

**ART-2 AND MULTISCALE ART-2 FOR ON-LINE
PROCESS FAULT DETECTION - VALIDATION VIA
INDUSTRIAL CASE STUDIES AND MONTE CARLO
SIMULATION****H. B. Aradhye^{*1} J. F. Davis^{**} B. R. Bakshi^{***,2}**

** Machine Vision and Robotics Group, SRI International, Menlo
Park, CA*

*** Dept of Chem Eng, University of California, Los Angeles, CA*

**** Dept of Chem Eng, The Ohio State University, Columbus,
OH*

Abstract:

Data from most industrial processes contain contributions at multiple scales in time and frequency. In contrast, most existing methods for fault detection are best for detecting events at only one scale. This paper provides experimental validation and insight into a new method of process fault detection based on the integration of multiscale signal representation and scale-specific clustering-based diagnosis. The multiscale ART-2 (MSART-2) algorithm models normal process operation as clusters of wavelet coefficients at different scales. It detects a process change when one or more wavelet coefficients of test data violate similarity thresholds with respect to clusters of normal data at that scale. Especially in industrial situations where the nature of the abnormal features is not known a priori, MSART provides better average performance due to its ability to adapt to the scale of the features. In contrast to most other multiresolution schemes, this framework exploits clustering behavior of wavelet coefficients of multiple variables for the purpose of scale selection and feature extraction. Detailed performance comparisons, based on rigorous Monte-Carlo simulations as well as industrial data from a large scale petrochemical process, are provided. Our results show that MSART-2 significantly improves the detection performance of the ART-2 detection algorithm over a broad range of process anomalies. Results are compared with single-scale and multiscale versions of PCA for benchmarking purposes.

Keywords: Wavelets, Adaptive Resonance Theory (ART), Principal Component Analysis (PCA), Process Monitoring and Fault Diagnosis

1. INTRODUCTION

In an environment where most process maneuvers are automated, algorithms to detect and classify

abnormal trends in process measurements are of critical importance from the point of view of safe and economical plant operation. These algorithms use information extracted from previously annotated process data for predicting, preferably in real time, the state of the process when only unannotated measurements are available. This task is referred to as *fault diagnosis* or *anomaly detection*

¹ H. B. Aradhye was a graduate student at the Department of Chemical Engineering and the Department of Electrical Engineering, The Ohio State University, when this research was being done.

² Author to whom all correspondence should be directed.

and isolation in the statistical process monitoring community. Clearly, one can draw close parallels to the above objective from fields as diverse as e-commerce (fraud detection), network security (intrusion detection), and wireless communication (signal detection). It is not surprising, then, that algorithms designed for each of these varied applications often rely on the same repository of pattern recognition/statistical modeling methods, such as neural networks and PCA, for learning the characteristics of the data. This work focuses on one such method, namely Adaptive Resonance Theory (ART), and reports significant performance gains in terms of faster, noise-tolerant detection under the proposed multiscale framework. The current work, however, is not specific to ART and has the potential to benefit other parallel applications across different domains and modeling methods listed above.

Common approaches to fault detection can be classified as causal analysis-based or clustering-based. Causal analysis-based approaches operate on discretized or fuzzified measurements and residuals. Examples of methods belonging to this category are the signed directed graph approach, fault tree approach, belief networks, etc. These methods model the fault-symptom relationship as a directed causal link (Wilcox and Himmelblau, 1994; Park and Seong, 1994; Qian, 1990; Yu and Lee, 1991; McDowell *et al.*, 1991). The links can be used to model qualitative and semi-qualitative relationships as well as logical and probabilistic relations. They are primarily used for fault identification. For example, a signed directed graph (SDG) represents the system state variables and malfunctions as nodes that are joined by arcs that represent the causal relationships. Fuzzy logic is often used in these architectures to represent the uncertainty and the graded presence or absence of faults. Bayesian Belief Networks are also one of the successful approaches (Rojas-Guzman and Kramer, 1994).

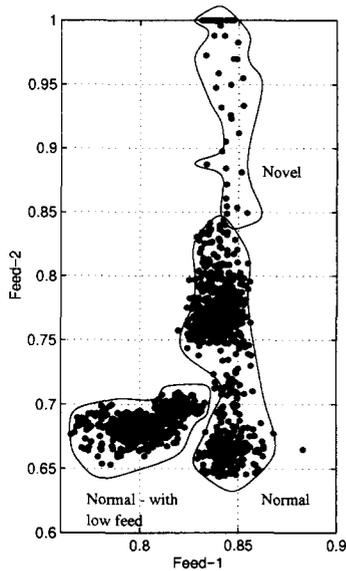
In contrast, clustering-based approaches are based on the observation that most real-world large-scale industrial processes, by their inherent nature, are not precisely defined in the space of sensor measurements. Within a loosely defined region, any given process may follow any of the several possible paths depending on a large number of known or unknown factors. There may exist several such regions, possibly disjoint, because of factors such as various combinations of input feed characteristics, changes in the desired nature of output, variations in the environmental conditions, and so on as shown in Figure 1a. Clustering-based models approximate these complex, multivariate modes of operation as *regions* in sensor space as opposed to deriving a precise functional relationship and are, thus, well suited

for diagnosis of industrial processes (Whiteley *et al.*, 1996; Kavuri and Venkatasubramanian, 1993) and form the basis of this work.

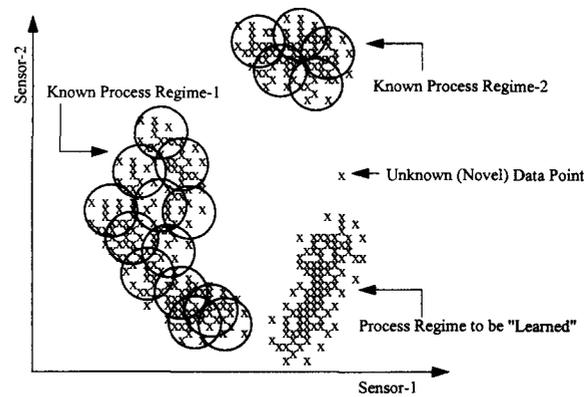
1.1 ART-based Fault Detection

The ART family of networks (Carpenter and Grossberg, 1987; Carpenter *et al.*, 1991c; Carpenter *et al.*, 1991a; Carpenter *et al.*, 1991b; Carpenter *et al.*, 1992) includes some of the few clustering algorithms that explicitly address the issue of stable adaptation and incremental learning with changing process behavior. Typical real world processes often drift from one operating regime to the other, exploring previously unknown equilibria in response to the ever-changing environment. When new information is available in terms of the latest process data, an ART-based fault detector can choose to modify its current clusters or add new clusters. This incremental modification takes place in a way which ensures that the network remains stable as well as capable of adaptation to the changing process conditions. ART and ARTMAP-based networks have been investigated for process modeling and diagnosis of multivariate chemical data by several researchers such as Wienke and co-workers (Wienke and Buydens, 1995; Wienke and Buydens, 1996; Wienke *et al.*, 1996), Hopke and co-workers (Song *et al.*, 1998), as well as Wang and co-workers (Wang *et al.*, 1999), in addition to the previous work by the authors (Whiteley and Davis, 1992) (Figure 1b).

ART-based clustering algorithms are especially sensitive to noise because of the inherent feature enhancement ability of ART coupled with the ability to remember rare events. The work by Frank *et al.* (Frank *et al.*, 1998) studied the clustering performance of fuzzy ART and ART-2 in the presence of noise and concluded that responsiveness to novel behavior can lead to non-optimal mapping because of the uncertain distinction between “novelty” and “noise”. Thus, the properties of Adaptive Resonance Theory that led to advantages in a noise-free environment do not necessarily offer similar benefits for noisy mappings (Marriott and Harrison, 1995). Several ART and ARTMAP variants have been proposed in the past to tackle this issue. The PROBART network proposed by Marriott and Harrison (Marriott and Harrison, 1995) stores probabilistic information about the node associations between ART layers to achieve a better performance in noisy mappings. A modified ARTMAP by Lim and Harrison (Lim and Harrison, 1997) was shown to approach Bayes optimal classification rates. The work by Srinivasa (Srinivasa, 1997) proposed a PROBART variant that improved its generalization ability in



(a) Nonlinear Mapping from a Chemical Process



(b) ART-2 for Incremental Diagnosis

Fig. 1. Fault Detection for Industrial Processes

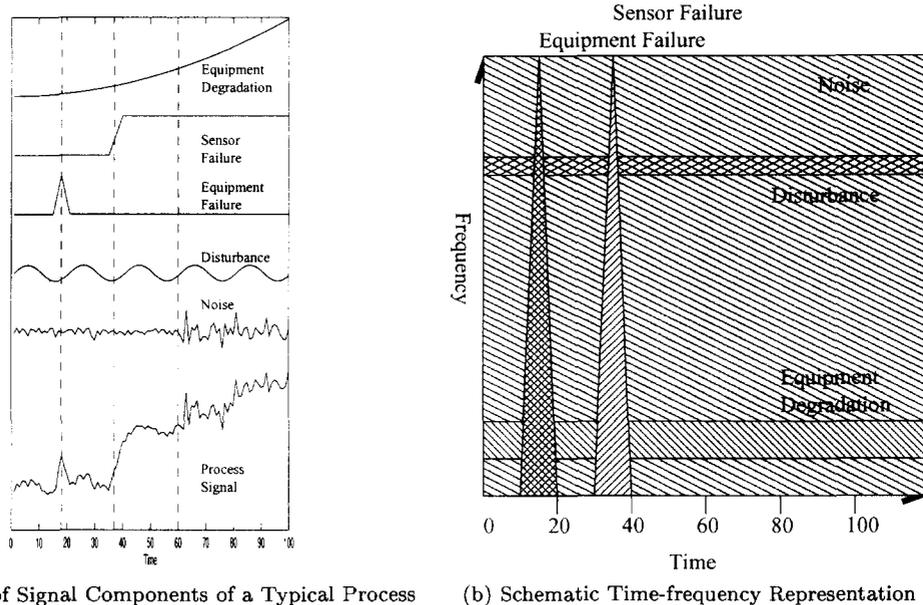
the context of high noise. Gaussian ARTMAP by Williamson (Williamson, 1996) combined a Gaussian classifier and an ARTMAP network by appropriately changing the definitions of ART *choice* and *match* functions. Recently, Wang and co-workers (Wang *et al.*, 1999) have proposed the use of wavelet feature extractors in place of the original data preprocessing and feature enhancement units within ART-2.

The current work approaches the problem of noise in ART mappings of digital signals in a manner fundamentally different than the research efforts discussed above. The proposed multiscale hierarchy of ART networks does not modify the internals of ART-2 in any way. As a result, the benefits of our mechanism are likely to be applicable even if any of the above ART variants were used as the basic unit of the hierarchy. Indeed, previous applications of our multiscale hierarchy have illustrated significant improvement in the performance of linear diagnosis methods based on PCA, Dynamic PCA, and a univariate Neyman-Pearson (NP) classifier (Bakshi, 1998; Bakshi *et al.*, 1999; Kano *et al.*, 2002). For an ideal case of a univariate Gaussian IID signal, the NP classifier can be theoretically proven to yield higher detection accuracy over a broad range of mean shifts if used with the proposed hierarchy (Aradhye *et al.*, 2000). This work combines the advantages of ART networks such as the ability to model nonlinear, disjoint process mappings and the incremental training ability with the benefits offered by multiresolution processing such as noise tolerance and quicker as well as more robust detection of events.

1.2 Wavelet Decomposition and Change Detection

Wavelets and multiresolution signal analysis (Mallat, 1989; Strang, 1989) have triggered developments in a range of process systems engineering related domains such as trend extraction (Bakshi and Stephanopoulos, 1994), process modeling (Palavajhala *et al.*, 1996), sensor validation (Luo *et al.*, 1998), noise reduction (Palavajhala *et al.*, 1996), etc. Advantages of these applications arise from the fact that most naturally occurring process signals are, in effect, a combination of various signal components corresponding to different events occurring at different localizations in time and frequency (Figures 2a and 2b). For example, equipment degradation occurs over wide time intervals and low frequencies. In contrast, sensor noise is spread across all frequencies and times. Events such as equipment failures are sharp, sudden changes that are localized in time but display components across all frequencies. As a result, specialized processing of the signal at different scales benefits tasks such as noise filtering and diagnosis.

A large body of published literature has investigated the use of wavelets for various forms of change detection. For example, the work by Crouse *et al.* (Crouse *et al.*, 1998) proposed a wavelet-domain Hidden Markov Model for univariate statistical signal processing. Swami, Sadler, and co-workers (Sadler *et al.*, 1998; Sadler and Swami, 1999; Sadler and Swami, 1998; Swami, 1996; Swami and Sadler, 1998b; Swami and Sadler, 1998a) have presented multiscale methods for step detection and estimation. Other researchers (Denjean and Castanie, 1994; Chou and Heck, 1994) have investigated wavelet-based shockwave



(a) Scales of Signal Components of a Typical Process (b) Schematic Time-frequency Representation

Fig. 2. Multiresolution Analysis of a Typical Process Signal

detection, mean value jump detection, monitoring of mechanical systems, and so on. These applications of multiresolution methods, including this work, are based on selection of wavelet coefficients for the purpose of retaining as much of the underlying process signal- and as little of the noise- as possible. Unlike these previous developments, however, the proposed multiscale hierarchy exploits clusters of wavelet coefficients of multiple process variables to provide a systematic way of selecting the most *relevant* scales. Because of fundamental functional relationships such as process chemistry, energy and mass balances, measurements in multivariate processes are correlated. If these intervariable correlations are linear, the resulting wavelet coefficients will be linearly correlated as well (Bakshi, 1998). Similarly, if the process variables are non-linearly correlated, the wavelet coefficients will be non-linearly correlated. The current work proposes to take advantage of these correlations and clustering behavior in the wavelet space for higher detection accuracy coupled with noise reduction.

2. BACKGROUND

2.1 Adaptive Resonance Theory

ART-2 is an unsupervised clustering mechanism proposed by Carpenter and Grossberg (Carpenter and Grossberg, 1987). Conventional clustering algorithms were designed to be synthesized off-line and lack the mechanism to adapt to dynamically evolving patterns. The objective of the analog ART-2 network is to “self-organize stable pattern recognition codes in response to arbitrary sequences of input patterns”. It imparts human-

like memory attributes which result in significant information management and system maintenance benefits. Later developments in the ART family of algorithms, such as ARTMAP and Fuzzy ARTMAP (Carpenter *et al.*, 1991c; Carpenter *et al.*, 1991a; Carpenter *et al.*, 1991b; Carpenter *et al.*, 1992), extended the basic principles of adaptive resonance for the purpose of supervised classification and function approximation.

For the purpose of diagnosis, the ART input space corresponds to the measurements of multiple process variables available at any time. Functional dependencies and constraints across process variables can be modeled as clusters of training data in this space: the underlying assumption being that abnormal behavior violates either these functional dependencies or the operating constraints. In either case, measurement vectors corresponding to anomalous behavior lie outside the clusters of normal data. When enough labeled data are available about a previously unknown abnormal operation, the ART-2 cluster space can be incrementally updated with prototypes that characterize the new behavior. Each cluster is associated with a particular process behavior in the form of a lookup table. The output space is thus the discrete space of possible diagnoses or classes.

The similarity measure is an ART-2 distance metric used to quantify the extent of match between the current measurement vector and the nearest cluster prototype. A similarity measure of 1 indicates an exact match, whereas a similarity measure of 0 indicates no match. The vigilance parameter is a cut-off such that a similarity measure greater than or equal to the vigilance is considered an acceptable match. A similarity measure below the vigilance represents an “unknown” process

condition. Implementation of ART-2 for fault diagnosis by Davis and co-workers uses a variable number of hyper-spherical clusters which are of fixed size. The lack of any orientation, incremental training, and overlapping coverage are some of its features distinct from other clustering-based diagnosis algorithms (e.g., (Kavuri and Venkatasubramanian, 1993)). It has been shown to be able to work consistently well over a wide range of simulated as well as real-life process situations (Whiteley and Davis, 1992; Whiteley and Davis, 1993; Whiteley *et al.*, 1996).

Due to the feature enhancement abilities of ART-2 clustering mechanism, however, an ART-2 based fault detector is vulnerable to process noise. For example, consider a multivariate, linearly correlated, noisy simulated process shown in Figure 3a. Abnormal operation was simulated as a mean shift added to all four variables from time-steps 176 through 225. Only normal data were used for training, so that the abnormal data were expected to be detected as an unknown event. Due to noise, however, we can see that normal and abnormal operations were not clearly separated. An ART-2 network was trained with independently generated normal data and was subjected to the test data. At each time step, the ART-2 similarity measure between the current four-dimensional data vector and stored cluster prototypes of normal data formed the basis for anomaly detection. For the given test data, the ART-2 similarity measures versus time are shown in Figure 3b-top. A similarity measure below the vigilance parameter indicated the absence of an acceptable winner cluster, and hence an “abnormal” state (Detection Flag = 1), as shown in the bottom graph. A similarity measure above the vigilance parameter indicated that a matching normal cluster was, indeed, found (Detection Flag = 0). We can see that many abnormal points were classified as normal (missed alarms). ART-2 diagnosis for such a noisy mapping was, thus, not robust. There was one false flag.

The use of several types of noise reduction filters, including wavelet-based filters, presents itself as a potential solution to the above noise vulnerability. This solution encounters the following two problems. First, the noise reduction or filtering step is clearly separated from the multivariate diagnosis step. The filtering step, thus, does not benefit from intervariable clustering behaviors that are typically present in real-life multivariate processes as shown in Figure 1a. Secondly, the diagnosis step is indifferent to which signal components were retained in the filtering step. To work around these issues, our approach integrates filtering and non-linear modeling for diagnosis. It also offers specialized processing according to the scales of the signal components retained in the filtered signal.

2.2 Wavelets

A well-known representation of a family of wavelet basis functions is:

$$\Psi_{su}(t) = \frac{1}{\sqrt{s}} \Psi \left(\frac{t-u}{s} \right) \quad (1)$$

where s and u represent the dilation and translation parameters, respectively, and $\Psi(t)$ is the mother wavelet.

If the translation parameter in a family of wavelets is discretized dyadically, $u = 2^m k$, the wavelet decomposition downsamples the coefficients at each scale. By convolution with the corresponding filters, any signal can be decomposed into its contributions at multiple scales as a weighted sum of dyadically discretized orthonormal wavelets.

$$y(t) = \sum_{m=m_0}^L \sum_{k=1}^N d_{mk} \Psi_{mk}(t) + \sum_{k=1}^N a_{Lk} \Phi_{Lk}(t) \quad (2)$$

where, y is the measurement, m_0 is the finest scale, L is the coarsest scale, d_{mk} are the detailed signal coefficients at scale m , and a_{Lk} are the scaled signal coefficients. We have typically chosen m_0 to be 1 for the examples studied in this paper.

Figure 4 illustrates the potential of wavelet decomposition for the task of fault detection of industrial process signals. As stated earlier (Figure 2a and 2b), a typical process signal is composed of a superimposition of several components such as sensor noise, disturbances, equipment degradation, and so on. By projecting the signal at increasingly coarser levels of resolution, the wavelet transform allows us to analyze each of these components at their respective frequencies and at the appropriate locations in time. Figure 4 shows successive approximations of the signal from Figure 2a using Haar wavelets and dyadic discretization. Equipment degradation can be observed at the lowest scaled signal \mathbf{a}_4 . Sudden events such as sensor and equipment failures can be observed across all detailed signals. For instance, the equipment failure from time-steps 35 through 40 can be seen at $\mathbf{d}_2[10]$, $\mathbf{d}_3[5]$, and $\mathbf{d}_4[3]$.

Decomposition of a signal by wavelets with dyadic downsampling implies that every measurement cannot be decomposed as soon as it is obtained. This can cause a time delay in many on-line applications of wavelets such as on-line filtering and statistical process monitoring. This time delay can be eliminated by decomposing the signal without downsampling, i.e., by discretizing the translation parameter as $u = k$. The wavelet coefficients lose their orthonormality but permit the devel-

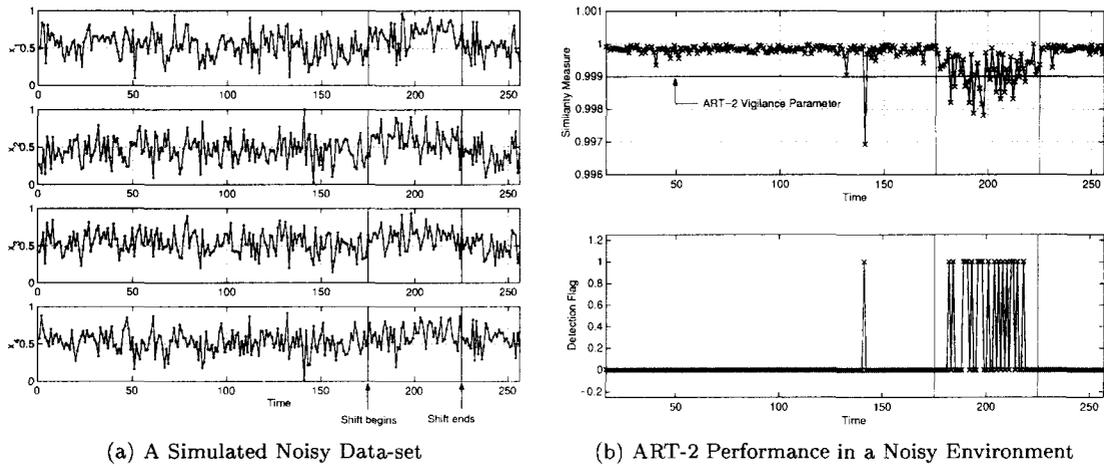


Fig. 3. An Example to Illustrate the Noise Sensitivity of the ART-2 Detector

opment of truly on-line multiscale methods. Our earlier work (Aradhye et al., 2000) has shown that wavelet decomposition with downsampling is more useful for monitoring of highly autocorrelated or non-stationary measurements, whereas, decomposition without downsampling is useful for diagnosis of uncorrelated or mildly autocorrelated measurements. In this work, we focus exclusively on transformations without downsampling as the emphasis here is on quick, online detection of faults.

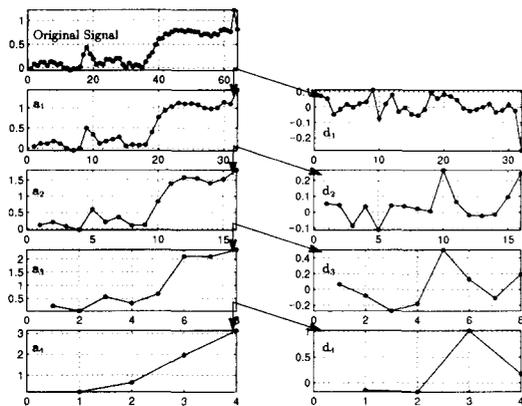


Fig. 4. Representation of Process Signals at Successive Levels of Approximation

3. ALGORITHM DESCRIPTION

Figure 5 shows a schematic diagram of the MSART-2 approach for online anomaly detection. Given the vigilance parameter ρ and the number of scales L , the following approach allows us to construct the ART-2 feature maps. Let P be the number of process variables in a multivariate process. All the constituent networks of the MSART-2 scheme cluster the data over a P -dimensional space of either the wavelet coefficients of these P variables on different scales, or the signals reconstructed by various combinations of wavelet coefficients.

3.1 Training

Consider an $N \times P$ matrix \mathbf{Y}^{train} of normal training data, where N is the number of training samples. During the training phase, the following steps synthesize normal clusters and thus capture the normal behavior of the process. We first apply the 1-D wavelet transform to each of the P variables to obtain detailed signal coefficients $d_{m,t,p}^{train}$ and the scaled signal coefficients $a_{L,t,p}^{train}$, where $m = 1, \dots, L$, $t = 2^L, \dots, N$, and $p = 1, \dots, P$. The illustration in Figure 5 uses a wavelet decomposition with $L = 4$. We then construct $L + 1$ training matrices that contain the corresponding detailed and scaled signal coefficients. ART-2 clustering is applied to each of these training matrices independently. Let the resulting cluster prototypes in the wavelet domain be represented as $ARTD_m$, $m = 1, \dots, L$, and $ARTA_L$, respectively. We thus have $L + 1$ ART-2 networks that constitute the *Scale Selection Layer* of wavelet-domain detectors. For example, Figure 5 shows a Scale Selection Layer composed of $ARTD_1$, $ARTD_2$, $ARTD_3$, $ARTD_4$, and $ARTA_4$, which represent clusters of wavelet coefficients of normal data at the respective scales.

At any time $t \geq 2^L$, the signal can be reconstructed in 2^{L+1} ways, depending on which of the $L + 1$ scales were selected for reconstruction. For each of the 2^{L+1} combinations, the coefficients corresponding to selected scales are retained for reconstruction. The remaining coefficients are reduced to zeros. Inverse wavelet transform is then applied. In this fashion, we generate training data matrices of reconstructed signals for each of the 2^{L+1} combinations.

Finally, we apply ART-2 clustering to each of these reconstructed training matrices independently to obtain cluster prototypes and associated weights in signal space filtered to retain the selected combination of scales. These 2^{L+1} ART-

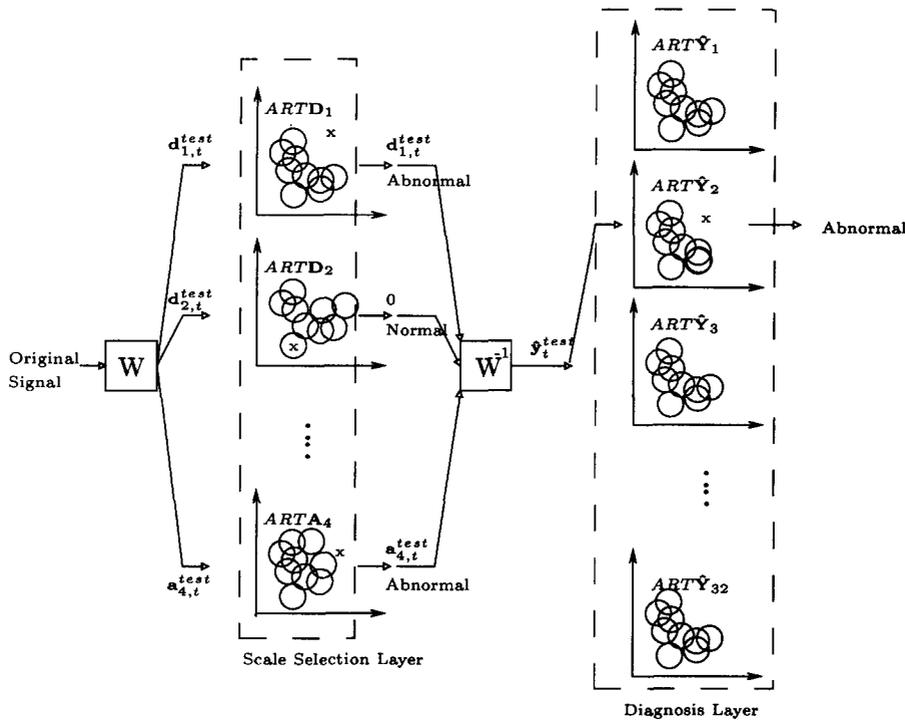


Fig. 5. The MSART-2 Architecture for Robust Fault Diagnosis

2 networks, $ART\hat{Y}_i, i = 1, \dots, 2^{L+1}$, constitute the *Diagnosis Layer* of detectors. In Figure 5, diagnoses of the 5 Scale Selection networks lead to $2^5 = 32$ possible ways in which the signal could be reconstructed. Correspondingly, the Diagnosis Layer in Figure 5 is composed of 32 ART-2 networks, each of which represents clusters of normal data reconstructed in one of the 32 possible ways.

When all scales are selected for reconstruction, the original signal matrix \mathbf{Y}^{train} is exactly reproduced for rows corresponding to $t \geq 2^L$. The corresponding Diagnosis Layer network is the same as the network used by Whiteley and Davis (1996). Hence the time-domain ART-based detector is a special case of the multiscale hierarchy presented in this work. A more detailed discussion of the algorithm and associated examples can be found in (Aradhye, 2001).

3.2 Online Testing

Having trained the Scale Selection Layer and Diagnosis Layer ART networks, we are now in a position to carry out online detection. At each time t , the following steps allow us to detect abnormalities using the proposed MSART-2 approach.

First, apply wavelet transform to decompose the P -dimensional signal vector \mathbf{y}_t^{test} into wavelet coefficients $d_{m,t,p}^{test}$ and $a_{L,t,p}^{test}$. Figure 5 shows a decomposition of a dyadic window of the test signal \mathbf{y}_t^{test} into coefficients $d_{1,t}^{test}, \dots, d_{4,t}^{test}$ and $a_{4,t}^{test}$.

Each of the scale selection networks then provides a diagnosis at the corresponding scale, based on whether the similarity between the input vector and the stored normal cluster prototypes is above the vigilance threshold. Only if the network $ARTD_m$ provides an “abnormal” diagnosis, the coefficients $d_{m,t,p}^{test}, p = 1, \dots, P$, are retained for reconstruction. Similarly, only if the network $ARTA_L$ provides an “abnormal” diagnosis, the coefficients $a_{L,t,p}^{test}$ are retained for reconstruction. For example, in Figure 5, the d_2 coefficient vector is deemed “normal” by $ARTD_2$. Hence, prior to the application of the inverse wavelet transform, the d_2 coefficients of all variables are reduced to zeros.

Lastly, apply inverse wavelet transform to the wavelet coefficients selected for reconstruction. The vector $\hat{\mathbf{y}}_t^{test}$, comprised of the reconstructed values for the P process variables, is presented as input to one of the 2^{L+1} $ART\hat{Y}$ Diagnosis Layer detectors corresponding to the combination of scales selected for reconstruction. For instance, the chosen Diagnosis Layer network in Figure 5 is trained on normal data that was wavelet-decomposed and reconstructed without the d_2 coefficients. Thus, the selected Diagnosis Layer network compares the reconstructed test signal at time t with prototypes of normal signals decomposed and reconstructed in exactly the same way. The resulting “normal” or “abnormal” diagnosis is provided to the user.

3.3 Parameter Selection

The tuning parameters in MSART include the depth of decomposition, the type of wavelet, and the vigilance values of individual ART networks.

With increasing depth of the wavelet decomposition, the ability of MSART to detect large shifts deteriorates due to the increasing time delay in obtaining the wavelet and scaling function coefficients at coarser scales. The ability to detect small shifts improves at greater depths due to greater separation between the stochastic variation and deterministic mean shift at coarser scales. As a rule of thumb, the depth of decomposition should correspond to the scale of the slowest/coarsest event expected in the data.

This paper has only focused on the use of Haar wavelets, but the approach may easily be used with other types of wavelets. The use of smoother, boundary-corrected wavelets may provide better performance than Haar wavelets due to better feature extraction and decorrelation abilities. Furthermore, extension to libraries of basis functions, such as wavelet packets, may permit MSART to automatically select the best family of basis functions from the library.

Different values of the vigilance parameter result in different overall false alarm rates. As is the case with most fault detection algorithms, if the false alarm rate for MSART is decreased (via increasing the vigilance parameter), the missed alarm rate increases. Tolerances to false alarms may vary for different process systems, and hence the choice of the vigilance value is application-specific. In general, it is possible to plot an operating curve of the false alarm rate vs. the missed alarm rate by experimenting with a range of vigilance values for the same process. According to process-specific tolerances, an appropriate value for the vigilance parameter can be chosen by selecting a point on this operating curve. The experiments discussed in this paper use the same set of ART configuration parameters, including the vigilance value, for all the Scale Selection Layer as well as Diagnosis Layer networks. All scales, thus, provide equally important information about detection of an event. As a result, the algorithm performs well as a general detection algorithm that can detect a broad range of events. With more specific information about the faults at hand, one may want to tailor the MSART detection system to specific types of events by adjusting the ART parameters at the relevant scales.

4. MONTE-CARLO PERFORMANCE COMPARISON

The Average Run-Length (ARL) curve is a standard for comparing the detection delays of two algorithms while keeping the false alarm rate constant. To generate the ARL curves, shifts of varying magnitudes were introduced at $t = 0$ to three illustrative process models provided in Table 1. For subsequent time-steps, simulated abnormal data were subjected to diagnosis by ART-2 as well as MSART-2 with $L = 2$. The time-step at which the shift was first detected (i.e., run-length) was recorded for each magnitude of shift for both detection algorithms. This simulation was repeated for 1000 instances and the run-lengths were averaged. Figure 6 shows a plot of ARL against the ratio of shift magnitude to the standard deviation of noise. As the shift magnitude increases, both algorithms take less time-steps for detection. For a broad range of shift magnitudes, however, MSART-2 detects the shift earlier. For small shifts, the process noise hampers the ability of ART-2 to consistently detect the shift. For very large shifts, however, ART-2 is seen to perform slightly better as the shift is easily separable from the inherent noise in the mapping.

5. INDUSTRIAL CASE STUDIES ON REAL-LIFE DATA

Results from the earlier section illustrated the utility of the proposed algorithms primarily with the help of simulated process models and/or specific detection limits. The analysis presented in this section displays the capability of our algorithms when compared with uniscale methods on univariate/multivariate real-life industrial data, collected from a large-scale petrochemical plant, over a broad range of false alarm rates. This analysis thus provides a detailed comparison of the performance of these methods on a broad range of detection operating regimes. In some cases, we also compare the performance of the human operator with that of the MSART algorithm. The depth of decomposition L was chosen to be 3.

Our detector implementations have been largely successful in tackling the above issues for a process of such a large scale. The following section describes a comparative study of the performance of the uniscale and multiscale versions of ART and PCA detectors. Note that for industrial case studies, the beginning and end of the events have to be manually determined with the help of operator annotations. This process is subjective and the results can be potentially influenced by the determination of the event duration. However, these examples help reinforce the conclusions drawn from the simulated case studies, for which exact

Table 1. Simulated Processes for Monte Carlo ARL Performance Comparison

Univariate Process	Linear Process	Nonlinear Process
$y(t) = x(t) = N(0, 1)$	$x_1(t) = N(0, 1)$	$r(t) = r(t - 1) - 0.001$ $\theta(t) = \theta(t - 1) + 2 * \pi * 0.006$ $x_1(t) = r(t) * \cos(\theta(t))$ $x_2(t) = r(t) * \sin(\theta(t))$ $y_i(t) = x_i(t) + \epsilon_i(t)$
	$x_2(t) = N(0, 1)$	
	$x_3(t) = \frac{x_1(t) + x_2(t)}{\sqrt{2}}$	
	$x_4(t) = \frac{x_1(t) - x_2(t)}{\sqrt{2}}$	
	$y_i(t) = x_i(t) + \epsilon_i(t)$	

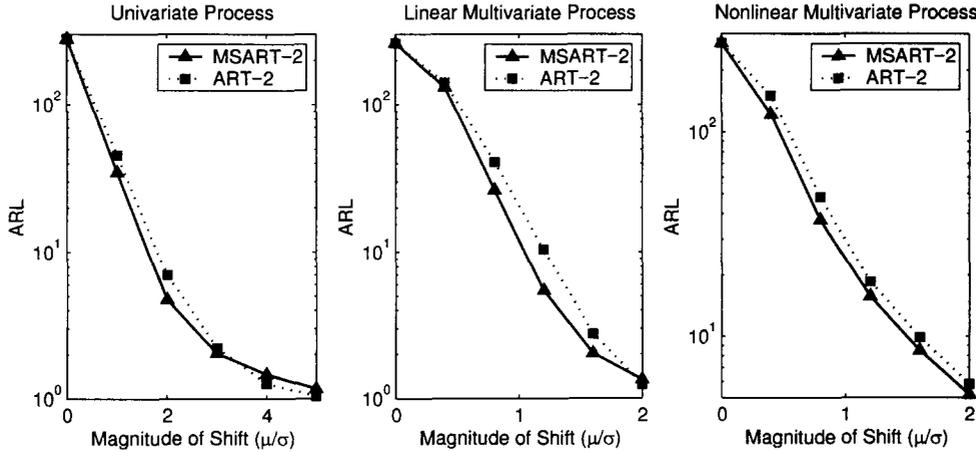


Fig. 6. Monte Carlo ARL Curves for Performance Comparison

information about the onset and reset of abnormal operation was available.

5.1 A Valve Leak Malfunction

Figure 7 shows the sensor readings during an abnormal process operation caused by a leaking valve. This event manifests itself primarily in three non-redundant sensors, trends of which are shown. As can be clearly observed, different sensors respond to the same root cause with different scales and delays due to differences in the underlying physical quantities being measured and also due to the process control scheme in place.

The performance curve from Figure 8 shows the average missed alarm rate plotted against the average false alarm rate. Different false alarm rates were achieved by changing the respective threshold parameters for the detection algorithms in question, namely PCA, MSPCA, ART, or MSART. For each false alarm rate, the corresponding average missed alarm rate was calculated for each abnormal event and plotted for each algorithm. It is obviously desirable to have the lowest possible missed alarm rate for a given false alarm rate.

The process variables involved in this event are linearly correlated, stationary, and approximately Gaussian. The underlying statistical assumptions for PCA are thus satisfied, leading to an expectation that it would model the process better than a generic neural, non-linear learning technique such

as ART. This expectation is seen to hold true, as PCA outperforms ART for the detection of this event. The proposed multiscale versions of these algorithms, MSPCA and MSART, outperform the respective single-scale versions. Due to the linear correlation between wavelet coefficients, corresponding to a linear correlation between process variables, MSPCA results in fewer false alarms than MSART.

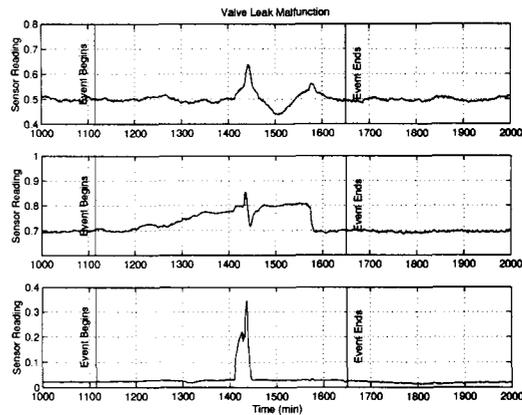


Fig. 7. Sensor Data for the ValveLeak Event

5.2 A Cold Weather Malfunction

Figure 9 shows our next industrial example, which involves a valve failure due to an unexpectedly lower ambient temperature. This is a univariate example which approximately obeys the assumptions of stationarity and unimodal Gaussian distribution. PCA is again seen to miss fewer alarms

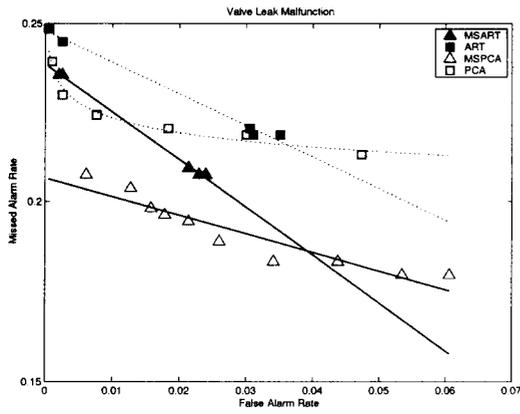


Fig. 8. Comparison of Performance for the Valve-Leak Event

when compared to ART as seen in Figure 10. Similarly, MSPCA performs better than MSART. As was the case of the previous example, the multiscale versions of PCA and ART outperformed the corresponding uniscale detectors.

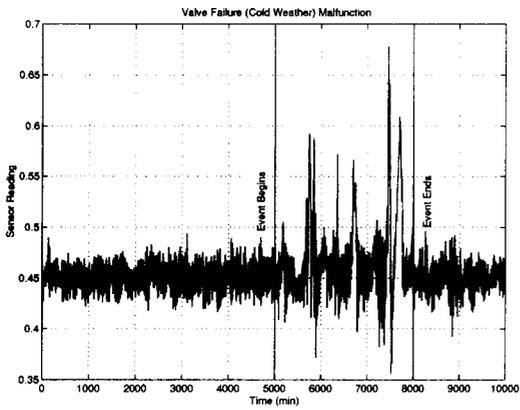


Fig. 9. Sensor Data for the Cold Weather Event

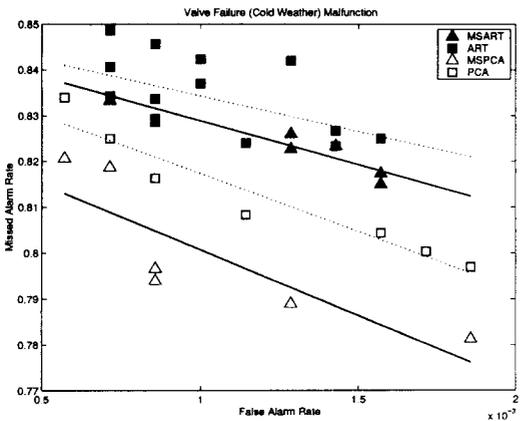


Fig. 10. Comparison of Performance for the Cold-Weather Event

5.3 A Change in Furnace Feed Event

Our third example corresponds to a “normal but unusual” univariate process operation associated

with a change in furnace feed. It was typically observed multiple times every day and lasted a few hours. Figure 11 shows several instances of this event. Although a high-level description of the sensor trend is seen to be similar for all instances, the event clearly showed characteristics for which multiscale analysis would be valuable, since the event lasted for varying intervals for different instances and rose to different magnitudes with differing approach and reset rates.

The performance curves for this event are shown in Figure 12. Unlike previous two examples, we observe that ART performs better than PCA, possibly due to deviations from ideal assumptions such as stationarity and Gaussian distribution. These deviations were seen to affect MSPCA more strongly than PCA, perhaps due to the strong auto-correlation in the data. However, MSART continues to outperform ART. The nonlinear modeling capability of ART makes it insensitive to non-stationarity and non-Gaussian behavior, and these benefits appear to hold in the wavelet domain as well.

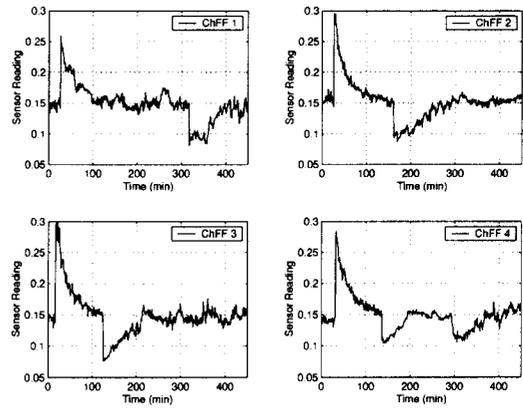


Fig. 11. Sensor Data for the Change in Furnace Feed Event

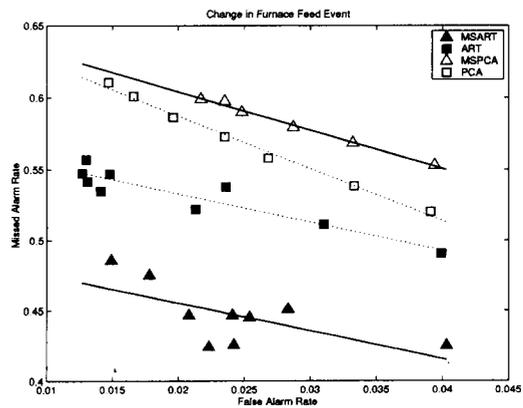


Fig. 12. Comparison of Performance for the Change in Furnace Feed Event

5.4 A Feed Malfunction

The data presented in Figure 13 show a distinctly nonlinear and multi-modal correlation in the bivariate sensor space. The process is seen to exist in three disjoint normal regimes. As a result, detector based on ART is expected to perform better than PCA for this process. Figure 14 shows this assertion to be true. Although the multiscale versions of these algorithms are seen to perform better, the improvement is more significant in the case of ART due to the nonlinear nature of the variable correlation.

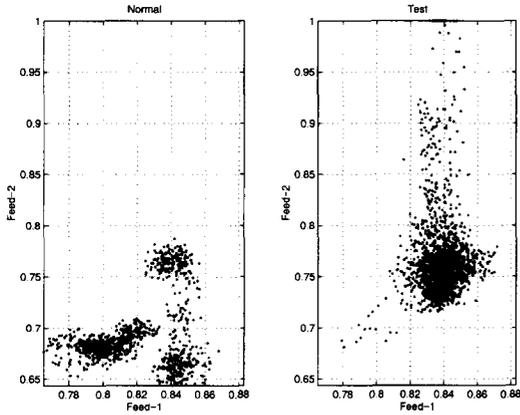


Fig. 13. Sensor Data for the FeedMalfunction Event

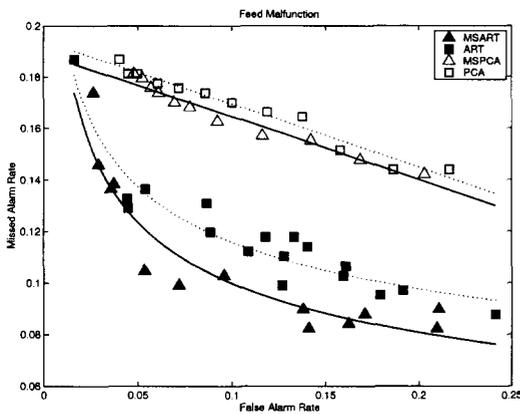


Fig. 14. Comparison of Performance for the Feed-Malfunction Event

6. COMPARISON WITH HUMAN OPERATOR

To provide a perspective of detection performance of the algorithms proposed in this work, we have provided two examples where the detection delay of a human operator is compared with that of our MSART algorithm. Figure 15 shows the detection flag of the MSART algorithm for the valve leak event (Figure 7) with 0 being normal and 1 being abnormal diagnosis. It can be seen that the

MSART algorithm detected the abnormal event hours in advance of the human operator.

Another example is provided with the help of Figures 16 and 17. This bivariate event was caused by a sensor “acting up”, i.e., providing erroneous reading. It can be seen that the MSART algorithm is a few minutes faster than the human operator, which is a significant fraction of the event duration.

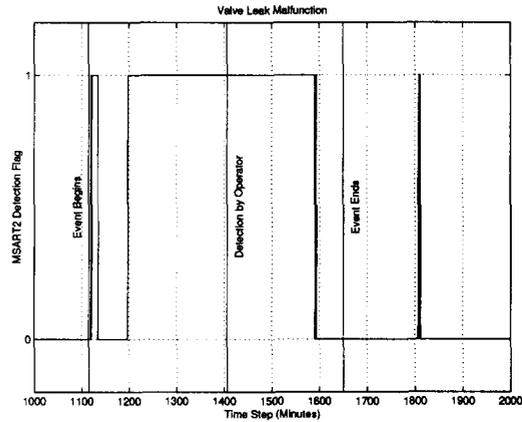


Fig. 15. Operator Detection of the ValveLeak Event

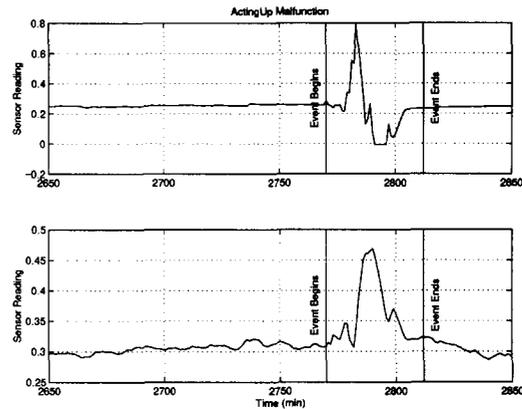


Fig. 16. Sensor Data for the ActingUp Event

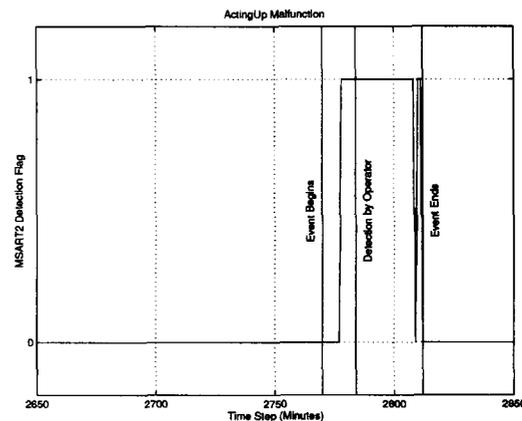


Fig. 17. Operator Detection of the ActingUp Event

7. CONCLUSION AND DISCUSSION

This work combines the advantages of ART networks such as the ability to model nonlinear, non-Gaussian, disjoint process mappings and the incremental training ability with the benefits offered by multiresolution processing such as noise tolerance, and quicker, more robust detection of events. Our results indicate that MSART-2, as compared to ART-2, is a general approach that is preferable for problems where it is necessary to detect all changes drawn from processes of various statistical characteristics. Our experiments with industrial data reveal that while ART-based diagnosis can comfortably detect unusual and abnormal operation earlier than the human operator, the MSART-2 algorithm can further reduce the detection delay while keeping the false alarm rate constant.

For a wavelet decomposition involving L scales, the worst-case computational requirement for MSART-2 is approximately $L + 2$ times the computation for the ART-2 detector. The worst-case storage requirement for MSART-2 is in fact approximately $L + 1 + 2^{L+1}$ times the storage requirement for ART-2.

Application of the ART-2 and PCA SPCs to industrial data brought out several features of interest regarding their function. The industrial data used in this work consisted of measurements from a total of 509 sensors distributed across different process units with close inter-connections. The readings of all the 509 sensors were provided every minute, with significant events annotated by the plant operators loosely as “normal”, “unusual”, and “abnormal”, and then specifically such as “furnace decoking”, “charge drier cool”, etc. SPC engines using PCA, ART, MSPCA, and MSART were built to detect these deviations from the normal process operation. This section provides a sense of the difficulties involved and lessons learned from the real-time industrial deployment of these SPC algorithms.

First of all, even with the best of efforts by the engineer, the sheer scale of the process often caused annotations to be inconsistent. An event listed as “unusual” for one day, for instance, may be omitted entirely for the next day. Hence, the data used for training had to be very carefully screened before use. Often, the information available to the operator came from sources outside of the 509 sensors and hence similar events ended up having different annotations. Although these issues adversely affected all SPCs being compared, the ART-based detectors were found to be specifically sensitive to contradicting annotations and overlapping class definitions. During incremental training, if new data overlapped in the sensor

space with previous clusters that belonged to a different class, the previous clusters were often completely overwritten, thus generating incorrect interpretations for future cases.

Secondly, some of the sensors may have been dysfunctional and exhibited erratic variations in their normal signatures, say of the order of 20% to 30% of the instrument span over a period of 24 hours. Removal of erratic or non-critical sensors and formulating the detector models again very significantly improved the detector performance in terms of false alarms. Selection of sensors that are a part of the detection models thus proved to be a very important aspect of the fault detection mechanisms.

Though most operations in the plant were continuous, the sensor trends were not entirely steady state. Often, there were slow drifts that were not completely captured by the data listed as “normal” in the training set, and hence were flagged by the detectors during the test phase. This generated a large number of notifications (on the order of 100 per day for a set of 370 sensors) that were unwelcome to the plant operators. Often, a sudden change, such as a set-point change, was deliberately introduced in the process by the plant operators. It was not possible to exhaustively provide examples of all such changes in the training data. Hence such changes were often flagged as abnormal by the detector, although they were in reality a part of the day-to-day plant operations. Exponential mean-filtering was used to make the ART and PCA detectors robust to mean-shifts and process drifts to a certain extent. However, it also masked genuine process anomalies that did need to be flagged by the detector. Hence the exponential filtering was later removed.

Since the sensor signals were being sampled per minute, the detector algorithms were expected to diagnose each snapshot of sensor data within a small fraction of a minute. However, the diagnosis operation in ART is an exhaustive search process. The time taken for a decision grows exponentially with the complexity of the search space, which is, in turn, a function of the dimensionality of the space in terms of the number of process variables and the nature of cross-variable relationships. However, as is the case with many large-scale problems, the relationships across process variables were mostly localized and related to the spatial position of the sensor in a process unit. To be able to manage scale and related issues such as speed, it was necessary to decompose the task of diagnosis into smaller subsets which were solved with smaller ART and PCA detectors. The decomposition was generated based on operator knowledge and reduced the associated structural and computational complexity to a large extent.

These *focused* SPCs resulted in a better diagnostic performance as well as were easier to train. In addition, the ART detector could adequately classify between different abnormal behaviors for input spaces with low dimensionality and less number of classes to distinguish between. However, it led to false alarms when the detector was trained with many different events. This problem was also solved with this focused detection approach.

Lastly, there was an expectation that the detector algorithm provide a list of sensors that contributed the most to the detector decision. A knowledge of the contributing sensors enables the operator to track the root cause of the abnormality. For PCA and MSPCA, a mechanism for calculating the contribution charts (Miller *et al.*, 1998) has been developed. For ART and MSART, similar calculations based on the directions in the sensor space that most contribute to the difference between actual and expected sensor readings lead to a list of three most contributing sensors. Although *ad hoc*, this method often provided an accurate list of contributing sensors.

The observations listed above are not specific to this particular industrial implementation. This work has been one of the first of its kind in terms of the complexity of the process and the scale of the deployment in terms of the number of sensors and process units involved. Future exercises of this kind can benefit significantly from the lessons learned from this deployment.

ACKNOWLEDGEMENTS

We gratefully acknowledge partial financial support from the Abnormal Situation Management Consortium, the Presidential Fellowship program of the Ohio State University, and the National Science Foundation through CAREER award, CTS-9733627.

NOMENCLATURE

Symbol	Description
s	Wavelet dilation parameter
u	Wavelet translation parameter
$\Psi(t)$	Mother wavelet function
y	Measured signal
m	Scale
m_0	Finest scale
L	Coarsest scale
N	Number of samples
Φ_{mk}	The k^{th} scaling function at scale m
d_{mk}	k^{th} detailed signal coefficient at scale m
a_{Lk}	k^{th} scaled signal coefficient
\mathbf{d}_m	Detailed signal coefficient vector at scale m
\mathbf{a}_L	Scaled signal coefficient vector
ρ	ART vigilance parameter
P	Number of process variables
$\mathbf{Y}^{\text{train}}$	Training data matrix
$d_{m,t,p}^{\text{train}}$	The detailed signal coefficient of training data for variable p , at scale m , and time t
$a_{L,t,p}^{\text{train}}$	The scaled signal coefficient of training data for variable p and time t
ARTD_m	The <i>Scale Selection Layer</i> ART network trained on the detailed signal coefficient vector at scale m
ARTA_L	The <i>Scale Selection Layer</i> ART network trained on the scaled signal coefficient vector
$\text{ART}\hat{\mathbf{Y}}_i$	The i^{th} <i>Diagnosis Layer</i> network
$d_{m,t,p}^{\text{test}}$	The detailed signal coefficient of test data for variable p , at scale m , and time t
$a_{L,t,p}^{\text{test}}$	The scaled signal coefficient of test data for variable p and time t

8. REFERENCES

- Aradhye, H. B. (2001). Anomaly Detection Using Multiscale Methods. PhD thesis. The Ohio State University.
- Aradhye, H. B., B. R. Bakshi, J. F. Davis and S. C. Ahalt (2000). Clustering in wavelet domain - a multiresolution art network for anomaly detection. *IEEE Transactions on Systems, Man and Cybernetics*.
- Bakshi, B. R. (1998). Multiscale PCA with application to multivariate statistical process monitoring. *AIChE Journal* **44**(7), 1596.
- Bakshi, B. R. and G. Stephanopoulos (1994). Representation of process trends- iv. induction of real-time patterns from operating data for diagnosis and supervisory control. *Computers and chemical Engineering*. **18**(4), 303.
- Bakshi, B. R., H. Aradhye and R. Strauss (1999). *Process Monitoring by PCA, Dynamic PCA,*

- and Multiscale PCA - Theoretical Analysis and Disturbance Detection in the Tennessee Eastman Process. AICHE Annual Meeting, Dallas, TX.
- Carpenter, G. A. and S. Grossberg (1987). A massively parallel architecture for a self-organizing neural pattern recognition machine. *Computer Vision, Graphics and Image Processing* **37**, 54.
- Carpenter, G. A., S. Grossberg and D. B. Rosen (1991a). Fuzzy ART: An adaptive resonance algorithm for rapid, stable classification of analog patterns. In: *Proceedings of the International Joint Conference on Neural Networks*. Vol. 2.
- Carpenter, G. A., S. Grossberg and D. B. Rosen (1991b). Fuzzy ART: Fast, stable learning and categorization of analog patterns by an adaptive resonance system. *Neural Networks* **4**, 759.
- Carpenter, G. A., S. Grossberg and J. H. Reynolds (1991c). ARTMAP: Supervised real-time learning and classification of nonstationary data by a self-organizing neural network. *Neural Networks* **4**, 565.
- Carpenter, G. A., S. Grossberg, N. Markuzon, J. H. Reynolds and D. B. Rosen (1992). Fuzzy ARTMAP: A neural network architecture for incremental supervised learning of analog multidimensional maps. *IEEE Transactions on Neural Networks* **3**(5), 698.
- Chou, K. C. and L. P. Heck (1994). A multiscale stochastic modeling approach to the monitoring of mechanical systems. In: *Proceedings of the IEEE International Symposium on Time-Frequency, Time-Scale Analysis*.
- Crouse, M. S., R. D. Nowak and R. G. Baraniuk (1998). Wavelet-based statistical signal processing using hidden Markov models. *IEEE Transactions on Signal Processing* **46**(4), 886.
- Denjean, A. and F. Castanie (1994). *Wavelets: Theory, Algorithms, and Applications*. Chap. Mean value jump detection: A survey of conventional and wavelet based methods. Newkoski Academic.
- Frank, T., K. Friedrich and T. Kuhlen (1998). Comparative analysis of fuzzy ART and ART-2A network clustering performance. *IEEE Transactions on Neural Networks* **9**(3), 544.
- Kano, M., K. Nagao, S. Hasebe, I. Hashimoto, H. Ohno, R. Strauss and B. Bakshi (2002). Comparison of statistical process monitoring methods: Application to the eastman challenge problem. *Computers and Chemical Engineering, Special Issue on Selected Papers from PSE 2000 (accepted)*.
- Kavuri, S. N. and V. Venkatasubramanian (1993). Representing bounded fault classes using neural networks with ellipsoidal activation functions. *Computers and Chemical Engineering* **17**, 139.
- Lim, C. P. and R. F. Harrison (1997). Modified fuzzy ARTMAP approaches Bayes optimal classification rates: An empirical demonstration. *Neural Networks* **10**(4), 755.
- Luo, R., M. Misra, S. J. Qin, R. Barton and D. M. Himmelblau (1998). Sensor fault detection via multiscale analysis and nonparametric statistical inference. *Industrial Engineering and Chemistry Research* **37**, 1024.
- Mallat, S. G. (1989). A theory for multiresolution signal decomposition: The wavelet representation. *IEEE Transactions and Pattern Analysis and Machine Intelligence* **11**(7), 674.
- Marriott, S. and R. F. Harrison (1995). A modified fuzzy ARTMAP architecture for the approximation of noisy mappings. *Neural Networks* **8**(4), 619.
- McDowell, J.K., M.A. Kramer and J.F. Davis (1991). Knowledge-based diagnosis in process engineering. *IEEE Expert - Intelligent Systems and Their Applications*.
- Miller, P., R. E. Swanson and C. E. Heckler (1998). Contribution plots: A missing link in multivariate quality control. *Applied Mathematics and Computational Science* **8**(4), 775.
- Palavajhala, S., R.L. Motard and B. Joseph (1996). Process identification using discrete wavelet transforms: Design of prefilters. *AICHE Journal* **42**(3), 777.
- Park, J.H. and P.H. Seong (1994). Nuclear power plant pressurizer fault diagnosis using fuzzy signed digraph and spurious faults eliminations methods. *Annals of Nuclear Energy*.
- Qian, D.Q. (1990). An improved method for fault location of chemical plants. *Computers and Chemical Engineering*.
- Rojas-Guzman, C. and M.A. Kramer (1994). Multi-stage bayesian networks subsume digraph and residual pattern approaches to fault diagnosis. In: *Proceedings of the Fifth International Symposium on Process Systems Engineering*.
- Sadler, B. M. and A. Swami (1998). On multiscale wavelet analysis for step estimation. *International Conference on Acoustics, Speech, and Signal Processing, Seattle, WA* **3**, 1517.
- Sadler, B. M. and A. Swami (1999). Analysis of multiscale products for step detection and estimation. *IEEE Transactions on Information Theory* **45**(3), 1043.
- Sadler, B. M., T. Pham and L. C. Sadler (1998). Optimal and wavelet-based shockwave detection and estimation. *Journal of the Acoustic Society of America* **104**(2), 955.
- Song, X., P. K. Hopke, M. Bruns D. A. Bossio and K. M. Scow (1998). A fuzzy adaptive resonance theory- supervised predictive mapping neural network applied to the classification

- of multivariate chemical data. *Chemometrics and intelligent laboratory systems* **41**, 161.
- Srinivasa, N. (1997). Learning and generalization of noisy mappings using a modified PRO-BART neural network. *IEEE Transactions on Signal Processing* **45**(10), 2533.
- Strang, Gilbert (1989). Wavelets and dilation equations: A brief introduction. *Society for Industrial and Applied Mathematics* **31**(4), 614.
- Swami, A. (1996). Cramer-Rao bounds for deterministic signals in additive and multiplicative noise. *Signal Processing* **53**, 231.
- Swami, A. and B. M. Sadler (1998a). Cramer-Rao bounds for step-change localization in additive and multiplicative noise. In: *9th IEEE Statistical Signal and Array Processing Workshop*. p. 403.
- Swami, A. and B. M. Sadler (1998b). Step-change localization in additive and multiplicative noise via multiscale products. In: *32nd Asilomar Conference on Signal, System, and Computation*. Pacific Grove, CA.
- Wang, X. Z., B. H. Chen, S. H. Yang and C. McGreavy (1999). Application of wavelets and neural networks to diagnostic system development, 2, an integrated framework and its application. *Computers and Chemical Engineering* **23**, 945.
- Whiteley, J. R. and J. F. Davis (1992). Knowledge-based interpretation of sensor patterns. *Computers and Chemical Engineering* **16**(4), 329.
- Whiteley, J. R. and J. F. Davis (1993). Qualitative interpretation of sensor patterns. *IEEE Expert- Intelligent Systems and Their Applications* **8**(2), 54.
- Whiteley, J. R., J. F. Davis, A. Mehrotra and S. C. Ahalt (1996). Observations and problems applying ART2 for dynamic sensor pattern interpretation. *IEEE Transactions on Systems, Man and Cybernetics Part A- Systems on Humans* **26**(4), 423.
- Wienke, D. and L. Buydens (1995). Adaptive resonance theory based neural networks - the 'ART' of real-time pattern recognition in chemical process monitoring?. *Trends in analytical chemistry* **14**(8), 398.
- Wienke, D. and L. Buydens (1996). Adaptive resonance theory based neural network for supervised chemical pattern recognition (fuzzy ARTMAP), part 1: Theory and network properties. *Chemometrics and intelligent laboratory systems* **32**, 151.
- Wienke, D., W. van den Broek, L. Buydens, T. Huth-Frehre and R. Feldhoff (1996). Adaptive resonance theory based neural network for supervised chemical pattern recognition (fuzzy ARTMAP), part 2: Classification of post-consumer plastics by remote NIR spectroscopy using an InGaAs diode array. *Chemometrics and intelligent laboratory systems* **32**, 165.
- Wilcox, N.A. and D.M. Himmelblau (1994). The possible cause and effect graphs model for fault diagnosis. *Computers and Chemical Engineering*.
- Williamson, J. R. (1996). Gaussian ARTMAP: A neural network for fast incremental learning of noisy multidimensional maps. *Neural Networks* **9**(5), 881.
- Yu, C.C. and C. Lee (1991). Fault diagnosis based on qualitative/quantitative knowledge. *AIChE Journal*.