Clustering in Wavelet Domain: A Multiresolution ART Network for Anomaly Detection

Hrishikesh B. Aradhye
Machine Vision and Robotics Group, SRI International, Menlo Park, CA

Bhavik R. Bakshi
Dept. of Chemical Engineering, The Ohio State University, Columbus, OH 43210

James F. Davis
Dept. of Chemical Engineering, University of California at Los Angeles, Los Angeles, CA 90095

Stanley C. Ahalt
Dept. of Electrical Engineering, The Ohio State University, Columbus, OH 43210

DOI 10.1002/aic.10245
Published online in Wiley InterScience (www.interscience.wiley.com).

A method for process fault detection is presented, based on the integration of multiscale signal representation and scale-specific clustering-based diagnosis. Previous work has demonstrated the utility of our multiscale detection scheme applied to linear projection-based methods, such as PCA and Dynamic PCA. This work further demonstrates the use and method independence of the multiscale scheme by applying it to a nonlinear modeling method, namely Adaptive Resonance Theory-2. The multiscale ART-2 (MSART-2) algorithm detects a process change when one or more wavelet coefficients violate the similarity thresholds with respect to clusters of wavelet coefficients under normal process operation at that scale. 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. By reconstructing the signal with only the relevant scales, MSART-2 can automatically extract the signal feature representing the abnormal operation under consideration. Illustrative examples as well as Monte Carlo bases for these claims via a comparative performance analysis over several case studies are provided. Comparison of average detection delays or run-lengths of MSART-2 with those of ART-2 for a variety of processes with different statistical characteristics is provided. Comparative results on real industrial case studies from a petrochemical process plant are also presented. Results indicate that MSART-2, as compared to ART-2, is a general approach that may be preferable for problems where it is necessary to detect all changes drawn from processes of various statistical characteristics. © 2004 American Institute of Chemical Engineers *AIChE J,* 50: 2455–2466, 2004

Keywords: statistical process control; process monitoring; adaptive resonance theory; fault detection; wavelets; average run-length

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 and isolation.* 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,

Correspondence concerning this article should be addressed to B. R. Bakshi at bakshi2@osu.edu.

© 2004 American Institute of Chemical Engineers

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 investigates its performance within 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. Our results compare the performance of ART-2 diagnostic models with and without the proposed multiscale framework to emphasize the utility and versatility of the hierarchy. However, comparison of the diagnostic ability ART-2 with other modeling/machine learning techniques is beyond the scope of this article. The reader is referred to the article by Carpenter et al. (Carpenter and Grossberg, 1987; Carpenter et al., 1991a, b, c, 1992) for a comparative analysis of ART-based methods applied to various problems of theoretical and/or practical interest.

Most real-world large-scale industrial processes, by their inherent nature, are not precisely defined in the space of sensor measurements. Clustering-based models approximate 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). Specifically, the ART family of networks (Carpenter and Grossberg, 1987; Carpenter et al., 1991a, b, c, 1992) includes some of the few clustering algorithms that explicitly address the issue of stable adaptation and incremental learning with changing process behavior. 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, 1996; Wienke et al., 1996), Hopke and coworkers (Song et al., 1998), as well as Wang and coworkers (Wang et al., 1999), in addition to the previous work by the authors (Whiteley and Davis, 1992). However, 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 nonoptimal 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 (Marriott and Harrison, 1995; Lim and Harrison, 1997; Srinivasa, 1997; Williamson, 1996; Wang et al., 1999).

This work approaches the problem of noise in ART mappings of digital signals in a manner fundamentally different than the research efforts discussed earlier. 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 earlier 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; Aradhye et al., 2003). 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., 2003). 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.

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 (Nounou and Bakshi, 1999), sensor validation (Luo et al., 1998), noise reduction (Palavajjhala et al., 1996), and so on. 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. Unlike 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 nonlinearly correlated, the wavelet coefficients will be nonlinearly correlated. This work proposes to take advantage of these correlations and clustering behavior in the wavelet space for higher detection accuracy coupled with noise reduction. Our earlier work (Aradhye et al., 2003) has shown that wavelet decomposition with downsampling is more useful for monitoring of highly autocorrelated or nonstationary 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 in a nonlinear system.

Description of the Proposed MSART-2 Algorithm

Figure 1 shows 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 that constitute the MSART-2 architecture. 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.

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

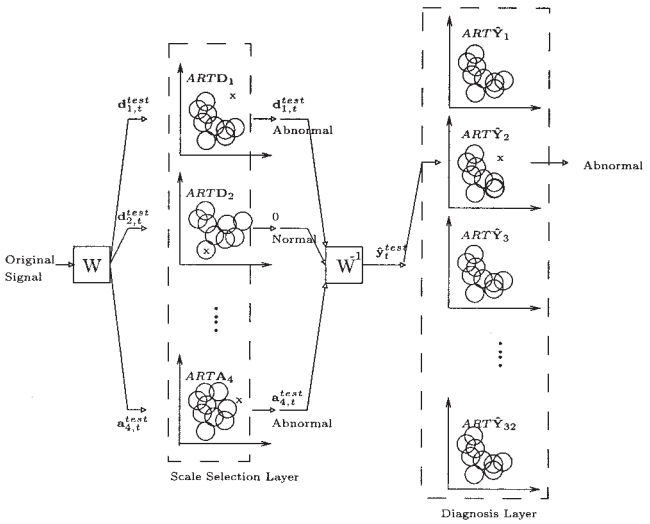


Figure 1. MSART-2 architecture for robust fault diagnosis.

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 $m = 1, \ldots, L$, $t = 2^L, \ldots, N$, and $p = 1, \ldots, P$. The illustration in Figure 1 used a wavelet decomposition with $L = 4$. We then construct $L + 1$ training matrices $\mathbf{D}_m^{train}$, $m = 1, \ldots, L$, and $\mathbf{A}_L^{train}$, each of size $N - 2^L + 1 \times P$, that contain the corresponding detailed and scaled signal coefficients. ART-2 clustering is independently applied to each of these training matrices. Let the resulting cluster prototypes in the wavelet domain be represented as $ART\mathbf{D}_m$, $m = 1, \ldots, L$, and $ART\mathbf{A}_L$, respectively. We, thus, have $L + 1$ ART-2 networks that constitute the *scale selection layer* of wavelet-domain detectors. For example, Figure 1 shows a scale selection layer composed of $ART\mathbf{D}_1$, $ART\mathbf{D}_2$, $ART\mathbf{D}_3$, $ART\mathbf{D}_4$, and $ART\mathbf{A}_4$, which represent clusters of wavelet coefficients of normal data at the respective scales.

A crucial feature of the MSART-2 architecture is the reconstruction of the signal based on only the relevant scales. By replacing all except the relevant scales by zeros before applying the inverse wavelet transform, the reconstructed signal is made to conform to the nature of the change under consideration in terms of its magnitude and rate of change. We, thus, filter out the unnecessary details of the process from the point of view of the change under consideration. 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. Let these matrices be $\hat{\mathbf{Y}}_1^{train}$, $\hat{\mathbf{Y}}_2^{train}$, $\ldots$, $\hat{\mathbf{Y}}_{2^{L+1}}^{train}$, each of which is of size $(N - 2^L + 1) \times P$. The data points for $t < 2^L$ are not reconstructed since all the wavelet coefficients are available only for $t = 2^L, \ldots, N$.

Finally, we apply ART-2 clustering independently to each of these reconstructed training matrices to obtain cluster prototypes and associated weights in signal space filtered to retain the selected combination of scales. These 2^{L+1} ART-2 networks, $ART\hat{\mathbf{Y}}_i$, $i = 1, \ldots, 2^{L+1}$, constitute the *diagnosis layer* of detectors. In Figure 1, 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 1, 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. Hence, the time-domain ART-based detector is a special case of the multiscale hierarchy presented in this work.

Online Testing

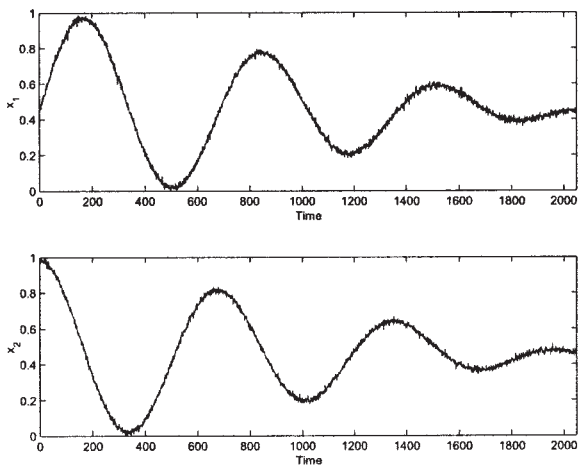
Having trained the scale selection layer and diagnosis layer ART networks, we are now in a position to carry out online detection. At each time t, the following steps allow us to detect abnormalities using the proposed MSART-2 approach:

(1) 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 1 shows a decomposition of a dyadic window of the test signal y_t^{test} into coefficients $d_{1,t}^{test}, \ldots, d_{4,t}^{test}$ and $a_{4,t}^{test}$. For each scale m, construct a P-dimensional vector $\mathbf{d}_{m,t}^{test}$, comprising of coefficients $d_{m,t,p}^{test}$ with $p = 1, \ldots, P$. This vector is presented as input to detector $ART\mathbf{D}_m$ of the Scale Selection Layer. Similarly, construct the vector $\mathbf{a}_{L,t}^{test}$ to be presented to the detector $ART\mathbf{A}_L$.

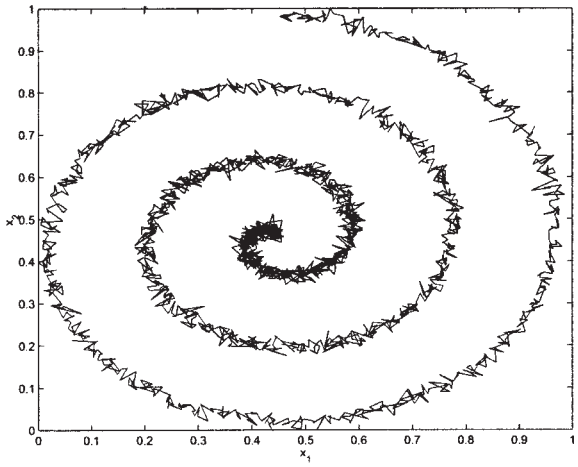
(2) Each of these networks provides a diagnosis at the corresponding scale, based on whether the similarity between the input vector and the stored normal cluster prototypes is above the vigilance threshold. Only if the network $ART\mathbf{D}_m$ provides an "abnormal" diagnosis, the coefficients $d_{m,t,p}^{test}$, $p = 1, \ldots, P$, are retained for reconstruction. Similarly, only if the network $ART\mathbf{A}_L$ provides an "abnormal" diagnosis, the coefficients $a_{L,t,p}^{test}$ are retained for reconstruction. For example, in Figure 1, the d_2 coefficient vector was deemed "normal" by $ART\mathbf{D}_2$. Hence, prior to the application of the inverse wavelet transform, the d_2 coefficients of all variables were reduced to zeros.

(3) 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} $ART\hat{\mathbf{Y}}$ diagnosis layer detectors corresponding to the combination of scales selected for reconstruction. For instance, the chosen diagnosis layer network in Figure 1 was 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.

The added benefits of our method come at a cost of increased computation and storage requirements. 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



(a) Normal Data for Cases One and Two



(b) Variable Correlations Under Normal Conditions

Figure 2. A bi-variate process for illustration of the MSART-2 approach.

requirement for MSART-2 is in fact approximately $L + 1 + 2^{L+1}$ times the storage requirement for ART-2.

Illustration of the MSART-2 Algorithm

In this section, we present three case studies that illustrate the advantages of our approach in more detail. To facilitate a visual representation, let us limit ourselves to two variables, although the method is general, and can be applied to data with any number of variables. The three cases differ in terms of noise and the extent of separation between normal and abnormal operation.

The experiments discussed in this article use the same set of parameters 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.

Figure 2a shows the normal behavior of the process considered in this illustration. The input vector $x(t)$ consisted of measurements of two nonlinearly correlated process variables $x_1(t)$ and $x_2(t)$. A bivariate problem was chosen for visual simplicity, although the algorithms considered are multivariate. Gaussian noise was superimposed on the data to simulate noisy conditions. Figure 2b illustrates the nonlinear correlation between these two process variables when plotted against each other. Simulated faults included shifts of differing magnitudes among differing levels of noise, followed by resumption of normal behavior. The test signals were subjected to online diagnoses by applying (1) an ART-2 detector, (2) a moving average (MA) filter, followed by an ART-2 detector (referred to as ART-2+MA), and (3) an MSART-2 detector. Comparative analyses brought out the strengths and weaknesses of the current approach with respect to the basic ART-2 based detection/diagnosis.

Please note that the noisy mapping in this case is random, and the exact diagnoses may differ for different instantiations of the random process. We go on to establish the utility of the current approach in a Monte Carlo fashion in the next section.

Case One: A low-noise process with a clearly separable shift

Figure 3a shows the test data used for diagnosis in this section. As can be seen as the left side of the outermost arm of the spiral, a shift was introduced to simulate abnormal behavior from time-step 501 to time-step 550. The number of scales, L, was chosen to be 4.

Scale Selection. The detection flags of the decomposed signals provide an insight into the mechanism of scale selection in the MSART-2 architecture. Figure 3b shows the diagnoses by the resulting 5 Scale selection layer networks for a part of the test signal. As explained earlier, the scale selection layer subjects each wavelet coefficient of the test data ($d_1, \ldots, d_4$, and a_4 in this case) independently to an ART-2 network trained exclusively on the corresponding coefficients of training normal data. In Figure 3b, a detection flag of 0 indicates a "normal" diagnosis, whereas a detection flag of 1 implies an "abnormal" diagnosis. The overall diagnosis, that is, the diagnosis on the reconstructed signal (Figure 3b: bottom-most graph), illustrates the effect of simultaneous selection of multiple scales. Figure 3b shows that when the abnormal region started at time-step 501, the mean shift was detected immediately by $ART\mathbf{D}_1$, the network trained with the finest detailed component of normal data. Scale selection networks at the subsequent (coarser) detailed scales, $ART\mathbf{D}_2, \ldots, ART\mathbf{D}_4$, detected the shift at subsequent points in time. Since the level of detail became coarser at lower frequencies, the duration for which the shift was detected increases from 1 time-step to 16 time-steps as we go from d_1 to d_4.

Similarly, when the normal operation resumed at time-step 550, the transition was detected in the order of the finest to the coarsest scale. Except for the transitional region, the fault was reflected only in the residual signal (a_4) for most parts. The residual signal is equivalent to that generated after applying 16

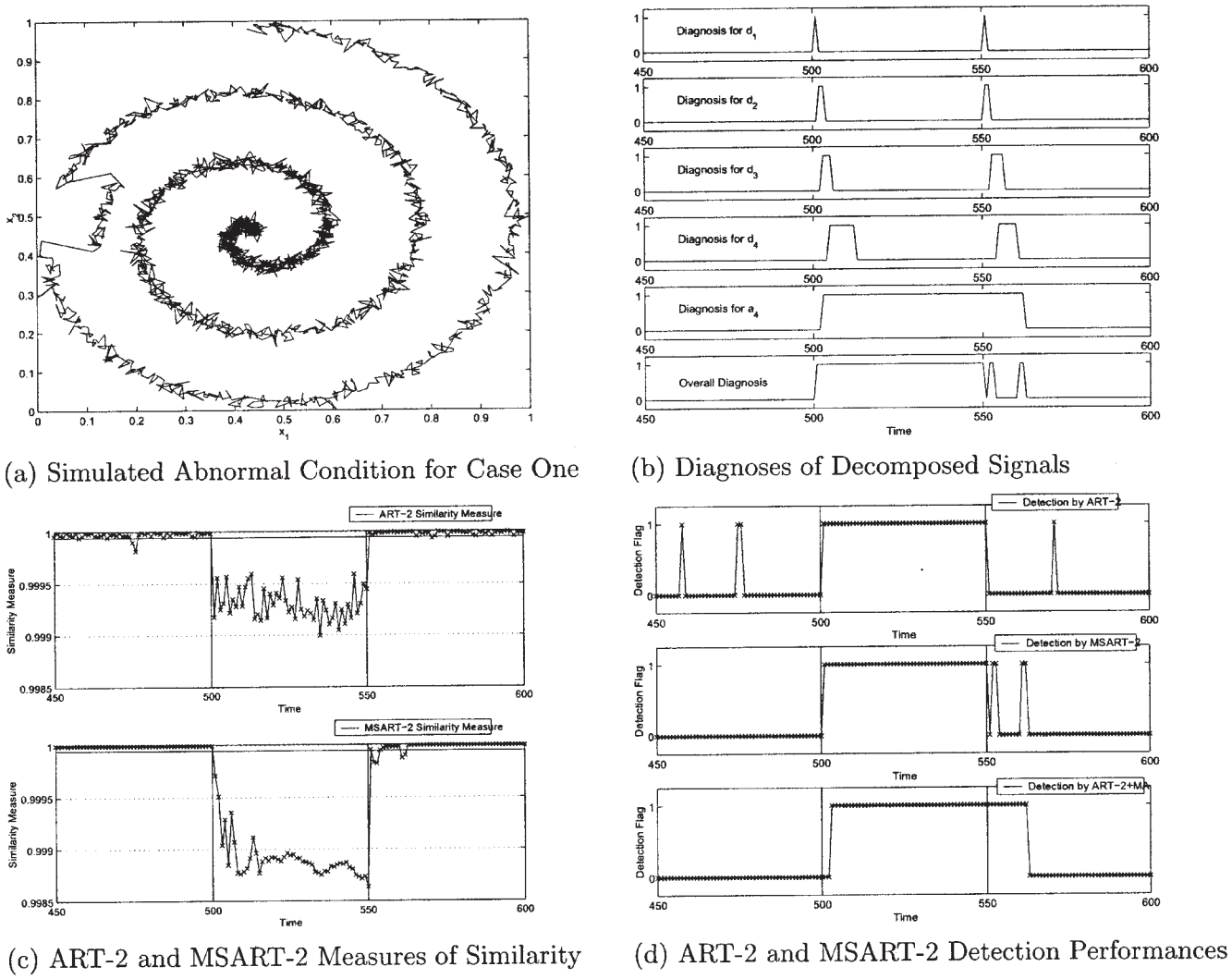


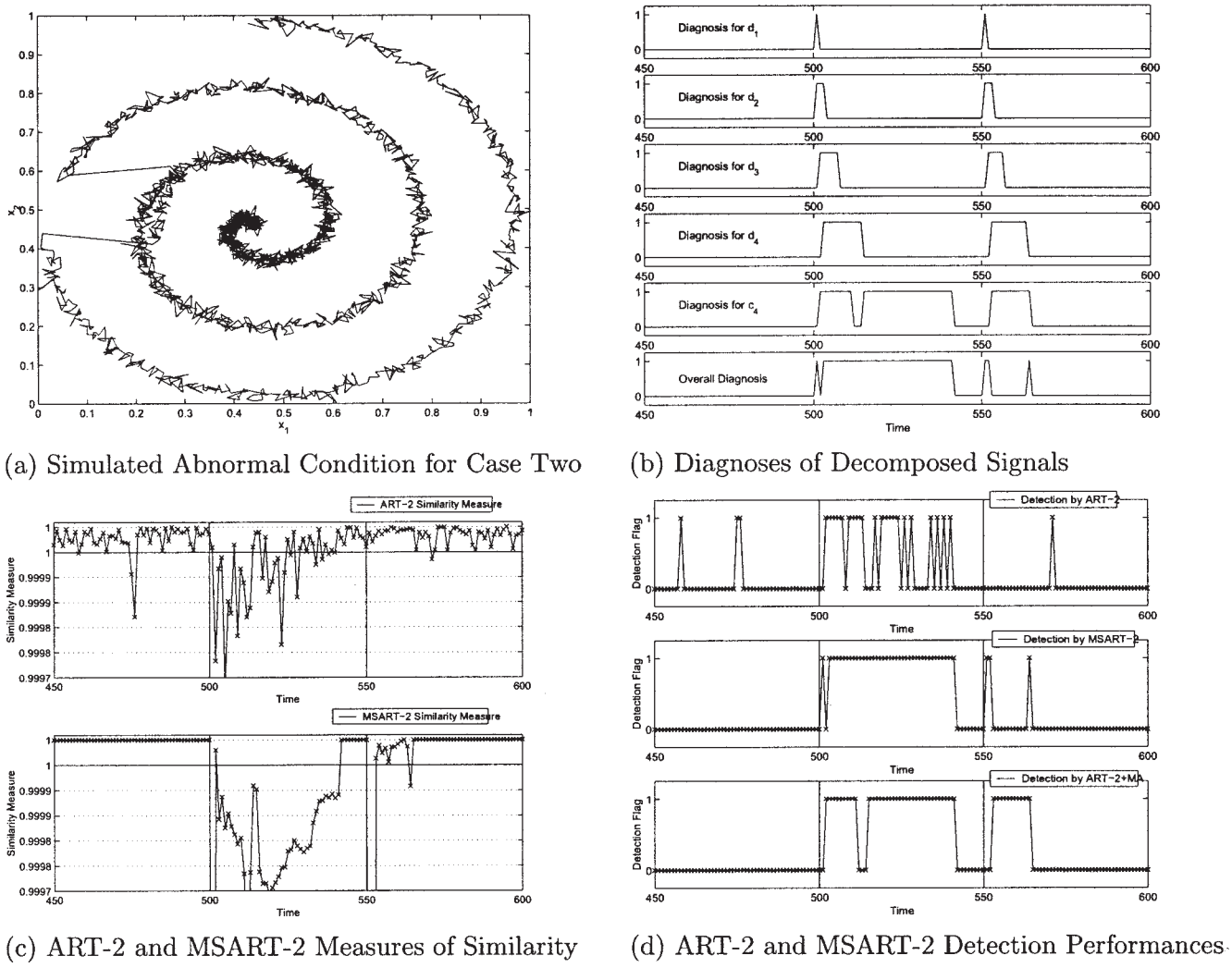
Figure 3. Comparative performance for test case one.

tap moving average filter, and, hence, it is less sensitive to noise than the original time-domain signal. However, it continued to report the fault for roughly up to 16 time-steps after the fault was over (Figure 3b: fifth plot from the top).

Reconstruction and Overall Diagnosis. Based on the outcomes of the Scale Selection Layer networks, a reconstructed signal was appropriately generated. For example, at time-step 551 in Figure 3b, only the networks $ART\mathbf{D}_1$ and $ART\mathbf{A}_4$ reported a non-normal operation. Hence, the reconstructed signal at time-step 551 was generated by applying the inverse wavelet transform with all other coefficients, except d_1 and a_4, replaced by zeros. Similarly, at time-step 556, the reconstructed signal was generated by applying the inverse wavelet transform after retaining only the coefficients d_3, d_4, and a_4, and replacing all other coefficients by zeros. This reconstructed signal was then subjected to an ART-2 network from the diagnosis layer. At each time-step, of the 32 diagnosis layer networks, the network trained on data generated by carrying the same reconstruction on normal data was chosen. The detection flag of the diagnosis layer network chosen at each time-step are plotted against time in the bottom-most graph.

The last scaled signal (a_4) was the only coefficient selected to reconstruct the signal for time-steps 509 through 550 (Figure 3b), because only $ART\mathbf{A}_4$ detected the fault in this time interval. The reconstructed signal was thus a scalar multiple of a_4, implying a consistent detection of sustained faults and less false alarms. Beyond time-step 550, however, multiple scales were selected for reconstruction. Since the transition at time-step 550 was negative, that is, from a positive shift to no shift, the scales d_1 through d_4 tend to neutralize the continuing positive deviation of the residual a_4 due to filter lag. As a result, the diagnosis based on the reconstructed signal (bottom-most graph) did not lead to as many false alarms following the resumption of a normal state as the diagnosis based solely on a_4 (4^{th} graph from the top). The reconstruction operation is, thus, crucial for avoiding false flags at the end of the abnormal operation, and at the same time maintain consistent detection of sustained shifts.

Analysis. The utility of the MSART-2 approach over ART-2 with or without moving average filtering is seen from Figures 3c and 3d. Figure 3c shows the similarity measures, and the associated vigilance parameters, for the current test



(a) Simulated Abnormal Condition for Case Two

(b) Diagnoses of Decomposed Signals

(c) ART-2 and MSART-2 Measures of Similarity

(d) ART-2 and MSART-2 Detection Performances

Figure 4. Comparative performance for test case two.

data using ART-2 and MSART-2. A similarity measure below the vigilance parameter (solid horizontal line) indicates an "abnormal" diagnosis. While both ART-2 and MSART-2 detected the fault for its entire duration, the multiscale approach managed to achieve a larger separation between the normal and faulty behavior without as many false alarms (Figure 3c).

Figure 3d shows the detection flags of three-fault detectors: an ART-2 detector, an MSART-2 detector with a 16 tap wavelet filter, and an ART-2 detector that uses a 16 tap moving average filter for noise removal. The ART-2 detection performance (top-most graph) reiterates the fact that ART-2 based diagnosis without any preprocessing is prone to noise and, hence, false alarms. MA smoothening filter achieves reduction in noise, and, hence, reduction in false alarms, during continued normal operation (bottom-most graph). However, it did not detect the fault immediately (time-step 501), and it lead to a set of false alarms immediately following the malfunction (time-step 550). The MSART-2 approach (middle graph) was successful in reducing both of these disadvantages by focusing on only the smoothened (a_4) component of the signal during sustained shift, and a combination of relevant scales during the transitional phases.

Case Two: A low-noise process with a narrowly separable shift

We now present a case where the faulty data were narrowly separated from the normal data by changing the magnitude of the shift (Figure 4a). The shift lasted for time-steps 501 through 550, similar to the earlier case. Toward the end, the shifted data completely overlapped with the other arm of the spiral, known to be normal.

Figures 4b through 4d illustrate the performance of MSART-2 relative to that of ART-2 with or without MA filtering. The individual outcomes of the scale selection layer networks were similar to Case One. Toward the end of faulty operation (time-steps 545 through 550), due to complete overlap of shifted data, and another arm of the normal spiral, none of the scale selection layer networks detected the fault. The sudden shift back to normal, however, was detected clearly (time-steps 551 and 552).

Figure 4c shows the similarity measures for abnormal operation for ART-2 and MSART-2. When compared to Case One (Figure 3c, top graph), ART-2 can be observed to achieve considerably less separation between the normal, and the ab-

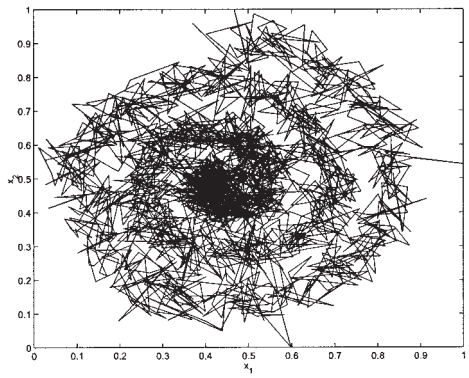


Figure 5. Variable correlations under normal conditions for case three.

normal operations in terms of the similarity measure (Figure 4c, top graph). Similar reduction in the extent of separation is seen with MSART-2 as well (Figures 3c and 4c, bottom graphs), although MSART-2 continued to outperform ART-2. The similarity measure for MSART-2 remained well below the vigilance for most parts. Toward the end of the abnormal operation, close match of the test and normal data affected the similarity measure.

As can be seen from the diagnoses reported in Figure 4d, ART-2 did not detect the fault consistently because of the smaller distinction between normal and abnormal data with respect to the extent of normal noise. Use of the MA filter alleviated the chattering and also reduced the number of false alarms during sustained normal behavior. This added advantage, however, came at the cost of delay in detecting the resumption of normal operation at time-step 551. The MSART-2 approach, similar to Case One, successfully managed to reduce the chattering as well as the inaccurate classification at the transitional regions. For Case Two, MSART-2 can thus be seen to provide quicker and more consistent detection than both ART-2 and ART-2+MA in spite of the narrow separation between abnormal and normal process operation.

Case Three: A high-noise process

In this section, we analyze the effect of multiscale architecture on anomaly detection in the presence of large extent of noise. The training and testing data used for this case are provided in Figures 5 and 6a. Similar to Case One, a shift was simulated from time-step 501 through 550, although it is difficult to visually detect the shift because of the presence of more noise.

In contrast with the earlier cases, the high-noise in this case hampered the detection of the transient phases in finer scales. Analysis of the decomposed scales (Figure 6b) shows that the two finest ART-2 detectors (d_1 and d_2) did not detect the shift at all, unlike the earlier cases. The overall diagnosis was based on only the coarsest scaled signal a_4, for most parts. Scales d_3 and d_4 detected the transition back to normal with the expected delay. This selection of multiple scales for reconstruction reduced, to a small extent, the lagged alarm at the resumption of normal operation. Figures 6c and 6d compare ART-2 and MSART-2 detectors for this test case. The similarity measure plot (Figure 6c, top graph) shows that ART-2 was unable to separate the normal and abnormal process operation. Thus, the ART-2 detector led to many missed and false alarms (Figure 6d, top graph). The detection flag for MSART-2 closely resembled that for ART-2+MA. Both detectors detected the transition away from normal, and resumption of normal operation at a lag approximately equal to the width of the filter used (16 in this case).

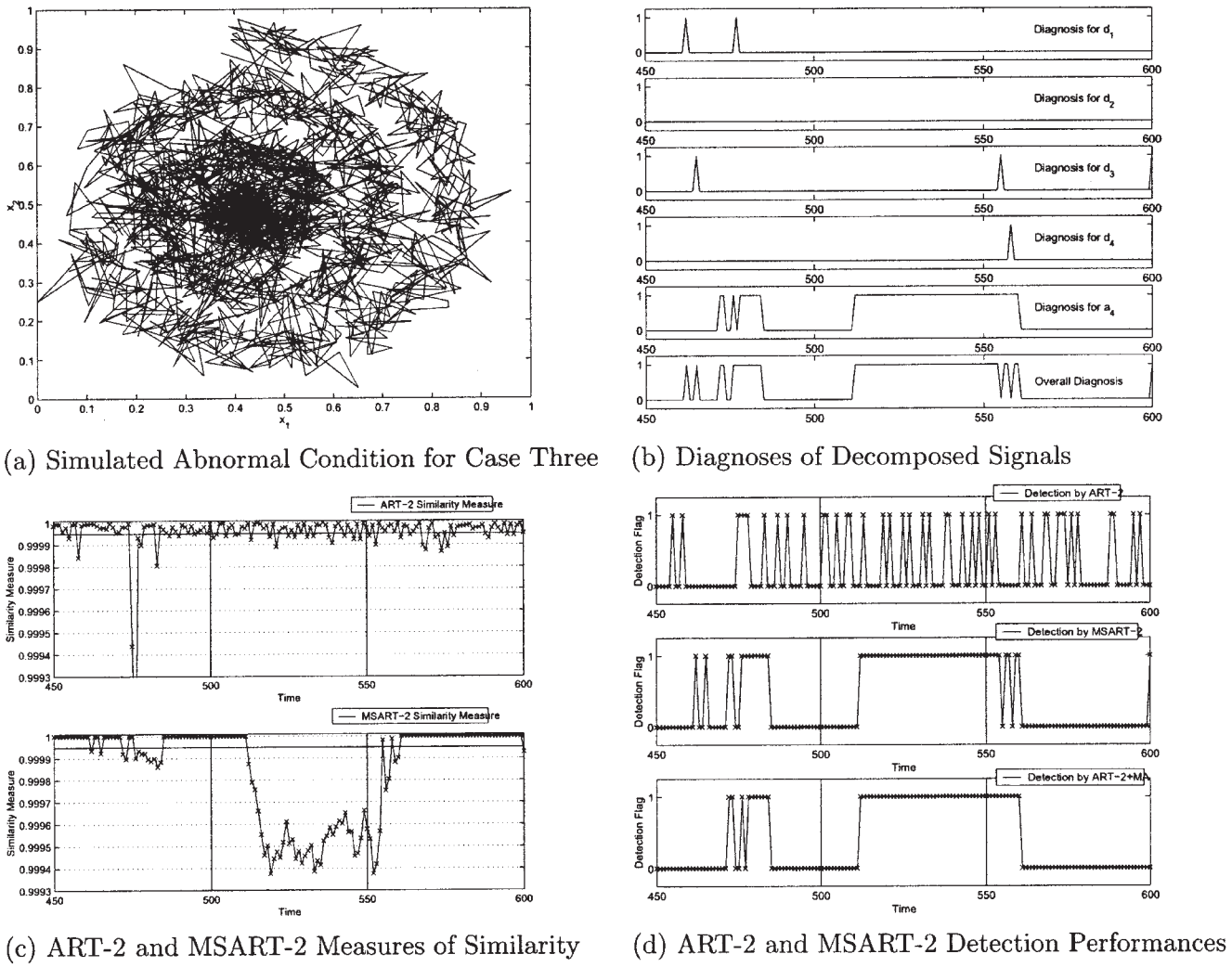
Because of the presence of high-noise for Case Three, it is not surprising that the ART-2+MA approach worked better than ART-2. The close resemblance of MSART-2 and ART-2+MA for this test case attests to our claim that the multiscale detection approach conforms to the best scale for the fault at hand.

These three representative cases illustrate that the multiscale approach is a generic approach that works well on various different changes. On the other hand, single-scale methods, such as ART-2 with and without moving average filter work, best only for specific situations. For example, the unfiltered ART-2-based approach works best only for low-noise mappings (or large shifts), with clearly separated normal and abnormal modes of operation. Similarly, the moving average based approach works better for very noisy mappings (or small shifts).

Average Run-Length Performance Analysis

Having presented illustrations that bring out the strengths of the proposed MSART-2 architecture, we now provide a statistically sound comparative performance analysis via Monte Carlo simulations on three types of processes. For each of the problems discussed later, mean shifts of varying magnitudes were superimposed on the normal data. The fault detection technique under investigation was then applied. The number of time-steps taken before the fault was detected for the first time, referred to as run-length, is noted for each algorithm. Run-lengths may vary in different instances of the random process for the same shift size, and the same detection mechanism. The average run-length (ARL), computed across multiple instances of the random process, was tabulated against each magnitude of mean shift for each detection algorithm. When the magnitude of shift is zero, the corresponding ARL value is indicative of the false alarm rate of the detection technique, and is referred to as the in-control run-length. For the same in-control run-length, it is desirable to have the lowest possible ARL values for nonzero mean shifts. This mechanism provides a standard way of comparing the relative performance of different monitoring techniques (Montgomery, 1996). When plotted against the magnitude of the shift, the ARL curve is expected to be nonincreasing, and typically converges to 1 as the magnitude of shift tends to infinity.

In the experiments presented in the following sections, the vigilance for ART-2 was varied for a fixed vigilance parameter of MSART-2, until the in-control run-lengths matched. We can then compare the MSART-2 and ART-2 detection performance, while keeping the average false alarm rate equal for both detectors, in a Monte Carlo fashion. Since repeated experimentation is required to calculate ARLs, the MSART-2 detector was limited to the minimum level of wavelet decomposition (that is, $L = 1$) to reduce computational time. For higher levels of wavelet decomposition, the difference between



(a) Simulated Abnormal Condition for Case Three

(b) Diagnoses of Decomposed Signals

(c) ART-2 and MSART-2 Measures of Similarity

(d) ART-2 and MSART-2 Detection Performances

Figure 6. Comparative performance for case three.

ART-2 and MSART-2 performances will be even more significant.

A univariate process

In this section, we consider the following simple univariate process model

$$x(t) = N(0, 1) \tag{1}$$

where $N(0, 1)$ is the output of an IID Gaussian random number generator with zero mean and unit variance, and $x(t)$ is the process under measurement. Process data were normalized so as to lie between the range 0 to 1 as required by ART-2. A data set of 1,000 samples was generated for this process, and used for training the ART-2 and MSART-2 detectors.

To generate the ARL curves, shifts of varying magnitudes were introduced at $t = 0$. For subsequent time-steps, simulated abnormal data were subjected to diagnosis by the algorithm under investigation (ART-2 or MSART-2), and time-step at which the shift was first detected (run-length) was recorded for each magnitude of shift for both detection algorithms. This process was repeated for 1,000 instances of the random process, and the run-lengths were averaged for each shift across these 1,000 simulations.

The ARL curves for ART-2 and MSART2 are provided as Figure 7. We can see that for a wide range of shift magnitudes, MSART-2 detects the shift with smaller average run-lengths. For small shifts, the process noise hampers the ability of ART-2 to consistently detect the shift. Thus, the multiscale architecture successfully improves on detection abilities of ART-2 without introducing significant delay. For large shifts, however, ART-2 is seen to perform slightly better as the shift is easily separable from the inherent noise in the mapping.

A multivariate, linearly correlated process

Consider the following linear multivariate process

$$x_1(t) = N(0, 1) \tag{2}$$

$$x_2(t) = N(0, 1) \tag{3}$$

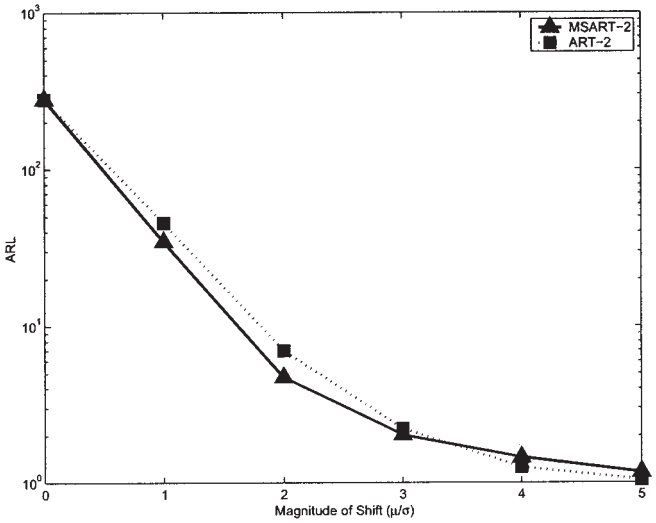


Figure 7. Comparison of ART-2 and MSART-2 performances based on ARL for a univariate process.

$$x_3(t) = \frac{x_1(t) + x_2(t)}{\sqrt{2}} \tag{4}$$

$$x_4(t) = \frac{x_1(t) - x_2(t)}{\sqrt{2}} \tag{5}$$

$$y_i(t) = x_i(t) + \varepsilon_i(t) \tag{6}$$

where $x_i(t)$, $i = 1, \ldots, 4$ are linearly correlated process variables under measurement. Simulated IID Gaussian noise $\varepsilon_i(t)$, of mean zero and standard deviation of 0.2, was superimposed on each variable to generate the measurements $y_i(t)$. Process data were normalized so as to lie between the range 0 to 1. Similar to the univariate process, a data-set of 1,000 measurement vectors was generated, and used for training the ART-2 and MSART-2 detectors.

Shifts were introduced to $y_i(t)$ at $t = 0$, with the magnitudes varying as multiples of the standard deviation of $\varepsilon_i(t)$. The linear correlation across the process variables is, thus, violated. In a manner similar to the univariate process above, ARL curves were generated, and are presented in Figure 8. Again, we observe that MSART-2 outperformed ART-2 for a wide range of shifts. Shifts of a given magnitude are applied across all process variables, and, hence, shifts are detected earlier (lower run-lengths) when compared to the univariate process (Figure 7). We observe that MSART-2 performs better than ART-2, except for large shifts when abnormal operation is well-separated from normal operation.

A multivariate, nonlinearly correlated process

We now present the ARL results for a nonlinear spiral process similar to the one used for section titled Illustration of the MSART-2 Algorithm

$$r(t) = r(t-1) - 0.001 \tag{7}$$

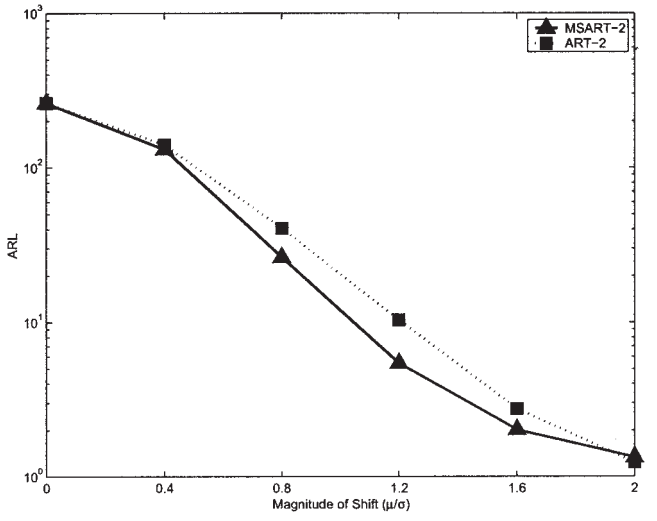


Figure 8. Comparison of ART-2 and MSART-2 performances based on ARL for a linear multivariate process.

$$\theta(t) = \theta(t-1) + 2 * \pi * 0.006 \tag{8}$$

$$x_1(t) = r(t) * \cos(\theta(t)) \tag{9}$$

$$x_2(t) = r(t) * \sin(\theta(t)) \tag{10}$$

$$y_i(t) = x_i(t) + \varepsilon_i(t) \tag{11}$$

The ARL results presented in Figure 9 show that, similar to the earlier results, the multiscale architecture is observed to improve the detection performance of ART-2 in noisy mappings (small shifts). Since the ARL curves are generated by averaging the run-lengths over a 1000 simulations, these results validate the illustrations provided in the section titled Average Run-Length Performance Analysis. When compared to Figure

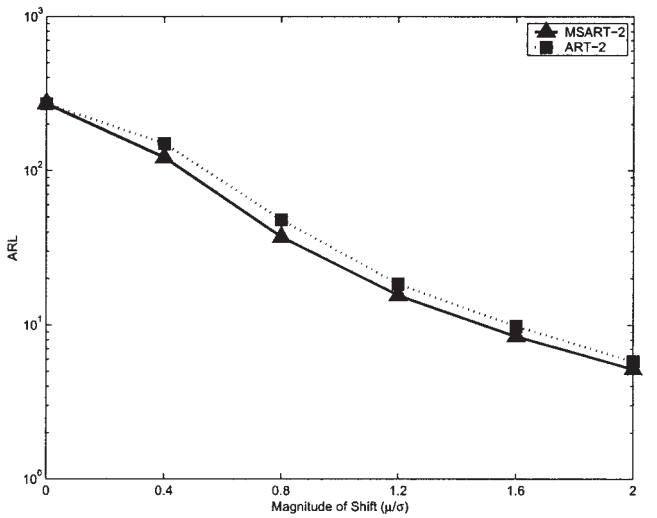
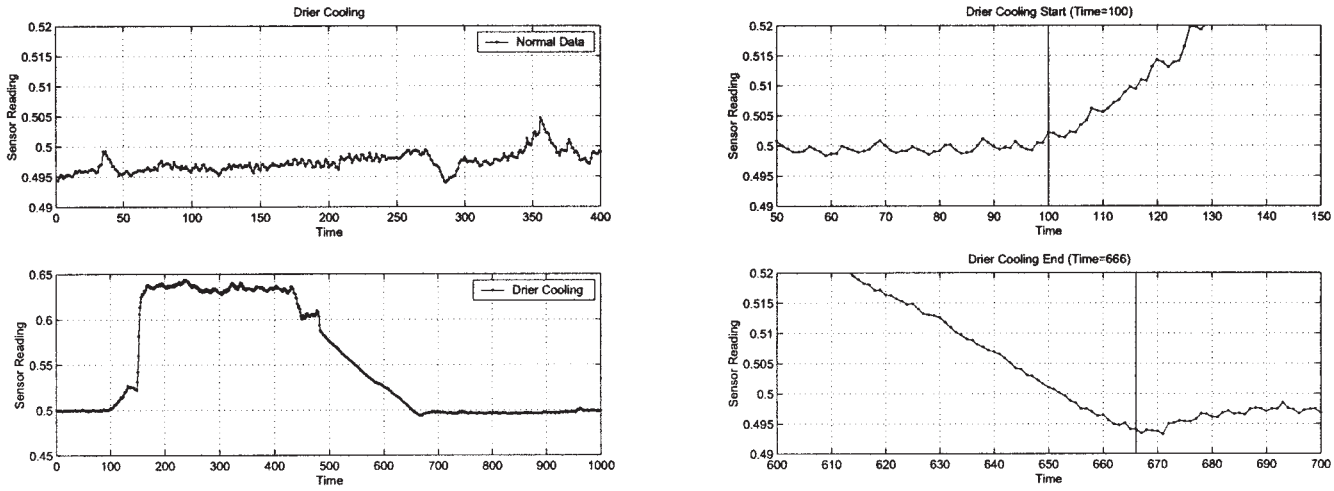


Figure 9. Comparison of ART-2 and MSART-2 performances based on ARL for a nonlinear multivariate process.



(a) The Drier Cooling Event (b) Detail of the Onset and End

Figure 10. Training and test data for the drier cooling event.

8, the reduced difference between the ARL curves can be attributed to the lower number of variables as well as the nonlinear nature of the process.

The ARL curves presented in this section confirm the utility of MSART-2 over ART-2 as established in the section titled Average Run-Length Performance Analysis. By exploiting wavelet-domain clusters, we see that MSART-2 can detect small shifts with smaller detection delays when compared with ART-2.

Industrial Case Studies

In this section, we present two univariate examples taken from sensor readings of a real large-scale petrochemical process. As claimed earlier, deviations from normality in real processes can be slow or fast. In addition, they may differ in the extent of noise, and random and/or deterministic nature of the change. We have chosen two representative process changes that exhibit these different characteristics. For each example case, ART-2, MSART-2, and ART-2+MA were trained with the same training data and same training parameters. Simulated to the illustration from the section titled Illustration of the MSART-2 Algorithm, the objective is to detect the deviations away from normality as soon as possible with the minimum number of both missed and false alarms. The results presented later support our claim that MSART-2 automatically conforms to the nature of the event at hand, and, hence, performs well as a general detection mechanism. A few more industrial case studies, including examples of multivariate systems, that investigate the utility of our framework can be found in (Aradhye et al., 2002).

Example 1: Drier cooling

Drier cooling is a typical "unusual" pattern in petrochemical processes where the coolant flow rate increases beyond the range of normal operation in response to the overheated unit. Figures 10a and 10b show the process data under normal and drier cooling conditions. With respect to the magnitude of the event, normal data are seen to be nearly of a constant mean.

Since the overall magnitude of the change is large compared to the extent of noise in the process, all three methods under investigation were expected to perform identically except for the transition phases. The onset of the deviation, as well as the return to normality, can be seen to be slow developing (relative to the window of 16 time-steps used in the MA and wavelet filters) and deterministic trends. Hence, the delay introduced by the MA filter was not significant with respect to the pace of change in the process signal. ART-2+MA was expected to have better detection accuracy, in the transient regions, than ART-2 because noise reduction capabilities of the MA filter outweighed the delay it introduced. This observation is reflected in the results presented in Figure 11.

The test data from Figure 10b were subjected to ART-2, MSART-2, and ART-2+MA detectors trained on the normal data from Figure 10a. After the onset of the event at time-step 100, all three methods detected the event at approximately

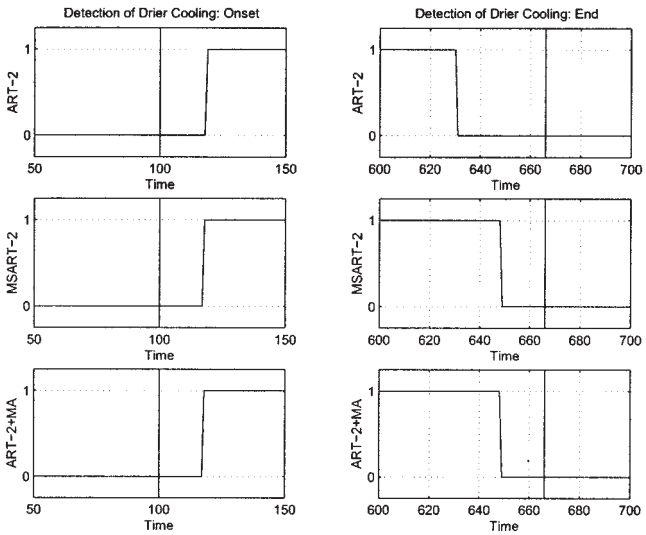


Figure 11. Comparative detection performance for the drier cooling event

Table 1. Industrial Case Studies: Summary

Event Onset	Type	True Time	ART2	MSART2	ART2+MA
Drier Cooling	slow, deterministic	100	120	119	119
Oil Accumulation	slow, stochastic	720	730	730	768
Event End	**Type**	**True Time**	**ART2**	**MSART2**	**ART2+MA**
Drier Cooling	slow, deterministic	666	630	648	648
Oil Accumulation	fast, deterministic	826	825	825	840

equal times (Table 1, row 1), and continued to detect it consistently. Toward the end of the event, however, the ART-2 detector missed about 18 genuine alarms more than the ART-2+MA and MSART-2 detectors (Table 1, row 3). The MA filter reduced the noise in the data. On the other hand, due to the slow pace of the onset and end, the filter did not cause a significant lag. These factors contributed to the better performance of ART-2+MA over ART-2. It can be seen that MSART-2 performance was equivalent to that of ART-2+MA, because it automatically selected the low-resolution scales for this slow-paced deterministic event. MSART-2, thus, successfully adapted to the slow, deterministic nature of the change.

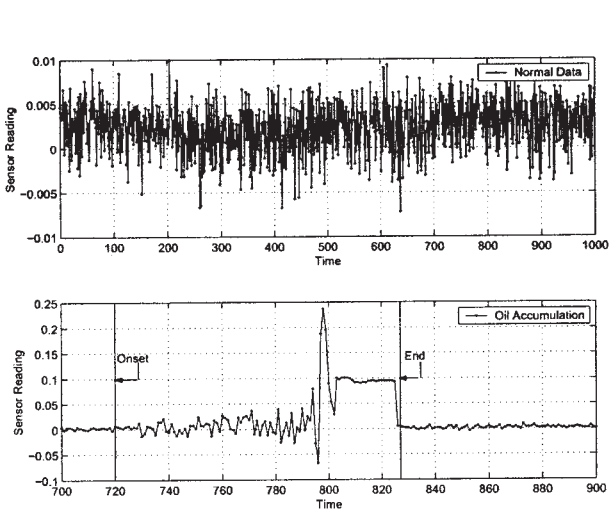
Example 2: Sensor malfunction due to oil accumulation

Redundant sensors are often used for critical measurements for the detection of faulty sensor operation. Failure of any one of the redundant sensors is typically diagnosed by increased magnitude of the difference between the sensor readings. In this example, we present an occurrence of sensor failure due to oil accumulation. The difference between a faulty sensor and its coupled redundant sensor is shown in Figure 12a. Under normal conditions, the difference was seen to be random and nearly zero mean (top graph). In the neighborhood of time-step 720 in the test data (bottom graph), oil began to accumulate in the actuator of one of the sensors, causing it to report erroneous readings. This error can be seen to have a nearly zero-mean, stochastic component in the beginning, and a strong deterministic component after time-step 805. The event ended with a sudden return to normality when the cause of the sensor failure was eliminated by a human operator (time-step 826).

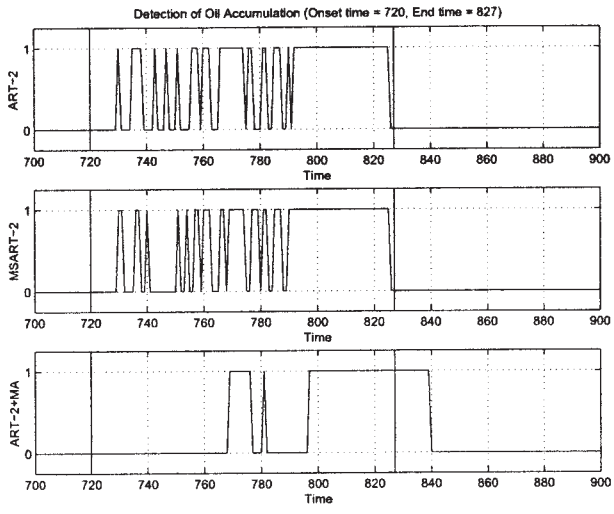
Since the MA filter was set to calculate the average over a window of 16 consecutive time-steps, we expected the ART-2+MA to be ineffective in detecting the initial zero-mean stochastic part of the failure pattern (Table 1, row 2). Also, in this case, the return to normality was a sudden, sharp change of large magnitude. Due to the change in question taking place over a time-span much smaller than the averaging window, we expected the ART-2+MA to result in a large number of false alarms immediately after the end of the sensor failure (Table 1, row 4). Indeed, we find that ART-2 detector resulted in a smaller number of false flags *and* a smaller number of missed flags for this event, when compared to the ART-2+MA detector (Figure 12b, Table 1). Similar to Figure 11, we observe that MSART-2 conforms to the scale of the change under consideration and mimics the best performance for the event at hand.

Conclusion

Previous work by the authors established ART-2 as a mechanism for efficiently and adaptively capturing linear and nonlinear mappings between process variables for the purpose of fault diagnosis and sensor trend analysis. The multiscale architecture proposed in this work was shown to significantly enhance the range of applicability of the ART-2-based diagnosis algorithm. Process malfunctions naturally occur across multiple scales. Single scale approaches, which can be shown to be special cases of the proposed scheme, are often limited to



(a) The Oil Accumulation Event



(b) Comparative Detection Performance

Figure 12. Detection of oil accumulation (sensor failure).

specific types of faulty operation depending on their scales. For example, ART-2 without any filtering, a finest scale detector, is best for detection only in cases where the shifts are large, the changes are sudden, or the event is stochastic. Similarly, ART-2 with moving average filtering, a coarsest scale detector, is best for small shifts, gradual changes, and deterministic events. Our approach integrates scale selection and clustering-based diagnosis. The results presented in this paper show that MSART-2 is a general detection algorithm that chooses the scale most appropriate for the malfunction at hand, and, hence, it delivers a good performance for noisy events with a wide range of shift magnitudes and paces.

Acknowledgments

The authors gratefully acknowledge the support of the Abnormal Situation Management Consortium, the National Science Foundation (NSF Grant CTS-9733627), and the Army Research Laboratory for this research.

Literature Cited

Aradhye, H. B., B. R. Bakshi, R. Strauss, and J. F. Davis, "Multiscale Statistical Process Control Using Wavelets—Theoretical Analysis and Properties," *AIChE J.,* **49,** 939 (2003).

Aradhye, H. B., J. F. Davis, B. R. Bakshi, and S. C. Ahalt, "Art-2 and Multi-scale Art-2 for On-Line Process Fault Detection—Validation via Industrial Case Studies and Monte Carlo Simulation," *Annual Reviews in Control,* **26,** 113 (2002).

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

Bakshi, B. R., H. Aradhye, and R. Strauss, *Process Monitoring by PCA, Dynamic PCA, and Multiscale PCA—Theoretical Analysis and Disturbance Detection in the Tennessee Eastman Process,* AIChE Annual Meeting, Dallas, TX (1999).

Bakshi, B. R., and G. Stephanopoulos, "Representation of Process Trends—iv. Induction of Real-Time Patterns from Operating Data for Diagnosis and Supervisory Control," *Comp. and Chem. Eng.,* **18,** 303 (1994).

Carpenter, G. A., and S. Grossberg, "A Massively Parallel Architecture for a Self-Organizing Neural Pattern Recognition Machine," *Comp. Vision, Graphics and Image Proc.,* **37,** 54 (1987).

Carpenter, G. A., S. Grossberg, N. Markuzon, J. H. Reynolds, and D. B. Rosen, "Fuzzy ARTMAP: A Neural Network Architecture for Incremental Supervised Learning of Analog Multidimensional Maps," *IEEE Transactions on Neural Networks,* **3,** 698 (1992).

Carpenter, G. A., S. Grossberg, and J. H. Reynolds, "ARTMAP: Supervised Real-Time Learning and Classification of Nonstationary Data by a Self-Organizing Neural Network," *Neural Networks,* **4,** 565 (1991a).

Carpenter, G. A., S. Grossberg, and D. B. Rosen, "Fuzzy ART: An Adaptive Resonance Algorithm for Rapid, Stable Classification of Analog Patterns," *Proc. of the Int. Joint Conf. on Neural Networks,* Seattle, WA, Vol. 2 (1991b).

Carpenter, G. A., S. Grossberg, and D. B. Rosen, "Fuzzy ART: Fast, Stable Learning and Categorization of Analog Patterns by an Adaptive Resonance System," *Neural Networks,* **4,** 759 (1991c).

Frank, T., K. Friedrich, and T. Kuhlen, "Comparative Analysis of Fuzzy ART and ART-2A Network Clustering Performance," *IEEE Transactions on Neural Networks,* **9,** 544 (1998).

Kavuri, S. N., and V. Venkatasubramanian, "Representing Bounded Fault Classes Using Neural Networks with Ellipsoidal Activation Functions," *Comp. and Chem. Eng.,* **17,** 139 (1993).

Lim, C. P., and R. F. Harrison, "Modified Fuzzy ARTMAP Approaches Bayes Optimal Classification Rates: An Empirical Demonstration," *Neural Networks,* **10,** 755 (1997).

Luo, R., M. Misra, S. J. Qin, R. Barton, and D. M. Himmelblau, "Sensor Fault Detection via Multiscale Analysis and Nonparametric Statistical Inference," *Ind. Eng. and Chem. Res.,* **37,** 1024 (1998).

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

Marriott, S., and R. F. Harrison, "A Modified Fuzzy ARTMAP Architecture for the Approximation of Noisy Mappings," *Neural Networks,* **8,** 619 (1995).

Montgomery, D. C., *Introduction to Statistical Quality Control,* Wiley, New York (1996).

Nounou, M. N., and B. R. Bakshi, "Online Multiscale Filtering of Random and Gross Errors Without Process Models," *AIChE J.,* **45,** 1041 (1999).

Palavajjhala, S., R. Motard, and B. Joseph, "Process Identification Using Discrete Wavelet Transforms: Design of Prefilters," *AIChE J.,* **42,** 777 (1996).

Song, X., P. K. Hopke, M. Bruns, D. A. Bossio, and K. M. Scow, "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 (1998).

Srinivasa, N., "Learning and Generalization of Noisy Mappings Using a Modified PROBART Neural Network," *IEEE Transactions on Signal Processing,* **45,** 2533 (1997).

Strang, G., "Wavelets and Dilation Equations: A Brief Introduction," *Soc. for Ind. and Appl. Mathematics,* **31,** 614 (1989).

Wang, X. Z., B. H. Chen, S. H. Yang, and C. McGreavy, "Application of Wavelets and Neural Networks to Diagnostic System Development, 2, an Integrated Framework and its Application," *Comp. and Chem. Eng.,* **23,** 945 (1999).

Whiteley, J. R., and J. F. Davis, "Knowledge-Based Interpretation of Sensor Patterns," *Comp. and Chem. Eng.,* **16,** 329 (1992).

Whiteley, J. R., J. F. Davis, A. Mehrotra, and S. C. Ahalt, "Observations and Problems Applying ART2 for Dynamic Sensor Pattern Interpretation," *IEEE Transactions on Systems, Man and Cybernetics, Part A—Systems on Humans,* **26,** 423 (1996).

Wienke, D., and L. Buydens, "Adaptive Resonance Theory Based Neural Networks—The 'ART' of Real-Time Pattern Recognition in Chemical Process Monitoring?" *Trends in Analytical Chemistry,* **14,** 398 (1995).

Wienke, D., and L. Buydens, "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 (1996).

Wienke, D., W. den Broek, L. Buydens, T. Huth-Frehre, and R. Feldhoff, "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 (1996).

Williamson, J. R., "Gaussian ARTMAP: A Neural Network for Fast Incremental Learning of Noisy Multidimensional Maps," *Neural Networks,* **9,** 881 (1996).

Manuscript received Feb. 7, 2003, and revision received Dec. 29, 2003.