 64 point-length window. After training, five sets of models for different operating conditions under 32 and 64 point-length windows are obtained. We used spectral analysis based on Thomson's multitaper method (Thomson, 1982) to differentiate the short- and long-term signal behavior, then chose the appropriate window size to analyze the test signal, since the high frequency components are more important in short-term behavior and the low frequency components dominate in long-term behavior. Window size selection for the test signal is depicted in Figure 8.

By calculating the probabilities of test data for each HMT model, the labels are assigned. Figure 9 depicts the classification result from likelihood determination, comparing the true and calculated probabilities (dark solid lines represent the true probability, while the dashed line represents the calculated probability). The y-axis for each plot is the probability

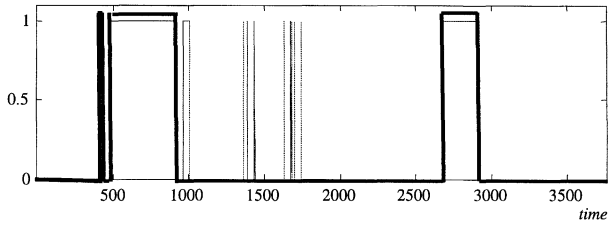

Figure 9b. Classification of trends using HMT for single variable. Solid lines represent the real probabilities and dashed lines the calculated probabilities.

Probabilities of abnormal operating condition **AI**.

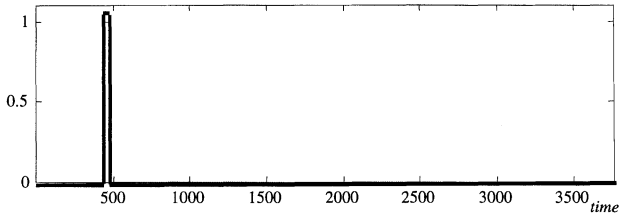

Figure 9c. Classification of trends using HMT for single variable. Solid lines represent the real probabilities and dashed lines the calculated probabilities.

Probabilities of intermediate (**I**) operating condition.

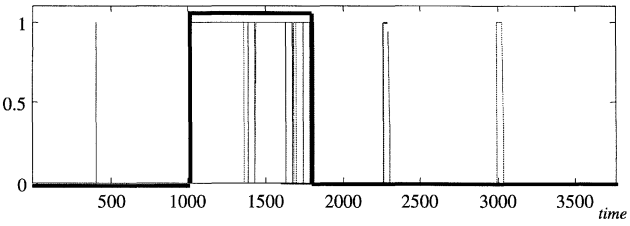

Figure 9d. Classification of trends using HMT for single variable. Solid lines represent the real probabilities and dashed lines the calculated probabilities.

Probabilities of abnormal operating condition **AII**.

(0 to 1) of belonging to that class and the x-axis has time in seconds. We can observe that the HMT model yields the correct classification for most part of the test signal.

- The abnormal condition **AI** (disturbance) and the abnormal condition **AIII** (sensor drift) can be recognized clearly (Figures 9c and 9e).
- The instances of misclassification between sensor noise (**AII**) and intermediate operating condition (**I**) (Figures 9b and 9d) are notable. As the level of noisy signal momentarily matches the level of the signal in the intermediate region, the method results in misclassification. Yet, since these misclassification instances are rather isolated, the overall trend can still be inferred. Nevertheless, the model may need further training to eliminate such instances.
- The brief misclassification between the normal condition (**N**) and the sensor noise (**AII**) (Figure 9a and 9d) is due to the switching of the window size from 32 to 64, and as the number of data points increase suddenly, this causes the method to consider the new signal section as noisy. As the window moves forward, this misclassification error is corrected immediately.

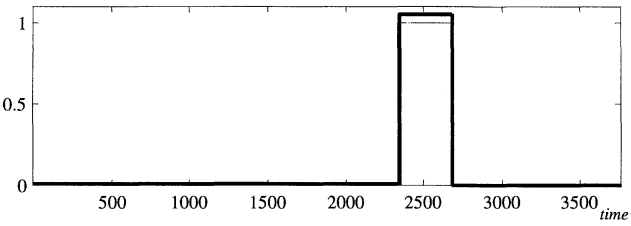

Figure 9e. Classification of trends using HMT for single variable. Solid lines represent the real probabilities and dashed lines the calculated probabilities.

Probabilities of abnormal operating condition **AIII**.

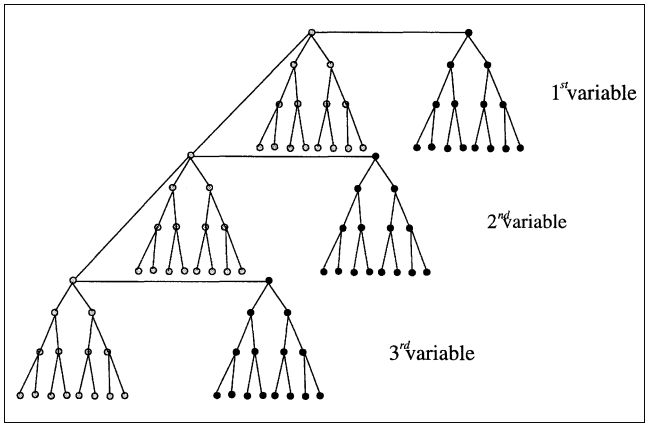

Figure 10. Multitree structure of extended HMT.

HMTs with Multiple Variables

In this section, we introduce an overall classification strategy to extract all possible information in the multiscale context as we did in single variable classification, which can now take into account the correlations among variables and the local dependencies within a single variable, as well as the random factors in all measurements. Different from our previous work (Bakhtazad et al., 2000), however, where we used PCA as the underlying multivariate analysis structure, we extend here the HMT model from the single tree structure to the multitree structure (MHMT), which is then used in the multivariate trend analysis. To illustrate the concept, a three variable multitree structure is depicted in Figure 10.

For each measured variable, we have two tree structures jointed together, one constructed by the scaling coefficients (gray nodes) and the other by the wavelet coefficients (black

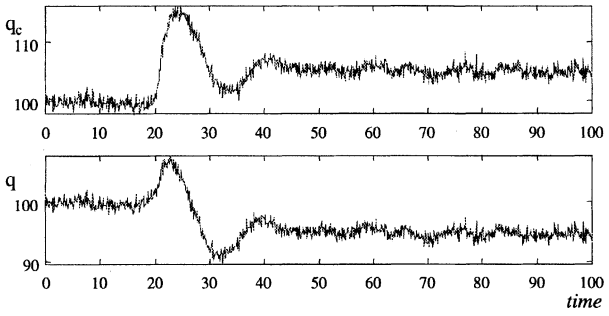

Figure 11a. Four abnormal conditions in CSTR reflected in the process manipulated variables.

Outlet flow rate and coolant flow rate responses to the inlet concentration stepup.

nodes). The root nodes of each tree from a single variable are connected together. The joint structure can be used for any single variable modeling, which includes all frequency (scale) components in this specific variable. To limit the computational complexity, we will only use the tree of scaling coefficients as we did in the previous section. In process monitoring, magnitude of a variable contains more information of the process, which is mainly described by its corresponding scaling coefficients. Also, scaling coefficients at each scale represent a smooth version of the signal with a different resolution; therefore, it is not difficult to argue that the scaling coefficients are sufficient for most of the process monitoring applications.

In our case, the root nodes of each tree structure are connected, which corresponds to each variable under consideration. In each single tree, we already accounted for the deterministic trend information and the random factors involved. The rationale behind using the multivariate tree structure is to be able to capture the correlations among variables. The connection here among variables is arbitrary, and the apparent parent-child connection does not really imply the parent-child dependence, but it is just a way to model the relation between two nodes. In principle, the multitree structure can be expanded indefinitely, but the computational complexity may restrict the number of variables.

The EM algorithm for the single variable case also applies in a straightforward manner to the multitree structure, since the binary structure remains unchanged in this structure expansion. For on-line process monitoring, the moving window is again used in the multivariable case. The complexity of the computation increases by the number of trees compared to the single tree at the same window size, but, as we will show later, the multivariable system contains more structural information, which makes it possible to reduce the window size, and, therefore, keep the computational complexity the same as, or less than, the single variable case. In principal, one can use different window sizes for variables, but the same number of data points needs to be used to each variable to keep the same weight of contribution to the MHMT from each variable. Using different window sizes makes the monitoring delay by the time of the longest window size used in the model, but it may reduce the computational complexity if a longer window size is needed for some of the variables. Longer window size usually is suggested for the slow dynamics, so a less number of data points within a certain time period is enough to characterize the variable trend. Similarly, a smaller window size is for the fast dynamics, so a greater number of data points within a certain time period is needed.

The monitoring procedure is the same as in the single variable case (Figure 5). The model training provides the model parameters for each known operating condition, and the process operating condition of any incoming data is labeled according to the likelihood value of each known operation model.

Case Study II

We chose the simulation of a constant volume, continuous stirred-tank reactor (CSTR) to demonstrate the multivariate trend analysis strategy. In the CSTR, a single irreversible, exothermic reaction A → B is assumed to occur and the model equations are published elsewhere (Pottmann and Seborg, 1992; Sun, 2001).

The system disturbances are the feed concentration and temperature c_{Af} and T_f respectively, and the control outputs are the tank concentration and temperature c_A and T, respectively. The two outputs are controlled by two PI controllers via feed and coolant flow rates q and q_c, respectively. For the concentration controller, the proportional gain is 0.02 and the integral time is 5 min. For the temperature controller, the proportional gain is 0.5 and the integral time is 0.2 min. White Gaussian noise is added to the outputs to simulate the real process signal. The sampling interval is assumed to be 0.1 min. The normal operating condition (**N**) is taken as the steady state (Pottman and Seborg, 1992)

$$c_{As} = c_A(0) = 8.235 \times 10^{-2} \text{ mol/L}$$

$$T_s = T(0) = 441.81 \text{ K}$$

Any deviations from the assumed normal operating condition are considered as abnormal operating conditions, which would need the operator's attention. In this case study, we define four abnormal operating conditions and monitor the

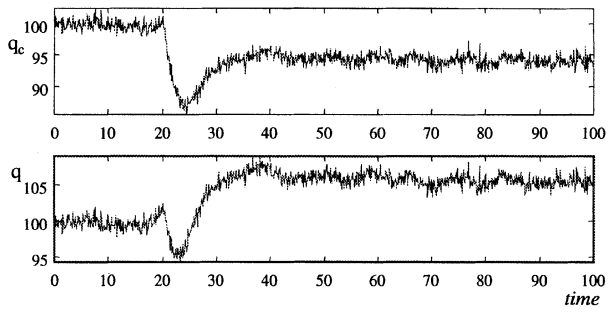

Figure 11b. Four abnormal conditions in CSTR reflected in the process manipulated variables.

Outlet flow rate and coolant flow rate responses to the inlet concentration stepdown.

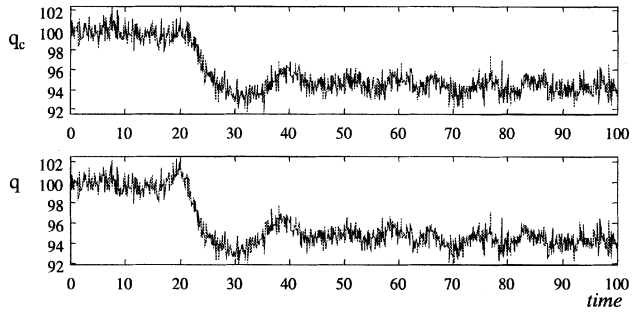

Figure 11c. Four abnormal conditions in CSTR reflected in the process manipulated variables.

Inlet flow rate and coolant flow rate responses to the preexponential factor decreasing.

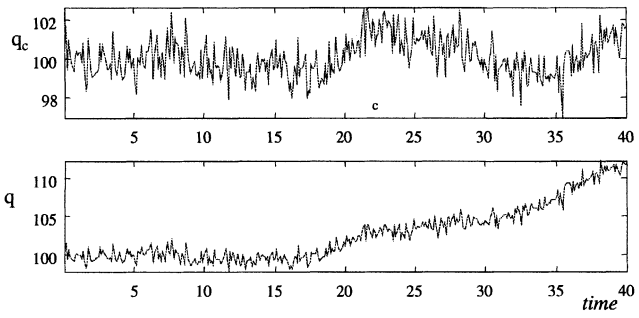

Figure 11d. Four abnormal conditions in CSTR reflected in the process manipulated variables.

Outlet flow rate and coolant flow rate responses to the outlet flow rate sensor drift.

two manipulated variables q and q_c. Sample simulations for the deviations from normal operating condition are depicted in Figure 11.

Tested cases are summarized below.

- In response to a sudden increase in inlet concentration c_{Af} (**AI**), q decreases and q_c increases (Figure 11a).
- In response to a sudden decrease in inlet concentration c_{Af} (**AII**), q increases and q_c decreases (Figure 11b).
- A sudden decrease in the pre-exponential factor k_0 (**AIII**) due to an unmeasured component variation in the inlet flow, both q and q_c decrease (Figure 11c).
- The flow sensor for q drifts high (**AIV**) without affecting the process, so other variables, including q_c, remain unchanged (Figure 11d).

Three simulations of each operating condition are carried out for model training. One simulation under each operating condition is used to test the monitoring result. Each test data includes the transient process from the normal operating condition to the abnormal operating condition. Two window sizes, 8-pt window and 16-pt window, are used and two wavelet functions, Haar and Symlet 8, are considered also for comparison purposes. The test results are shown in Figures 12–15. The y-axis for each plot is the probability (0 to 1) and the x-axis time in minutes. The results are for only 8-pt window and Haar wavelet combination as the others produced similar results indicating the insensitivity of the results to the choice of the window size and the wavelet function (Sun, 2001). To simplify the model structure and, therefore, to reduce the computational effort, only scaling coefficients are considered in this case study. The following remarks are in order:

- The method classifies **AI** correctly (Figure 12) with a small delay. During the delay, **AI** is temporally misclassified as **AIV** due to the similar response of these two abnormalities in the beginning.
- The method classifies **AII** correctly when the process nears steady-state (Figure 13). There is a brief period of mis-

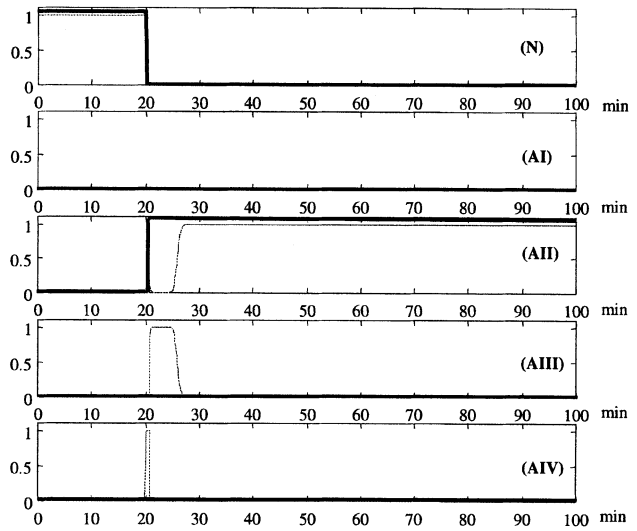

Figure 13. Classification result of AII.

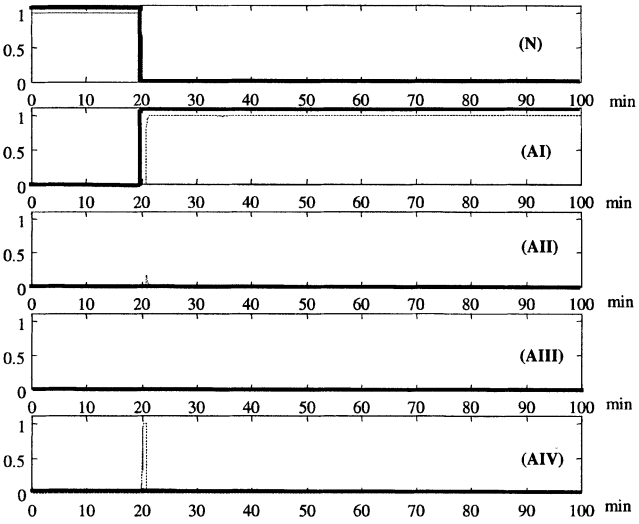

Figure 12. Classification result of AI.

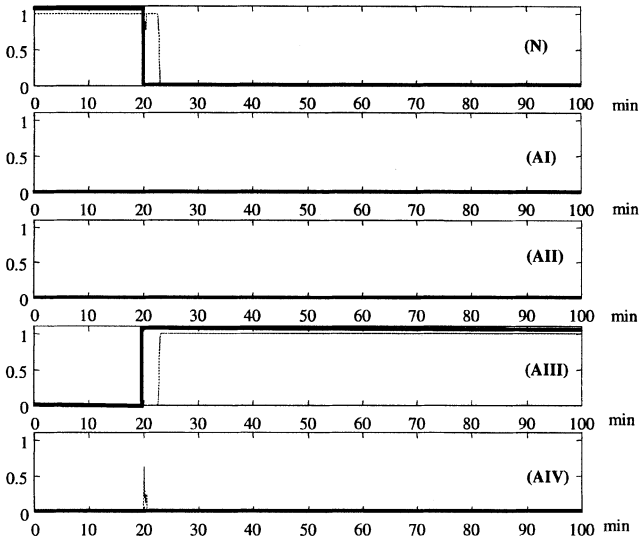

Figure 14. Classification result of AIII.

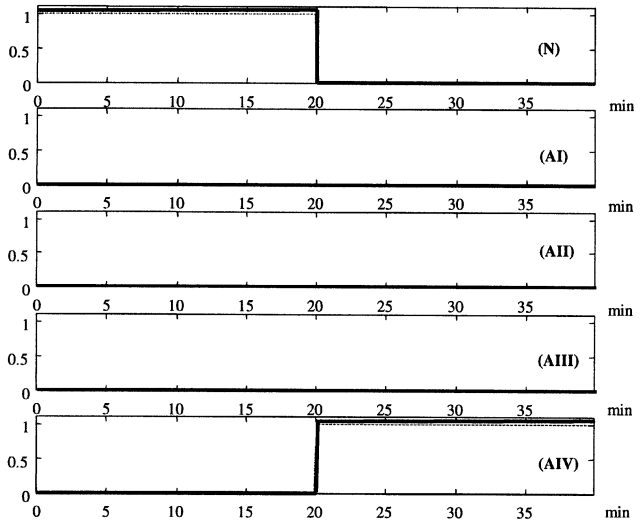

Figure 15. Classification result of AIV.

classification during the transient between **AII** and **AIII** due to the similar responses of the two monitored variables. With more process information, the misclassification can be possibly avoided. As in (a), there is a temporary misclassification between **AII** and **AIV**.

- The method classifies **AIII** correctly when process nears steady state (Figure 14). There is a short period of misclassification between **N** and **AIII** during the transient, which is considered as **N** initially.
- The method classifies **AIV** correctly.

From the results above, we can conclude that the MHMT method for trend analysis correctly classifies the different process operating conditions. There are a few ambiguities in the transient region, which result from the similar responses

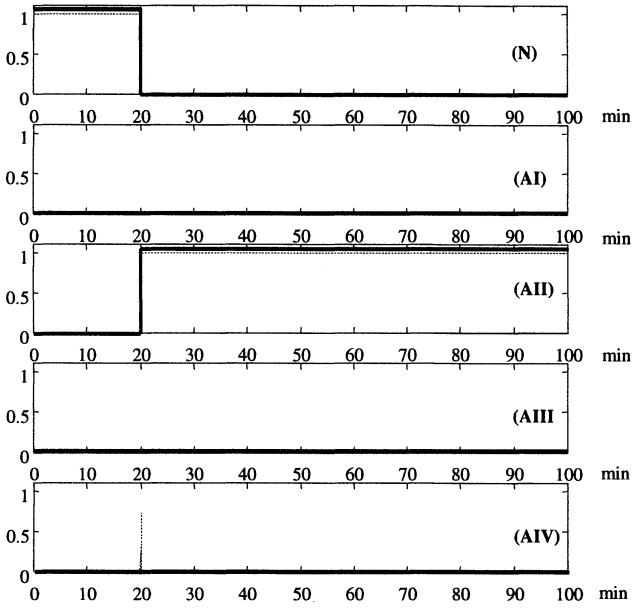

Figure 17. 8-pt Window and three-variable classification result of AII by using Haar wavelet.

of manipulated variables to different process events. In other words, the information contained in two variables is not enough to immediately discern the transient part of the process. To eliminate such ambiguities, additional variables may be needed in the HMT model.

It is also noted that the results do not show significant difference with respect to the different window sizes and wavelet functions in this particular case study. This indicates that a small window size and a simple wavelet function can be as efficient as a long window size and a complicated wavelet

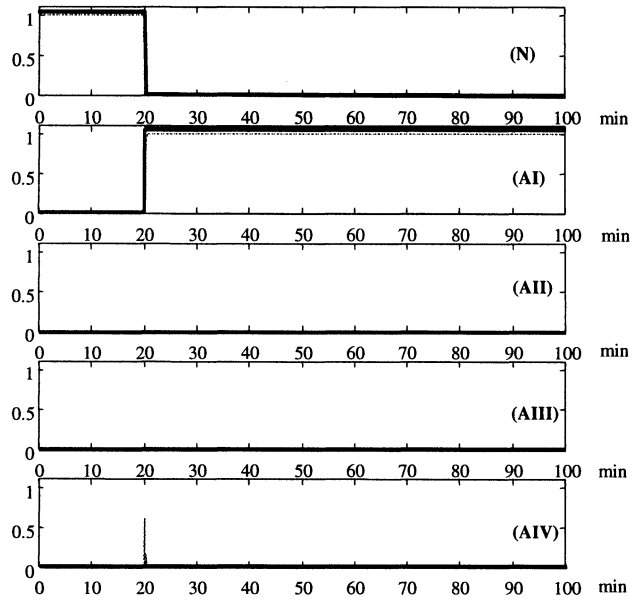

Figure 16. 8-pt Window and three-variable classification result of AI by using Haar wavelet.

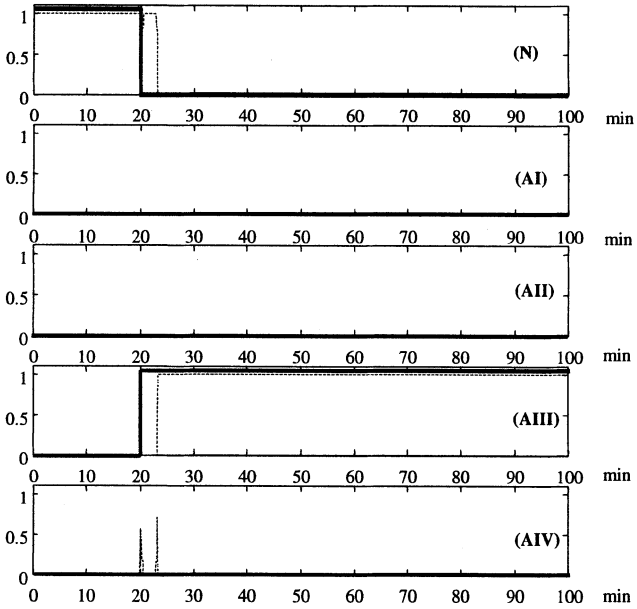

Figure 18. 8-pt Window and three-variable classification result of AIII by using Haar wavelet.

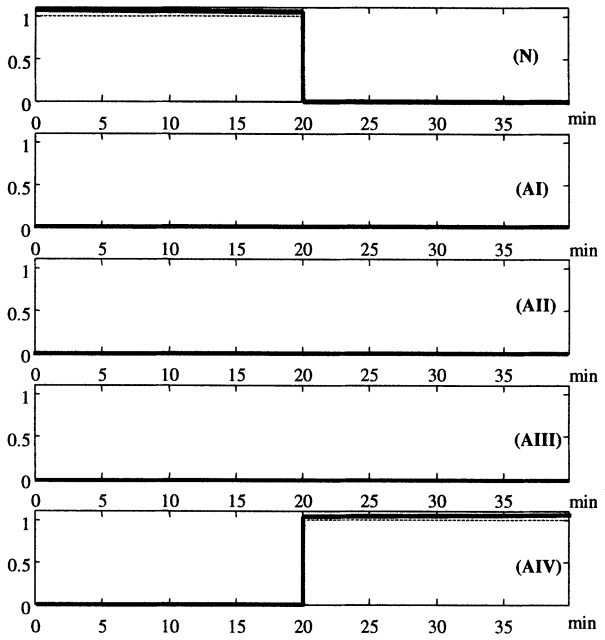

Figure 19. 8-pt Window and three variable classification result of AIV by using Haar wavelet.

function. The use of a small window size can reduce the delay of abnormal operating condition detection and save computer resources. It also makes it possible to consider more process variable and, therefore, extract more information from the process without increasing the computational effort.

To address the misclassification between **AII** and **AIII**, we next included the measurements of the inlet concentration. The result is shown in Figures 16–19. We can see that the model based on three variables renders the classification significantly sharper than the one based on only two variables (Figures 16–17), and that the misclassification between **AII** and **AIII** is successfully corrected (Figure 17). This is, of course, expected as the inlet concentration measurement provides direct information about the disturbance. We can also notice that the results of **AIII** and **AIV** remain identical as before, since there is no new information added to help identify **AIII** (Figure 18) and two measurements are enough to characterize **AIV** (Figure 19).

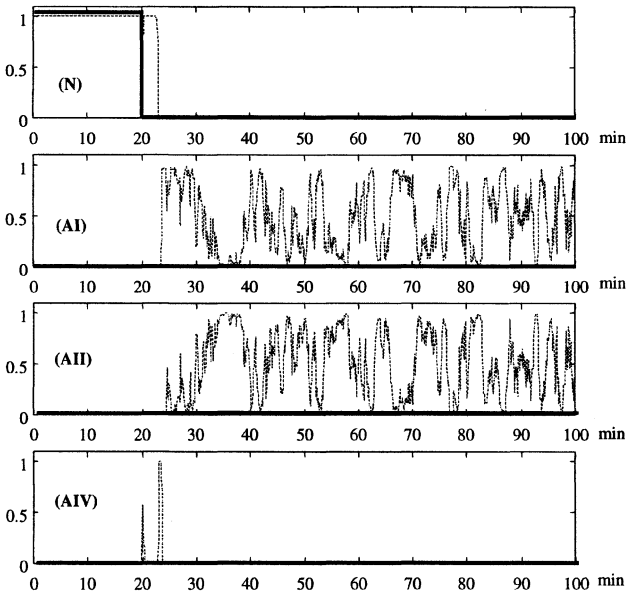

Figure 20. AIII classification result with AIII unknown.

So far, we assumed that all types of process operating conditions are known. For a real process, it is hard to know and model *a priori* all possible abnormalities that might happen. We will now test the performance of the method when it faces an unknown event. Here, we will assume that **AIII** is unknown (unmodeled), and test the MHMT classification performance in the presence of other known models. Figure 20 depicts the classification result that shows that the MHMT can not assign the cause of the unknown **AIII** operating condition. Based on the behavior of measured variables, the unknown operating condition is similar to the **AI** and **AIII**, but not exactly the same. The method returns an ambiguous classification result, which can be used effectively to alert the operators about a new event that needs to be modeled for future reference.

Conclusions

In this article, we presented a wavelet-domain HMT method for process trend analysis to help detect and classify process abnormalities. The classification results on simulated data show that this method can efficiently capture the inherent characteristics of process trends and produce a correct recognition of events. While multiple faults cannot be handled at this point, the methodology offers significant promise as the framework for the class of problems discussed here.

While on-line implementation of the trend detection strategy is indeed feasible, the training of HMT models may be rather time consuming and inefficient. The presence of local minima encountered in the EM algorithm limits the complexity of the problems that can be tackled at the present time. We have found that the amount of data required for training is extremely problem dependent, and, especially, when process events have somewhat similar features, more data would be required for training to ensure models can capture the subtleties associated with each event. A smaller window size helps training computationally and can improve detection time, yet the classification performance may deteriorate as trends may not have developed distinct features in a short time. Moreover, combining detail and approximate coefficients to build HMT models would be a natural next step in process trend detection, but this extension is currently hampered by computational difficulties as mentioned before. Furthermore, in the multivariable problem, we have found that three to five variables can be handled relatively easily, but one reaches a computational bottleneck for larger problems. All these can be addressed and possibly resolved by considering some of the new developments in HMM training algorithms (Romberg et al., 2001; Fan and Xia, 2001). These issues are currently under investigation.

Acknowledgments

A grant from the Australian Research Council is gratefully acknowledged. The authors also thank the anonymous reviewers for their valuable comments.

Literature Cited

Bakhtazad, A., A. Palazoglu, and J. A. Romagnoli, "Detection and Classification of Abnormal Process Situations Using Multidimensional Wavelet Domain Hidden Markov Trees," *Comput. & Chem. Eng.*, **24**, 769 (2000).

Bakshi, B. R., and G. Stephanopoulos, "Representation of Process Trends, Part III. Multi-Scale Extraction of Trends from Process Data," *Comput. and Chem. Eng.*, **18**, 267 (1994).

Bakshi, B. R., "Multiscale PCA with Application to Multivariate Statistical Process Monitoring," *AIChE J.*, **44**, 1596 (1998).

Crouse M. S., R. D. Nowak, and R. G. Baraniuk, "Wavelet-Based Statistical Signal Processing Using Hidden Markov Models," *IEEE Trans. on Signal Processing*, **46**, 886 (1998).

Daiguji, M., O. Kudo, and T. Wada, "Application of Wavelet Analysis to Fault Detection in Oil Refinery," *Comput. & Chem. Eng.*, **21**, s1117 (1997).

Daubechies, I., *Ten Lectures on Wavelets*, Society for Industrial and Applied Mathematics, Philadelphia, PA (1992).

Fan, G., and X-G Xia, "Improved Hidden Markov Models in the Wavelet-Domain," *IEEE Trans. on Signal Processing*, **49**, 115 (2001).

Galán, O., A. Palazoglu, and J. A. Romagnoli, "Robust H_∞ Control of Nonlinear Plants Based on Multi-linear Models—An Application to a Bench Scale pH Neutralization Reactor," *Chem. Eng. Sci.*, **55**, 4435 (2000).

Mallat, S. G., "A Theory for Multiresolution Signal Decomposition: The Wavelet Representation," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, **11**, 674 (1989).

Motard, R., and B. Joseph, *Wavelet Applications in Chemical Engineering*, Kluwer Academic, Boston, MA (1994).

Nimmo, I., "Adequately Addressing Abnormal Operations," *Chem. Eng. Prog.*, **91**, 36 (1995).

Pottmann, M., and D. E. Seborg, "Identification of Non-linear Processes Using Reciprocal Multiquadric Functions," *J. of Process Control*, **2**, 189 (1992).

Rabiner, L., and B. Juang, *Fundamentals of Speech Recognition*, Prentice Hall, Englewood Cliffs, NJ (1993).

Romberg, J. K., H. Choi, and R. G. Baraniuk, "Bayesian Tree Structured Image Modeling Using Wavelet-Domain Hidden Markov Models," *IEEE Trans. on Image Processing*, **10**, 1056 (2001).

Redner, R., and H. Walker, "Mixture Densities, Maximum Likelihood and the EM Algorithm," *SIAM Rev.*, **26**, 195 (1994).

Smyth, P., "Hidden Markov Models for Fault Detection in Dynamic Systems," *Pattern Recognition*, **27**, 149 (1994).

Sorenson, H. W., and D. L. Alspach, "Recursive Bayesian Estimation Using Gaussian Sums," *Automatica*, **7**, 465 (1971).

Sun, W., "Process Trend Analysis via Wavelet Domain Hidden Markov Models," PhD Diss., University of California, Davis (2001).

Thomson, D. J., "Spectrum Estimation and Harmonic Analysis," *Proc. of IEEE*, **70**, 1055 (1982).

Vedam, H., and V. Venkatasubramanian, "A Wavelet Theory—Based Adaptive Trend Analysis System for Process Monitoring and Diagnosis," *Proc. of American Control Conf.*, American Automatic Control Council, Evanston, IL, distributed through IEEE Service Center, Piscataway, NJ (1997).

MATLAB, *Wavelet Toolbox for Use with MATLAB*, User's Guide, Mathworks Inc., Natick, MA (1997).

Wong, J. C., K. A. McDonald, and A. Palazoglu, "Classification of Process Trends Based on Fuzzified Symbolic Representation and Hidden Markov Models," *J. of Process Control*, **8**, 395 (1998).

Wong, J. C., K. A. McDonald, and A. Palazoglu, "Classification of Abnormal Plant Operation Using Multiple Process Variable Trends," *J. of Process Control*, **10**, 409 (2001).

Manuscript received Oct. 1, 2001, and revision received Aug. 27, 2002.