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

Hrishikesh B. Aradhye, Bhavik R. Bakshi, James F. Davis, and Stanley C. Ahalt

Abstract

This paper presents a method for process fault detection based on the integration of multiscale signal representation and scale-specific clustering-based diagnosis. Previous work by the authors 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 its use in conjunction with a non-linear 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, the present 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. We provide illustrative examples as well as Monte Carlo bases for these claims via a comparative performance analysis over several case studies. 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. We also present comparative results on real industrial case studies from a petrochemical process plant. Our 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.

Keywords

Statistical Process Control; Process Monitoring; Adaptive Resonance Theory; Fault Detection; Wavelets; Average Run-Length.

I. Introduction

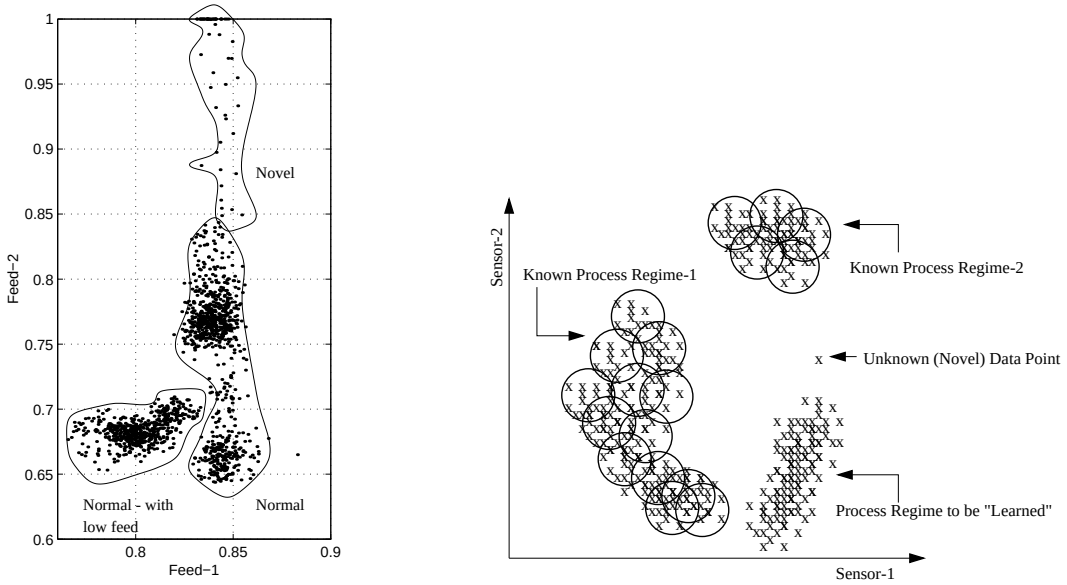
IN an environment where most process maneuvers are automated, algorithms to detect and classify abnormal trends in process measurements are of critical importance from the point of view of safe and economical plant operation. These algorithms use information extracted from previously annotated process 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* 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.

H. B. Aradhye is a student member of IEEE. He is pursuing a graduate education at the Department of Chemical Engineering and the Department of Electrical Engineering, The Ohio State University, Columbus, OH.

B. R. Bakshi is an Associate Professor at the Department of Chemical Engineering, The Ohio State University, Columbus, OH.

J. F. Davis is a member of IEEE. He is a Professor at the Department of Chemical Engineering, University of California at Los Angeles, Los Angeles, CA.

S. C. Ahalt is a member of IEEE. He is a Professor at the Department of Electrical Engineering, The Ohio State University, Columbus, OH.



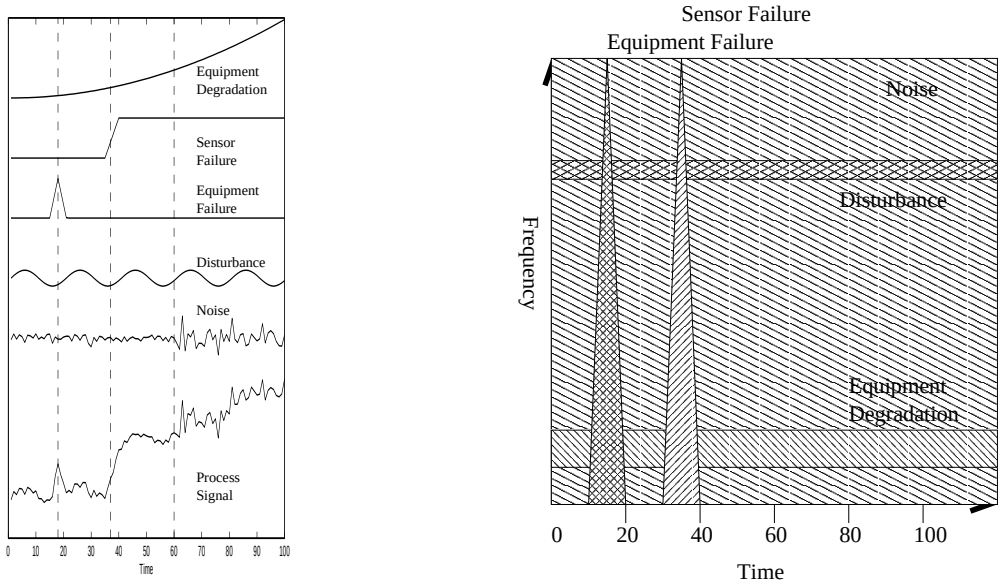
(a) Nonlinear Mapping from a Chemical Process (b) ART-2 for Incremental Diagnosis

Fig. 1. Fault Detection for Industrial Processes

A. Clustering-based Fault Detection

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 (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 [1][2]. Specifically, the ART family of networks [3][4][5][6][7] 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 [8][9][10], Hopke and co-workers [11], as well as Wang and co-workers [12], in addition to the previous work by the authors [13] (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. [14] 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 [15]. Several ART and ARTMAP variants have been proposed in the past to tackle this issue. The PROBART network proposed by Marriot and Harrison [15] 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 [16] was shown to approach Bayes optimal classification rates.



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

Fig. 2. Multiresolution Analysis of a Typical Process Signal

The work by Srinivasa [17] proposed a PROBART variant that improved its generalization ability in the context of high noise. Gaussian ARTMAP by Williamson [18] combined a Gaussian classifier and an ARTMAP network by appropriately changing the definitions of ART *choice* and *match* functions. Recently, Wang and co-workers [12] 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 [19][20]. 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 [21]. 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.

B. Wavelet Decomposition and Change Detection

Wavelets and multiresolution signal analysis [22][23] have triggered developments in a range of process systems engineering related domains such as trend extraction [24], process modeling [25], sensor validation [26], noise reduction [27], 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. [28] proposed a wavelet-domain Hidden Markov Model for univariate statistical signal processing. Swami, Sadler, and co-workers [29][30][31][32][33][34] have presented multiscale methods for step detection and estimation. Other researchers [35][36] have investigated wavelet-based shockwave 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 [19]. 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.

II. Background

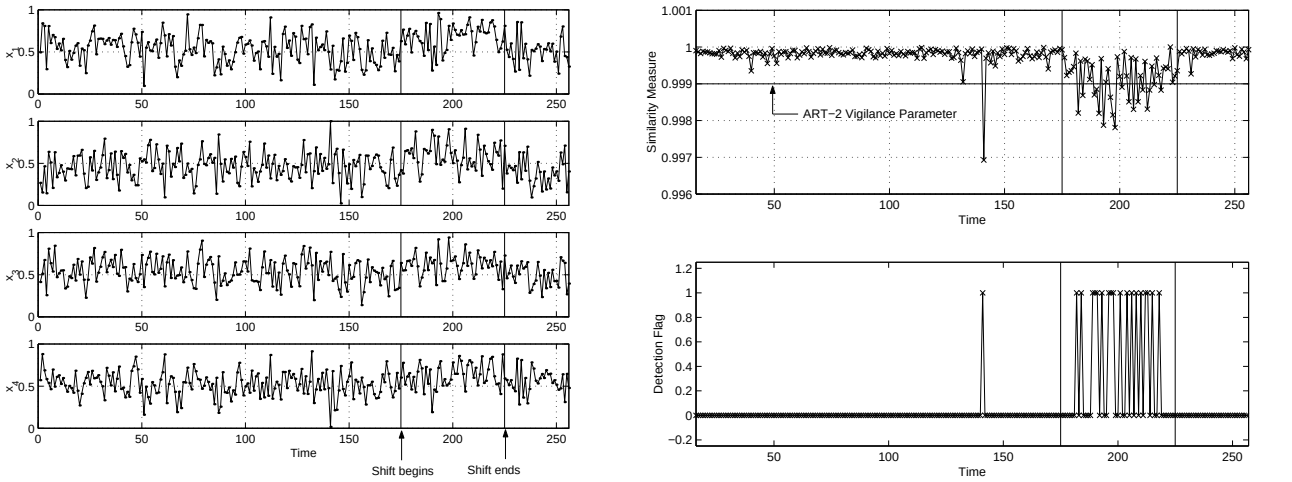
A. Adaptive Resonance Theory

ART-2 is an unsupervised clustering mechanism proposed by Carpenter and Grossberg [3]. 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 [4][5][6][7], 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 (Whiteley and Davis) 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., [2]). It has been shown to be able to work consistently well over a wide range of simulated as well as real-life process situations [13][37][1].

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



(a) A Simulated Noisy Data-set (b) ART-2 Diagnosis Performance in a Noisy Environment

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

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 (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 a lot of 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 (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.

B. Wavelets

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

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

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.

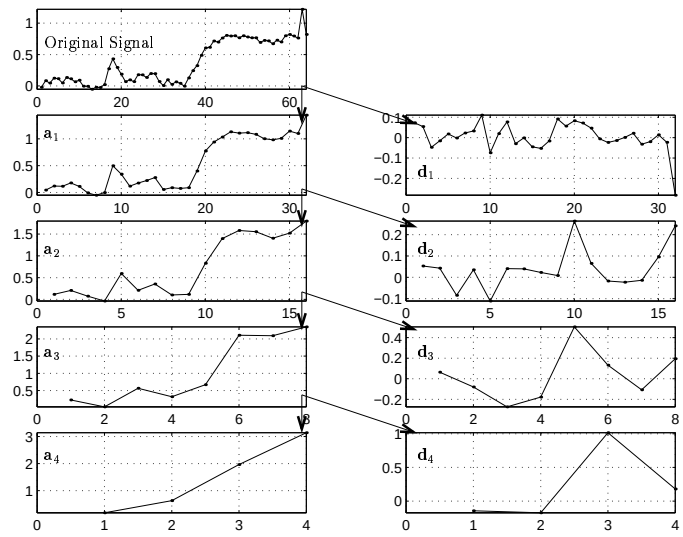


Fig. 4. Representation of Process Signals at Successive Levels of Approximation

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

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 development of truly on-line multiscale methods. Our earlier work [21] 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.

III. Description of the Proposed MSART-2 Algorithm

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 that constitute the MSART-2 architecture. Let P be the number of process

variables in a multi-variate 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.

A. 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 $d^{train}_{m,t,p}$ and the scaled signal coefficients $a^{train}_{L,t,p}$, where $m = 1, \ldots, L$, $t = 2^L, \ldots, N$, and $p = 1, \ldots, P$. The illustration in Figure 5 used a wavelet decomposition with $L = 4$. We then construct $L + 1$ training matrices $\mathbf{D}^{train}_m$, $m = 1, \ldots, L$, and $\mathbf{A}^{train}_L$, 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 5 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 $\hat{\mathbf{Y}}^{train}_1, \hat{\mathbf{Y}}^{train}_2, \ldots, \hat{\mathbf{Y}}^{train}_{2^{L+1}}$, 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 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. Hence the time-domain ART-based detector is a special case of the multiscale hierarchy presented in this work.

B. 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 $\mathbf{y}^{test}_t$ into wavelet coefficients $d^{test}_{m,t,p}$ and $a^{test}_{L,t,p}$. Figure 5 shows a decomposition of a dyadic window of the test signal y^{test}_t into coefficients $d^{test}_{1,t}, \ldots, d^{test}_{4,t}$ and $a^{test}_{4,t}$. For each scale m, construct a P-dimensional vector $\mathbf{d}^{test}_{m,t}$, comprising of coefficients $d^{test}_{m,t,p}$ with $p = 1, \ldots, P$. This vector is presented as input to detector $ART\mathbf{D}_m$ of the

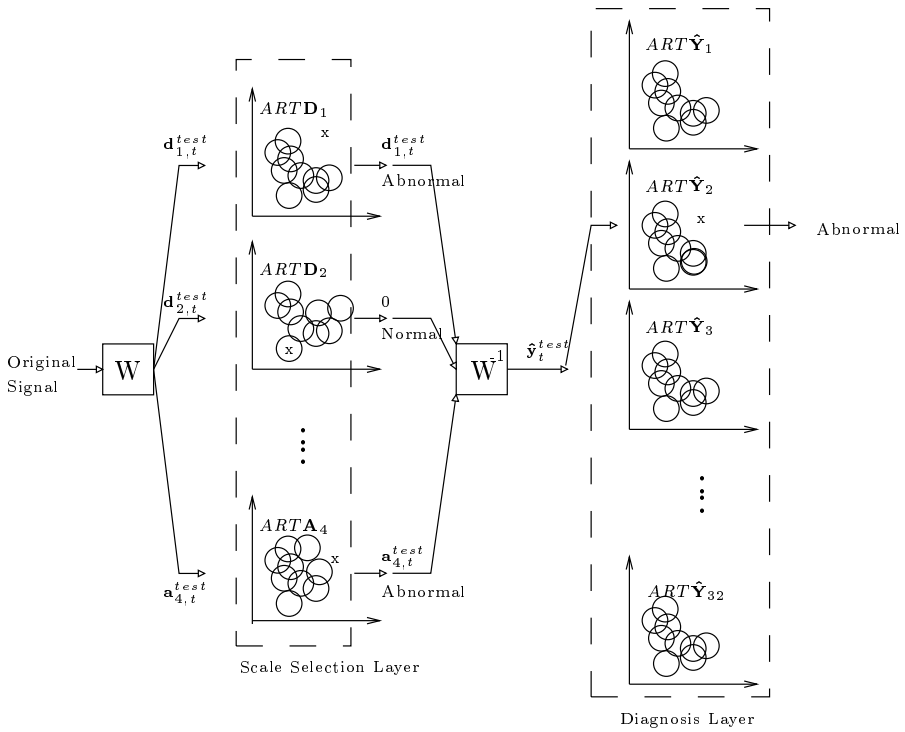


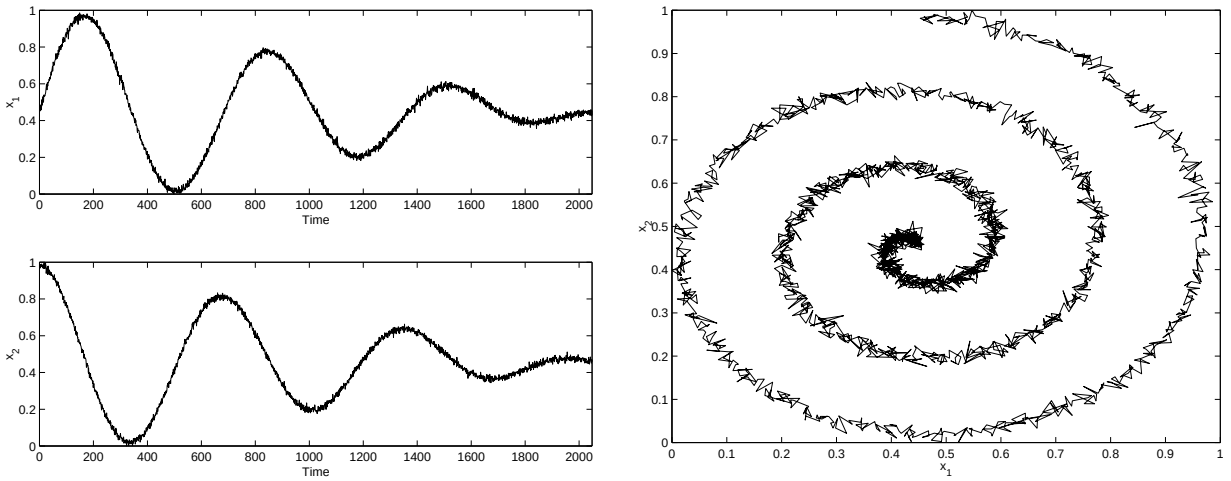
Fig. 5. The MSART-2 Architecture for Robust Fault Diagnosis

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 5, 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 5 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 requirement for MSART-2 is in fact approximately $L + 1 + 2^{L+1}$ times the storage requirement for ART-2.

IV. 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



(a) Normal Data for Cases One and Two (b) Variable Correlations Under Normal Conditions

Fig. 6. A Bi-variate Process for Illustration of the MSART-2 Approach

noise and the extent of separation between normal and abnormal operation.

The experiments discussed in this paper 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 6a shows the normal behavior of the process considered in this illustration. The input vector $\mathbf{x}(t)$ consisted of measurements of two nonlinearly correlated process variables $x_1(t)$ and $x_2(t)$. A bi-variate problem was chosen for visual simplicity, although the algorithms considered are multi-variate. Gaussian noise was superimposed on the data to simulate noisy conditions. Figure 6b illustrates the non-linear 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.

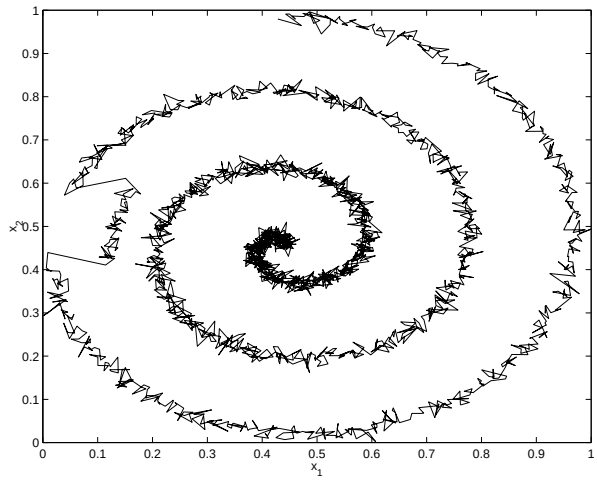
Please note that the noisy mapping in this case is random and the exact diagnoses may differ for different instantiations of the random process. We go on to establish the utility of the current approach in a Monte Carlo fashion in the next section.

A. Case One: A Low-noise Process with a Clearly Separable Shift

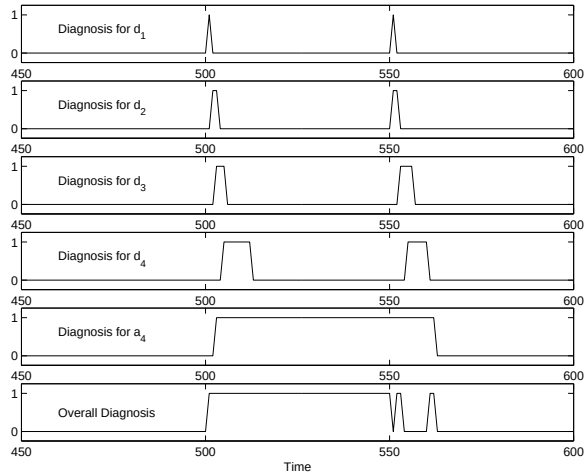
Figure 7a shows the test data used for diagnosis in this section. As can be seen as the left side of the outermost arm of the spiral, a shift was introduced to simulate abnormal behavior from time-step 501 to time-step 550. The number of scales, L, was chosen to be 4.

A.1 Scale Selection

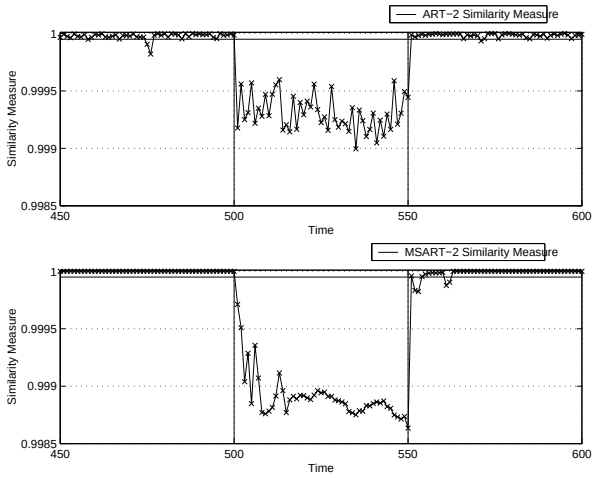
The detection flags of the decomposed signals provide an insight into the mechanism of scale selection in the MSART-2 architecture. Figure 7b shows the diagnoses by the resulting 5 Scale Selection Layer networks for a part of the test signal. As explained earlier, the Scale Selection Layer subjects each 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 7b, a detection



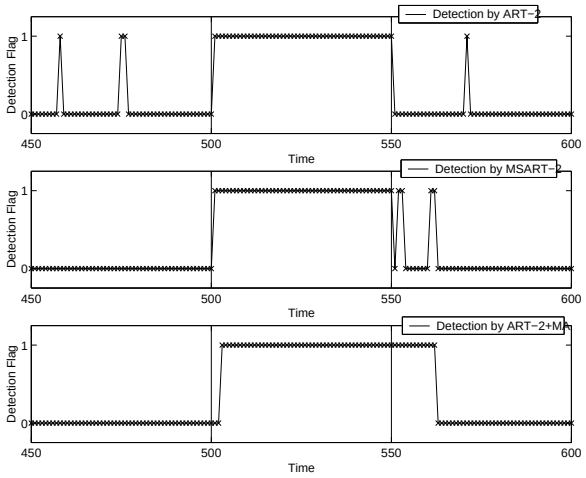
(a) Simulated Abnormal Condition for Case One



(b) Diagnoses of Decomposed Signals



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



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

Fig. 7. Comparative Performance for Test Case One

flag of 0 indicates a "normal" diagnosis, whereas a detection flag of 1 implies an "abnormal" diagnosis. The overall diagnosis, i.e., the diagnosis on the reconstructed signal (Figure 7b: bottom-most graph), illustrates the effect of simultaneous selection of multiple scales. Figure 7b 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 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 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 7b: fifth plot from the top).

A.2 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 7b, 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 7b), 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, i.e. 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-nost 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.

A.3 Analysis

The utility of the MSART-2 approach over ART-2 with or without moving average filtering is seen from Figures 7c and 7d. Figure 7c shows the similarity measures, and the associated vigilance parameters, for the current test 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 7c).

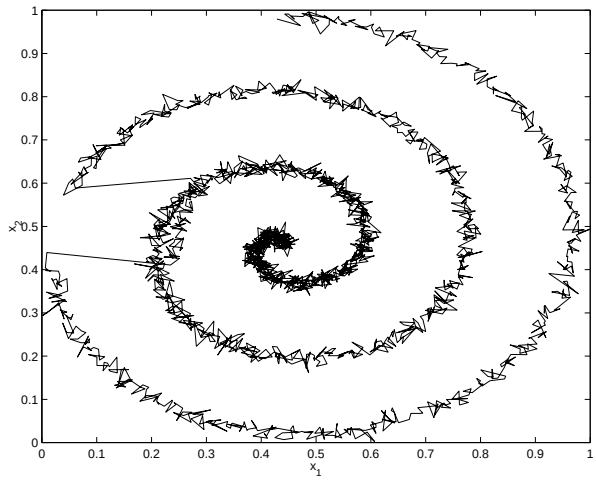
Figure 7d shows the detection flags of three fault detectors: an ART-2 detector, an MSART-2 detector with a 16 tap wavelet filter, and an ART-2 detector that uses a 16 tap moving average filter for noise removal. The ART-2 detection performance (top-most graph) reiterates the fact that ART-2 based diagnosis 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.

B. Case Two: A Low-noise Process with a Narrowly Separable Shift

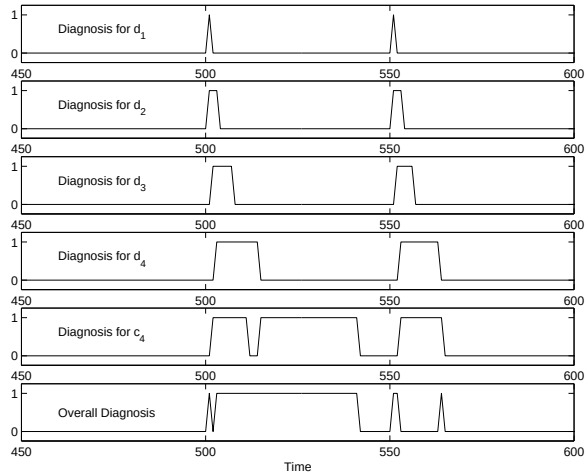
We now present a case where the faulty data were narrowly separated from the normal data by changing the magnitude of the shift (Figure 8a). The shift lasted for time-steps 501 through 550, similar to the earlier case. Towards the end, the shifted data completely overlapped with the other arm of the spiral, known to be normal.

Figures 8b through 8d 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. Towards 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).

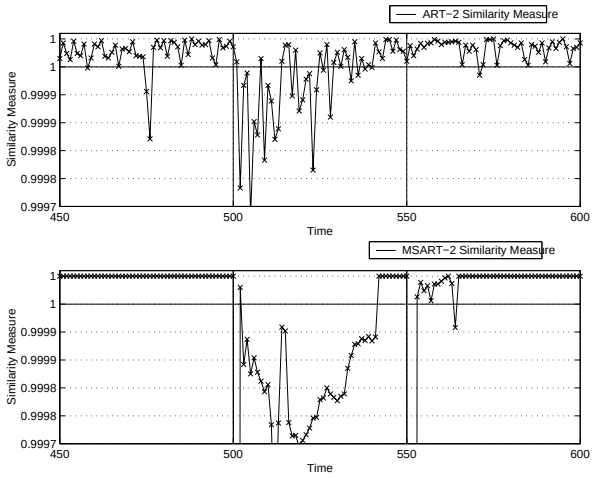
Figure 8c shows the similarity measures for abnormal operation for ART-2 and MSART-2. When



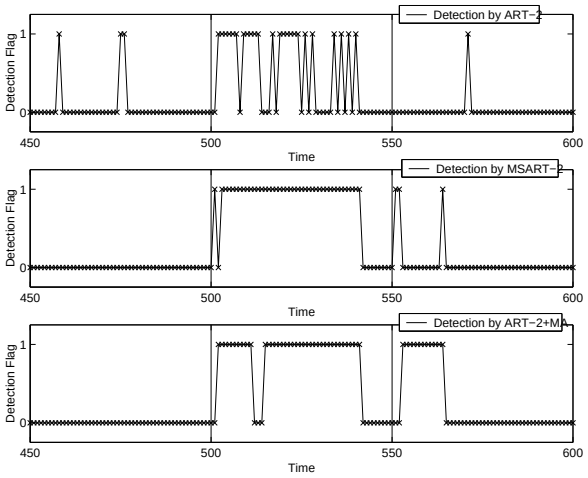
(a) Simulated Abnormal Condition for Case Two



(b) Diagnoses of Decomposed Signals



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



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

Fig. 8. Comparative Performance for Test Case Two

compared to Case One (Figure 7c, top graph), ART-2 can be observed to achieve considerably less separation between the normal and the abnormal operations in terms of the similarity measure (Figure 8c, top graph). Similar reduction in the extent of separation is seen with MSART-2 as well (Figures 7c and 8c, bottom graphs), although MSART-2 continued to outperform ART-2. The similarity measure for MSART-2 remained well below the vigilance for most parts. Towards 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 8d, 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.

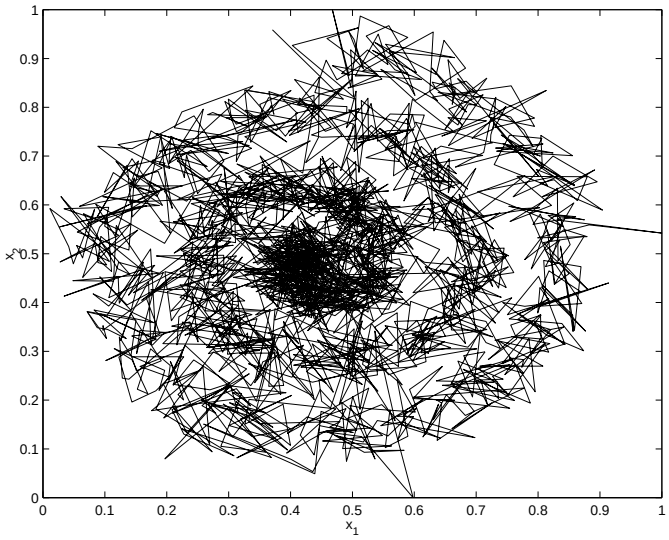


Fig. 9. Variable Correlations Under Normal Conditions for Case Three

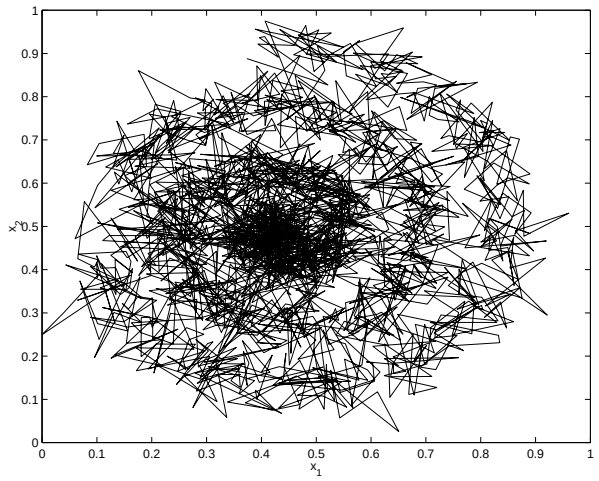
C. Case Three: A High-noise Process

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 9 and 10a. 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.

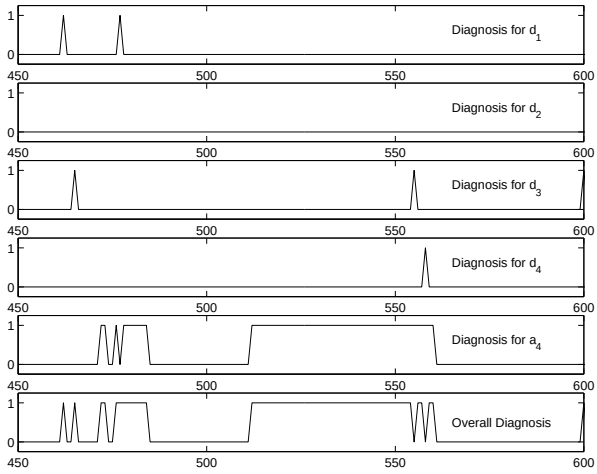
In contrast with the earlier cases, the high noise in this case hampered the detection of the transient phases in finer scales. Analysis of the decomposed scales (Figure 10b) shows that the two finest ART-2 detectors (d_1 and d_2) did not detect the shift at all, unlike the earlier cases. The overall 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 10c and 10d compare ART-2 and MSART-2 detectors for this test case. The similarity measure plot (Figure 10c, 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 10d, 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.

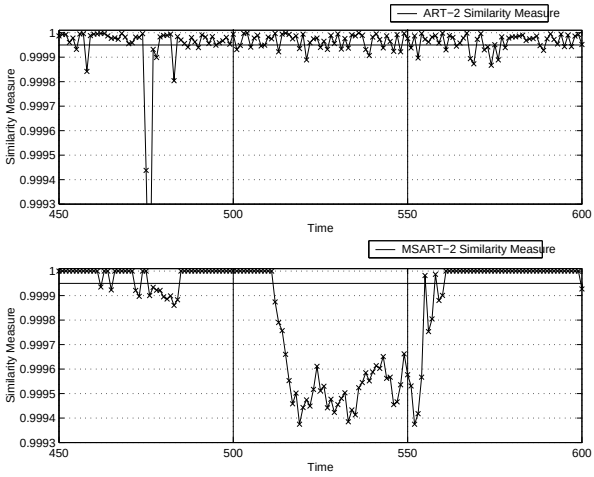
These three representative cases illustrate that the multiscale approach is a generic approach that works well on various different changes. On the other hand, single-scale methods such as ART-2 with and without moving average filter work best only for specific situations. For example, the unfiltered 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).



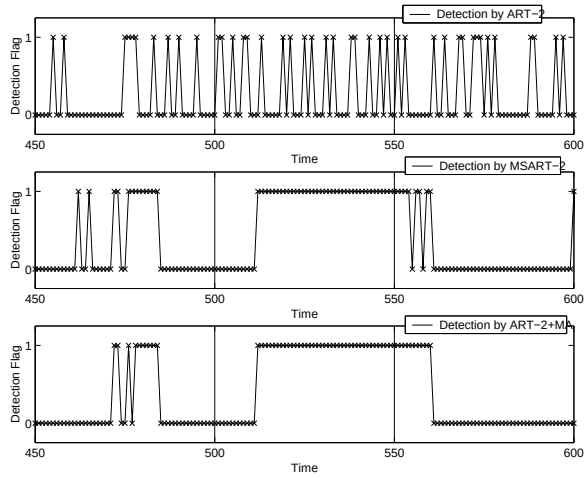
(a) Simulated Abnormal Condition for Case Three



(b) Diagnoses of Decomposed Signals



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



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

Fig. 10. Comparative Performance for Case Three

V. 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 below, 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 non-zero mean shifts. This mechanism provides a standard way of comparing the relative performance of different monitoring techniques [38]. When plotted against the magnitude of the shift, the ARL curve is expected to be non-increasing 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

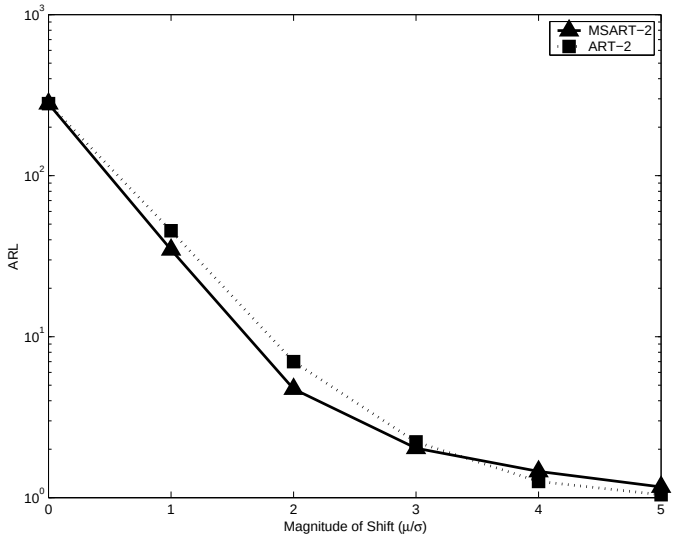


Fig. 11. Comparison of ART-2 and MSART-2 Performances based on ARL for a Univariate Process

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 (i.e., $L = 1$) to reduce computational time. For higher levels of wavelet decomposition, the difference between ART-2 and MSART-2 performances will be even more significant.

A. A Univariate Process

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

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

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 1000 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 1000 instances of the random process and the run-lengths were averaged for each shift across these 1000 simulations.

The ARL curves for ART-2 and MSART2 are provided as Figure 11. 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.

B. A Multivariate, Linearly Correlated Process

Consider the following linear multi-variate process:

$$x_1(t) \quad = \quad N(0, 1) \tag{4}$$

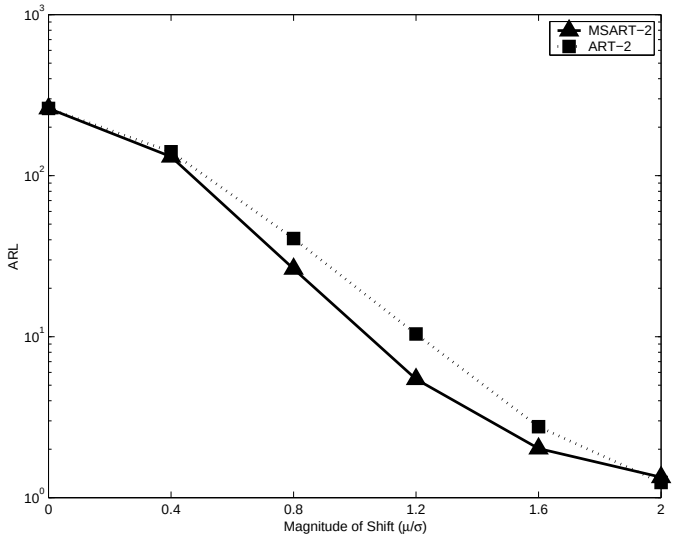


Fig. 12. Comparison of ART-2 and MSART-2 Performances based on ARL for a Linear Multivariate Process

$$x_2(t) = N(0,1) \quad (5)$$

$$x_3(t) = \frac{x_1(t) + x_2(t)}{\sqrt{2}} \quad (6)$$

$$x_4(t) = \frac{x_1(t) - x_2(t)}{\sqrt{2}} \quad (7)$$

$$y_i(t) = x_i(t) + \epsilon_i(t) \quad (8)$$

where $x_i(t), i = 1, \ldots, 4$ are linearly correlated process variables under measurement. Simulated IID Gaussian noise, $\epsilon_i(t)$, 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 1000 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 $\epsilon_i(t)$. 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 12. 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 11). We observe that MSART-2 performs better than ART-2, except for large shifts when abnormal operation is well-separated from normal operation.

C. A Multivariate, Nonlinearly Correlated Process

We now present the ARL results for a non-linear spiral process similar to the one used for Section 3.

$$r(t) = r(t-1) - 0.001 \quad (9)$$

$$\theta(t) = \theta(t-1) + 2 * \pi * 0.006 \quad (10)$$

$$x_1(t) = r(t) * cos(\theta(t)) \quad (11)$$

$$x_2(t) = r(t) * sin(\theta(t)) \quad (12)$$

$$y_i(t) = x_i(t) + \epsilon_i(t) \quad (13)$$

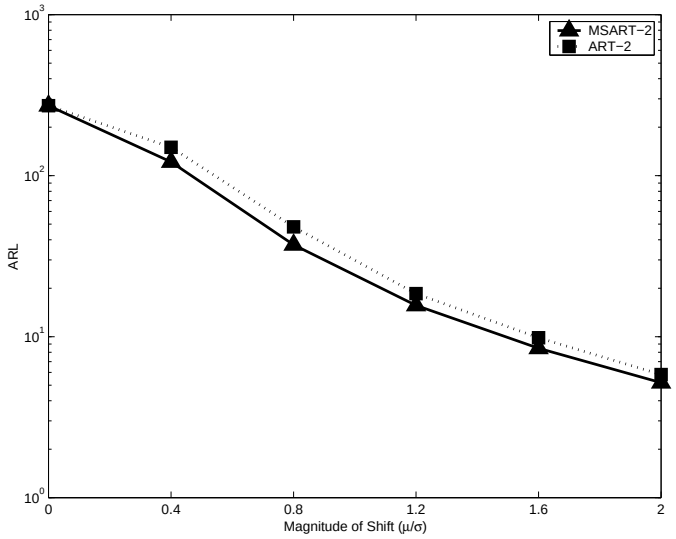


Fig. 13. Comparison of ART-2 and MSART-2 Performances based on ARL for a Nonlinear Multivariate Process

The ARL results presented in Figure 13 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 Section 4. When compared to Figure 12, 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 Section 4. By exploiting wavelet-domain clusters, we see that MSART-2 can detect small shifts with smaller detection delays when compared with ART-2.

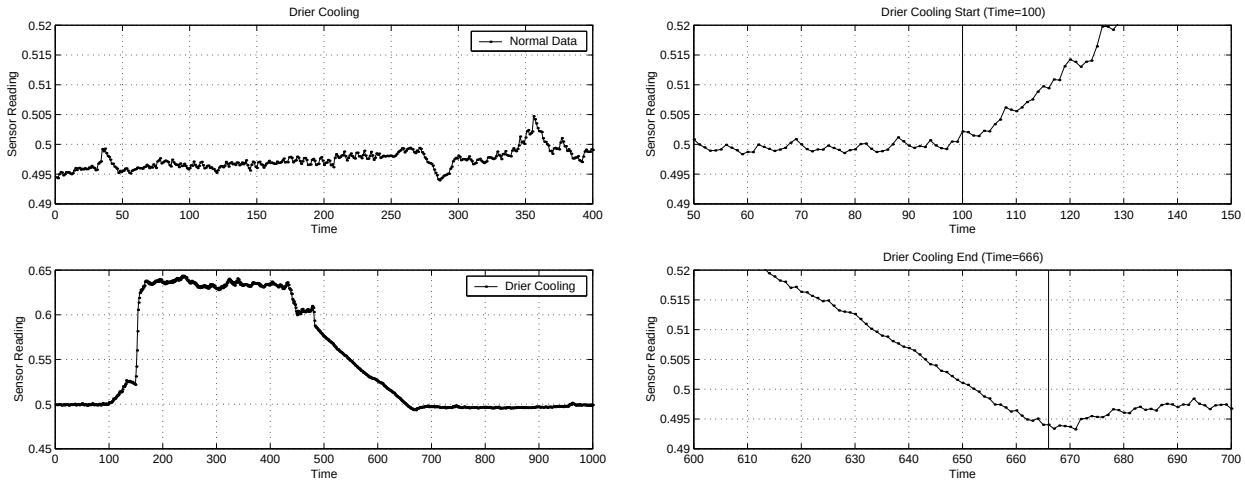
VI. Industrial Case Studies

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 Section 3, 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 below 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. 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 14a and 14b 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



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

Fig. 14. Training and Test Data for the Drier Cooling Event

(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 15.

The test data from Figure 14b were subjected to ART-2, MSART-2, and ART-2+MA detectors trained on the normal data from Figure 14a. After the onset of the event at time-step 100, all three methods detected the event at approximately equal times and continued to detect it consistently. Towards the end of the event, however, the ART-2 detector missed approximately 10 genuine alarms more than the ART-2+MA and MSART-2 detectors. 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.

B. 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 16a. 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. 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

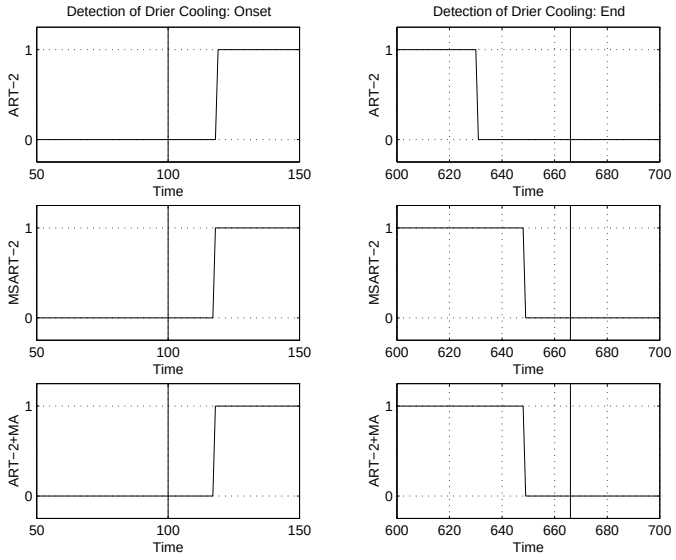
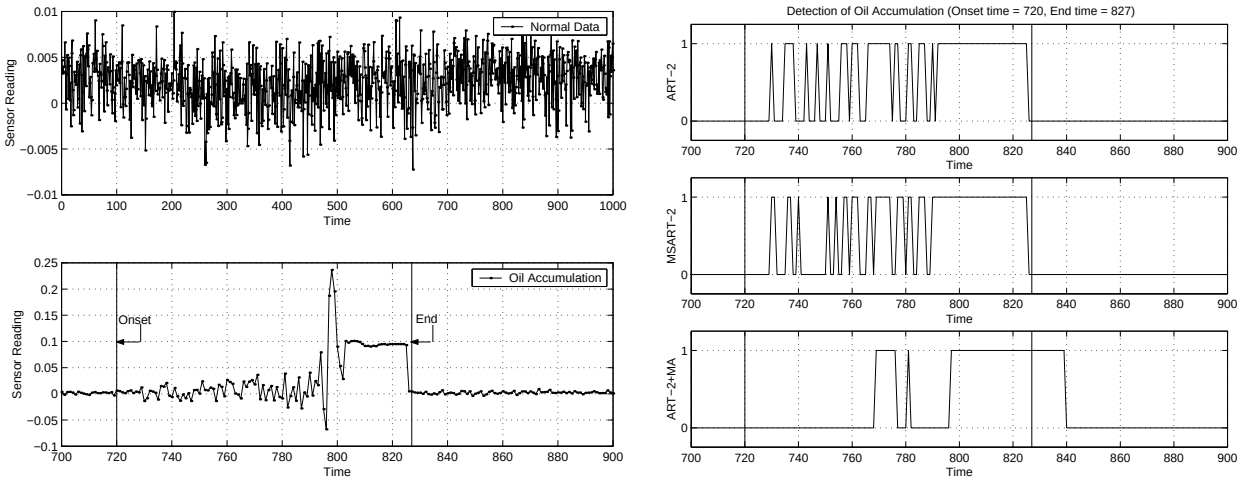


Fig. 15. Comparative Detection Performance for the Drier Cooling Event



(a) The Oil Accumulation Event

(b) Comparative Detection Performance

Fig. 16. Detection of Oil Accumulation (Sensor Failure)

after the end of the sensor failure. 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 16b). Similar to figure 15, we observe that MSART-2 conforms to the scale of the change under consideration and mimics the best performance for the event at hand.

VII. Conclusion

Previous work by the authors established ART-2 as a mechanism for efficiently and adaptively capturing linear and non-linear 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 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.

References

[1] J. R. Whiteley, 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 an Humans*, vol. 26, no. 4, pp. 423, 1996.

[2] S. N. Kavuri and V. Venkatasubramanian, "Representing bounded fault classes using neural networks with ellipsoidal activation functions," *Computers and Chemical Engineering*, vol. 17, pp. 139, 1993.

[3] G. A. Carpenter and S. Grossberg, "A massively parallel architecture for a self-organizing neural pattern recognition machine," *Computer Vision, Graphics and Image Processing*, vol. 37, pp. 54, 1987.

[4] G. A. Carpenter, S. Grossberg, and J. H. Reynolds, "ARTMAP: Supervised real-time learning and classification of nonstationary data by a self-organizing neural network," *Neural Networks*, vol. 4, pp. 565, 1991.

[5] G. A. Carpenter, S. Grossberg, and D. B. Rosen, "Fuzzy ART: An adaptive resonance algorithm for rapid, stable classification of analog patterns," in *Proceedings of the International Joint Conference on Neural Networks*, 1991, vol. 2.

[6] G. A. Carpenter, S. Grossberg, and D. B. Rosen, "Fuzzy ART: Fast, stable learning and categorization of analog patterns by an adaptive resonance system," *Neural Networks*, vol. 4, pp. 759, 1991.

[7] G. A. Carpenter, 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*, vol. 3, no. 5, pp. 698, 1992.

[8] D. Wienke and L. Buydens, "Adaptive resonance theory based neural networks - the 'ART' of real-time pattern recognition in chemical process monitoring?," *Trends in analytical chemistry*, vol. 14, no. 8, pp. 398, 1995.

[9] D. Wienke 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*, vol. 32, pp. 151, 1996.

[10] D. Wienke, W.van 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*, vol. 32, pp. 165, 1996.

[11] X. Song, 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*, vol. 41, pp. 161, 1998.

[12] X. Z. Wang, 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," *Computers and Chemical Engineering*, vol. 23, pp. 945, 1999.

[13] J. R. Whiteley and J. F. Davis, "Knowledge-based interpretation of sensor patterns," *Computers and Chemical Engineering*, vol. 16, no. 4, pp. 329, 1992.

[14] T. Frank, K. Friedrich, and T. Kuhlen, "Comparative analysis of fuzzy ART and ART-2A network clustering performance," *IEEE Transactions on Neural Networks*, vol. 9, no. 3, pp. 544, 1998.

[15] S. Marriott and R. F. Harrison, "A modified fuzzy ARTMAP architecture for the approximation of noisy mappings," *Neural Networks*, vol. 8, no. 4, pp. 619, 1995.

[16] C. P. Lim and R. F. Harrison, "Modified fuzzy ARTMAP approaches Bayes optimal classification rates: An empirical demonstration," *Neural Networks*, vol. 10, no. 4, pp. 755, 1997.

[17] N. Srinivasa, "Learning and generalization of noisy mappings using a modified PROBART neural network," *IEEE Transactions on Signal Processing*, vol. 45, no. 10, pp. 2533, 1997.

[18] J. R. Williamson, "Gaussian ARTMAP: A neural network for fast incremental learning of noisy multidimensional maps," *Neural Networks*, vol. 9, no. 5, pp. 881, 1996.

[19] B. R. Bakshi, "Multiscale PCA with application to multivariate statistical process monitoring," *AIChE Journal*, vol. 44, no. 7, pp. 1596, 1998.

[20] B. R. Bakshi, 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.

[21] H. B. Aradhye, B. R. Bakshi, R. Strauss, and J. F. Davis, "Multiscale statistical process control using wavelets - theoretical analysis and properties," *Technometrics*, under review.

[22] S. G. Mallat, "A theory for multiresolution signal decomposition: The wavelet representation," *IEEE Transactions and Pattern Analysis and Machine Intelligence*, vol. 11, no. 7, pp. 674, 1989.

[23] Gilbert Strang, "Wavelets and dilation equations: A brief introduction," *Society for Industrial and Applied Mathematics*, vol. 31, no. 4, pp. 614, 1989.

[24] B. R. Bakshi and G. Stephanopoulos, "Representation of process trends- iv. induction of real-time patterns from operating data for diagnosis and supervisory control," *Computers and chemical Engineering.*, vol. 18, no. 4, pp. 303, 1994.

[25] M. N. Nounou and B. R. Bakshi, "Online multiscale filtering of random and gross errors without process models," *AIChE Journal*, vol. 45, no. 5, pp. 1041, 1999.

[26] R. Luo, M. Misra, S. J. Qin, R. Barton, and D. M. Himmelblau, "Sensor fault detection via multiscale analysis and nonparametric statistical inference," *Industrial Engineering and Chemistry Research*, vol. 37, pp. 1024, 1998.

[27] S. Palavajjhala, R.L. Motard, and B. Joseph, "Process identification using discrete wavelet transforms: Design of prefilters," *AIChE Journal*, vol. 42, no. 3, pp. 777, 1996.
[28] M. S. Crouse, R. D. Nowak, and R. G. Baraniuk, "Wavelet-based statistical signal processing using hidden Markov models," *IEEE Transactions on Signal Processing*, vol. 46, no. 4, pp. 886, 1998.
[29] B. M. Sadler, T. Pham, and L. C. Sadler, "Optimal and wavelet-based shockwave detection and estimation," *Journal of the Acoustic Society of America*, vol. 104, no. 2, pp. 955, 1998.
[30] B. M. Sadler and A. Swami, "Analysis of multiscale products for step detection and estimation," *IEEE Transactions on Information Theory*, vol. 45, no. 3, pp. 1043, 1999.
[31] B. M. Sadler and A. Swami, "On multiscale wavelet analysis for step estimation," *International Conference on Acoustics, Speech, and Signal Processing, Seattle, WA*, vol. 3, pp. 1517, 1998.
[32] A. Swami, "Cramer-Rao bounds for deterministic signals in additive and multiplicative noise," *Signal Processing*, vol. 53, pp. 231, 1996.
[33] A. Swami and B. M. Sadler, "Step-change localization in additive and multiplicative noise via multiscale products," in *32nd Asilomar Conference on Signal, System, and Computation*, Pacific Grove, CA, 1998.
[34] A. Swami and B. M. Sadler, "Cramer-Rao bounds for step-change localization in additive and multiplicative noise," in *9th IEEE Statistical Signal and Array Processing Workshop*, 1998, p. 403.
[35] A. Denjean and F. Castanie, *Wavelets: Theory, Algorithms, and Applications*, chapter Mean value jump detection: A survey of conventional and wavelet based methods, Newkoski Academic, 1994.
[36] K. C. Chou and L. P. Heck, "A multiscale stochastic modeling approach to the monitoring of mechanical systems," in *Proceedings of the IEEE International Symposium on Time-Frequency, Time-Scale Analysis*, 1994.
[37] J. R. Whiteley and J. F. Davis, "Qualitative interpretation of sensor patterns," *IEEE Expert- Intelligent Systems and Their Applications*, vol. 8, no. 2, pp. 54, 1993.
[38] D. C. Montgomery, *Introduction to Statistical Quality Control*, John Wiley and Sons, Inc., New York, 1996.