Anomaly Detection Using Multiscale Methods

DISSERTATION

Presented in Partial Fulfillment of the Requirements for

the Degree Doctor of Philosophy in the

Graduate School of The Ohio State University

By

Hrishikesh Aradhye, B.Tech., M.S.

* * * * *

The Ohio State University

2001

Dissertation Committee:

Dr. James F. Davis, Adviser

Dr. Bhavik R. Bakshi

Dr. Stanley C. Ahalt

Dr. David Tomasko

Approved by

Adviser

Department of Chemical Engineering

ABSTRACT

In an environment where most process maneuvers are automated, algorithms to detect and classify abnormal trends in process measurements are of critical importance. The petrochemical industry in the United States loses billions of dollars annually due to improper abnormal situation management, and a staggering one in 16 plant accidents results in a fatality. Hence, Statistical Process Control and Monitoring (SPC) has been an active area of research for many decades and a variety of statistical and machine learning-based methods have been developed. However, most existing methods for process monitoring learn the signal characteristics at a fixed scale, and are best for detecting changes at that single scale. In contrast, data from most industrial processes are inherently multiscale in nature due to events occurring with different localization in time, space, and frequency. Unfortunately, existing techniques are unable to adapt automatically to the scale of these features. Many existing methods also require the measurements to be uncorrelated, whereas, in practice, autocorrelated measurements are very common in industrial processes.

In this work, we have investigated the use of multiscale techniques to improve upon these shortcomings of existing single-scale approaches. Because of fundamental functional relationships such as process chemistry, energy and mass balances, measurements in multivariate processes are correlated. Our approach learns these correlations and clustering behaviors in the wavelet space using machine learning

methods such as Adaptive Resonance Theory (ART-2) and Principal Component Analysis (PCA), resulting in higher detection accuracy coupled with noise reduction. The performance of our method, named Multi-Scale Statistical Process Control and Monitoring (MSSPC), is compared with existing methods based on the average detection delays for detecting shifts of different sizes. Our ART-2 based MSSPC detector is currently deployed in a large scale petrochemical plant to detect process anomalies in real time by incrementally learning normal process operation in the wavelet domain. Several case studies for the detection of real process malfunctions, including the comparison with the performance of human operators, are also presented in this work. These results indicate that MSSPC is a good method for monitoring of measurements with unknown and different types of changes.

This thesis is dedicated to my parents.

ACKNOWLEDGMENTS

I offer my sincere thanks to my adviser, Dr. James Davis, for his support, foresight, and guidance through many a frustrating moment in the course of this project.

I thank Dr. Bhavik Bakshi and Dr. Stanley Ahalt for introducing the field of signal processing to me.

I would like to gratefully acknowledge the efforts by Suresh Ramachandran, Iky Chan, Magaly Barroeta, Ramon Strauss, and Saravanarajan Rathinasabapathy in their assistance with some of the experiments and charts presented in this dissertation. Many thanks to Zhiguo (Godwin) Zhang for helping with the graduation procedures.

The Abnormal Situation Management Consortium and the Ohio State University (Presedential Fellowship Program) provided the funding for this research. I thank them sincerely for this financial assistance.

VITA

November 1, 1973	Born - Nashik, India
1995	B.Tech. (Chemical Engineering), Indian Institute of Technology, Bombay
1997	M.S. (Chemical Engineering), University of New Mexico, Albuquerque, NM
2000	M.S. (Electrical Engineering), The Ohio State University, Columbus, OH

PUBLICATIONS

Research Publications

V. Sreenivas, K. M. Moudgalya, H. B. Aradhye, and G. Sivakumar, "Quantifying goodness of initial guesses for automatic selection of nonlinear algebraic equation solvers," in *Proceedings of the 14th IMACS World Congress on Computational and Applied Mathematics*, 1994.

H. B. Aradhye and A. S. Heger, "Sensor fault diagnosis using hierarchical Bayesian belief networks," in *Proceedings of the American Nuclear Society Annual Winter Conference*, November 1997.

H. Aradhye and A. S. Heger, "Recognizing patterns in information retrieval: A memory-based classifier for inferring relevancy," *Artificial Intelligence in Engineering*, vol. 12, no. 1-2, pp. 99–106, 1998.

A. S. Heger and H. B. Aradhye, "Bayesian belief networks for sensor fault detection, isolation, and accommodation," Patent Application, University of New Mexico, 1999, Provisionally Accepted by the US Patent Office.

H. Aradhye and C. Dorai, "New efficient learning architecture for multi-class SVM-based classification," Invention Disclosure YOR8-2000-0756, IBM T. J. Watson Research Center, Yorktown Heights, NY, 2000.

H. Aradhye and C. Dorai, "Method for recognizing and classifying video text and video characters using kernel-space methods," Invention Disclosure YOR8-2000-1038, IBM T. J. Watson Research Center, Yorktown Heights, NY, 2000.

H. Aradhye and A. S. Heger, *Fuzzy Logic and Probability Applications*, chapter Signal Validation Using Bayesian Belief Networks, SIAM, 2001, In Print.

H. Aradhye, C. Dorai, and J. Shim, "Study of embedded font context and kernel space methods for improved videotext recognition," in *Proceedings of the IEEE International Conference on Image Processing*, 2001, Accepted.

C. Dorai, H. Aradhye, and J. Shim, "End-to-end videotext recognition for multimedia content analysis," in *Proceedings of the IEEE International Conference on Multimedia and Expo*, 2001, Accepted.

H. B. Aradhye, J. F. Davis, B. R. Bakshi, and S. C. Ahalt, "ART-2 and multiscale ART-2 for process fault diagnosis - validation via industrial case studies and Monte Carlo simulation," in *Proceedings of the 4th IFAC Workshop on On-Line Fault Detection and Supervision*, 2001, Accepted as a Keynote Presentation.

H. B. Aradhye, B. R. Bakshi, J. F. Davis, and S. C. Ahalt, "Clustering in wavelet domain: A multiresolution ART network for anomaly detection," *IEEE Transactions on Systems, Man, and Cybernetics*, 2000, Submitted.

H. B. Aradhye, B. R. Bakshi, R. Strauss, and J. F. Davis, "Multiscale statistical process control using wavelets - theoretical analysis and properties," *Technometrics*, 2000, Submitted.

H. Aradhye and A. S. Heger, "Use of Bayesian belief networks for sensor fault detection, isolation, and accommodation," *Artificial Intelligence in Engineering*, 2000, Submitted.

H. Aradhye and C. Dorai, "Augmented edit distance based temporal contiguity analysis for improved videotext recognition," in *Proceedings of the IEEE 2001 Workshop on Multimedia Signal Processing*, 2001, Submitted.

FIELDS OF STUDY

Major Field: Chemical Engineering

Studies in:

Machine Learning and Anomaly Detection	Prof. James Davis
Wavelets and Statistical Signal Processing	Prof. Bhavik Bakshi
Computer Vision and Image Processing	Prof. Stanley Ahalt

TABLE OF CONTENTS

Page

LIST OF TABLES

Table **Page**

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

1.1 Problem Statement

An estimated eleven billion dollars per year are lost in the chemical process industry as a result of ineffective management of abnormal situations [2]. While most abnormal situations result in lost products, lost productivity, or process inefficiency, a staggering 1 in 16 abnormal situations results in a fatality. Even with sophisticated data collection and computerized control systems, there remains a need to better interpret data and improve decision support with respect to detection and identification of faults. The problem of analyzing and interpreting process data is made difficult because there are typically hundreds of sensor measurements that are often correlated. Both the process and the sensors introduce noise that can mask the relevant information. The process can be in a wide range of disjoint conditions of operation which are continuously changing as result of process improvement.

More precisely, a process fault is defined as a "non-permitted deviation of a characteristic property which leads to inability to fulfill the intended purpose" [3]. In an environment where most process maneuvers are automated, development of procedures to detect and classify abnormal process behavior by observation of measured signals is of critical importance. This work presents a novel and now proven pattern

recognition approach that can readily identify complex, nonlinear, correlated, noisy, and multivariate process behaviors for the purpose of fault detection and diagnosis.

1.2 Motivation

Most industrial chemical processes, by their inherent nature, are not precisely defined because the underlying physical and chemical process transformations cannot be fully described. A chemical process may track any of several possible paths of operation depending on a large number of known and unknown factors, and may exist in multiple regions of operation. A process can move from one normal operation to another as result of conditions such as changing feed or changing weather. Likewise, the process can shift into abnormal regions of operation as result of equipment or operator failures. There can exist several such normal and abnormal regions and they can be contiguous or disjoint. Data clustering approaches for defining these regions are intuitively suitable for fault detection and identification.

A critical feature for a fault detection approach is the ability to adapt with changing process conditions and behaviors. New regions of normal operation are continuously sought and new regions of abnormal operation are always possible. The data analysis approach must therefore accommodate new information and knowledge about the process, and at the same time, keep the previously acquired information intact. The Adaptive Resonance Theory (ART) networks are among a very few clustering algorithms that explicitly address this information management capability. A framework for ART-based data analysis for detection and identification has been developed by Davis and co-workers and has been applied to large-scale industrial processes. ART-based clustering algorithms, however, are very sensitive to noise. They

require careful definition of regions identified, selection of data used to train, and configuration of the algorithm to achieve good performance. Analogously, Principal Component Analysis (PCA), which has been a popular unsupervised modeling technique for the statistical process control community for linearly correlated processes, also fails to adequately account for process behaviors and noise.

Noise sensitivity can be attributed in part to the fact that most naturally occurring process signals are a combination of components corresponding to different events occurring at different localizations in time and frequency. For example, equipment degradation occurs over long time intervals and low frequencies. In contrast, sensor noise is spread across all frequencies and times. Specialized processing of the signal at different scales benefits data analysis tasks such as detection and identification.

A multiscale hierarchy presented in this work that reflects sensor signals decomposed into time and frequency (wavelet) behaviors, provides a systematic way of selecting the signal resolutions most relevant at the current time for a particular kind of process event. For univariate and linearly correlated multivariate processes, this wavelet-based, signal decomposition scheme is shown in this work to provide significant improvement in detection performance over conventional statistical process monitoring techniques. This dissertation is thus focused on combining PCA and ART's ability to model linear or nonlinear disjoint process regions and manage new information, with the benefits of multiresolution data processing for greater noise tolerance and quicker, more robust detection.

1.3 Results and Contributions

Early applications of ART-based detection and identification were limited to relatively small process units and/or simulated data. The first phase of this work involved a large-scale process study involving the application of ART to petrochemical plant data over two years of operation but with no multiscale decomposition or specialized noise handling approaches. Results showed that ART can detect process faults online, and well in advance of the human operator. In some cases, ART was able to detect process malfunctions that went unnoticed by the plant engineers.

The next phase of this work involved the investigation of multiresolution data analysis techniques for noise tolerance and detection accuracy. One of the main contributions is the theoretical derivation of performance curves for the multiscale Statistical Process Monitoring (MSSPC) and multiscale Principal Component Analysis (MSPCA) methods. The theoretical curves have been validated against those calculated from Monte Carlo simulation studies for both univariate and linear multivariate processes. The significance of this is that there now exists a theoretical justification for the utility of these methods and a means of parameter configuration according to desired operational objectives.

An important extension of the multiscale scheme by Bakshi [4] to multivariate nonlinear modeling has been its integration with ART. We refer to the approach as multi-scale ART or MSART. In the final phase of this work, the MSART network has been used to successfully model nonlinear relationships in the wavelet domain as well as the reconstructed signal domain. In practice, detection is shown to be quicker and more consistent than the single scale ART detector by Whiteley and Davis. This result has been validated using the comparative performance curves for ART and

MSART generated over several simulated and real processes of varying statistical nature.

1.4 Impact on Process Systems Engineering

In general, this research has developed a comprehensive approach for state estimation of multi-sensor processes that provides robust detection and identification performance, and at the same time, is easy to configure and use. The approach accounts for multivariate, non-linear, correlated and noisy data and addresses a critical practical need of incorporating new information incrementally. It has been applied and demonstrated for large scale chemical processes. Success is to a point that the approach is being considered for integration into a commercial product. This work overlaps significantly with several areas including chemical process engineering, computer science, digital signal processing, and statistics. The general nature of the approach has both drawn upon work in these diverse disciplines and has led to its consideration in parallel application areas such as network security and intrusion detection, signal detection in wireless communications, and digital speech processing.

CHAPTER 2

BACKGROUND

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, for learning the characteristics of the data. This work has primarily focused on two such techniques: Principal Components Analysis (PCA) and Adaptive Resonance Theory (ART). We will present a method for multiscale fault detection using these two techniques and demonstrate significant performance enhancements on the basis of extensive simulated as well as industrial case studies.

To this end, this dissertation has built upon and enhanced the MSSPC hierarchy by Bakshi (1998) and the ART-based diagnostic model by Whiteley and Davis (1996). This chapter provides the basic background comprised of the previous work in this area, the description of the MSSPC algorithm, and the description of the ART-2 diagnosis model.

2.1 Previous Work

Statistical Process Control (SPC) has been an active area of research for many decades and a variety of methods have been developed. These include methods for univariate SPC such as, Shewhart, moving average (MA), exponentially weighted moving average (EWMA), and cumulative sum (CUSUM) charts. Methods for multivariate SPC include multivariate extensions of univariate methods, and methods that monitor latent variables obtained by combining the measured variables to a lower dimension space. Popular methods for reducing the dimensionality of the measured data include principal component analysis (PCA) and partial least squares regression (PLS). These dimensionality reduction methods have been very popular for multivariate SPC in the chemical industry, and many extensions and applications have been developed [5][6][7].

Most existing univariate and multivariate SPC methods operate at a fixed scale, and are therefore best for detecting changes at a single scale. For example, Shewhart charts analyze the raw measurements, thus representing them at the scale of the sampling interval or the finest scale, and are best for detecting localized and large changes. In contrast, MA, EWMA, and CUSUM charts consider the measurements

at a coarser scale through a transformation using a linear filter, and are best for detecting small shifts or features at coarser scales.

In contrast to the single-scale nature of SPC methods, data from most practical processes are inherently multiscale in nature due to events occurring with different localization in time, space and frequency. A typical example of such data from a petrochemical process is shown in Figure 2.1. Figure 2.1(top) shows data during normal operation, while Figure 2.1(bottom) represents unusual operation due to a drier cooling event. In Figure 2.1(bottom), the process change at approximately 150 time units is at a very fine scale and localized in time, but spans a wide range of frequencies. The steady portions of the signal are at coarser scales and span a wide temporal range. Finally, the change between 425 and 675 time units consists of a small sharp change followed by a short steady section and a slow ramp at an intermediate scale. Ideally, techniques for detecting changes at different scales, such as those shown in Figure 2.1(bottom), should adapt automatically to the scale of the features. In response to this need, many heuristic or ad hoc techniques have been proposed for overcoming the single-scale nature of SPC charts. These include the Western Electric rules [8] useful for identifying patterns in data and combined Shewhart and CUSUM charts [9] for identifying large and small shifts.

Another characteristic of many existing SPC methods is that they assume the measurements to be uncorrelated, or white, whereas, in practice, autocorrelated measurements are extremely common. A common approach for SPC of autocorrelated measurements is to decorrelate them by fitting a time series model, and monitoring

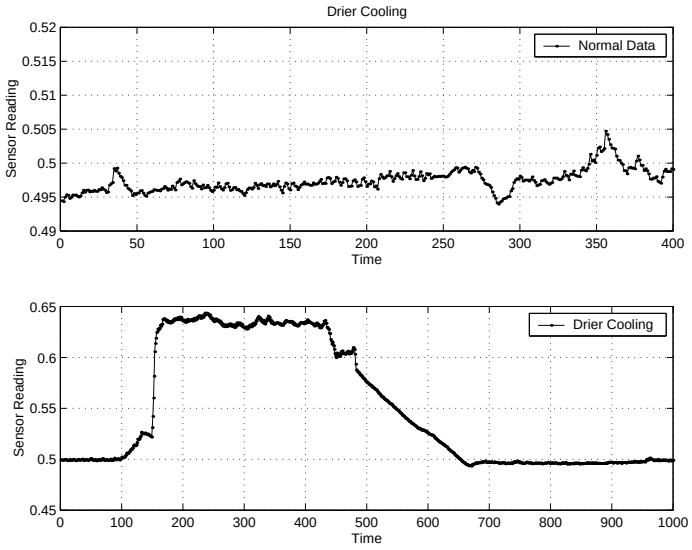


Figure 2.1: Data from a petrochemical process: (top) Normal Operation, and (bottom) Abnormal operation representing the drier cooling event

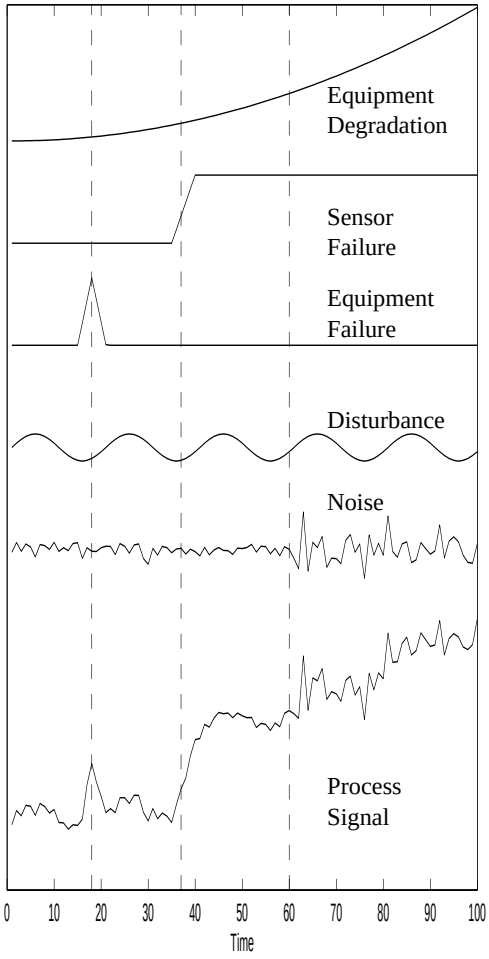
the residual error. However, this approach is not practical, particularly for multivariate processes. Other univariate approaches for decorrelating autocorrelated measurements without time-series modeling include taking the batch means [10], and finding the residuals between the measurements and their one-step ahead prediction by a moving centerline EWMA (MCEWMA) model [11]. Batch means control charts often result in a significant delay in detecting large shifts, while MCEWMA is best only for IMA stochastic processes. Consequently, neither of these approaches is broadly applicable to different types of stochastic processes and process shifts and they lack multivariate generalizations. For multivariate SPC, the measurements are usually decorrelated by including lagged values of the variables [6]. The resulting model obtained by methods such as PCA or PLS implicitly decorrelates the measurements by extracting the linear time-series model. Traditional T^2 and Q control charts are then

applied in the space of the selected latent variables and its residuals. This approach often works better than SPC by steady-state PCA or PLS for autocorrelated data.

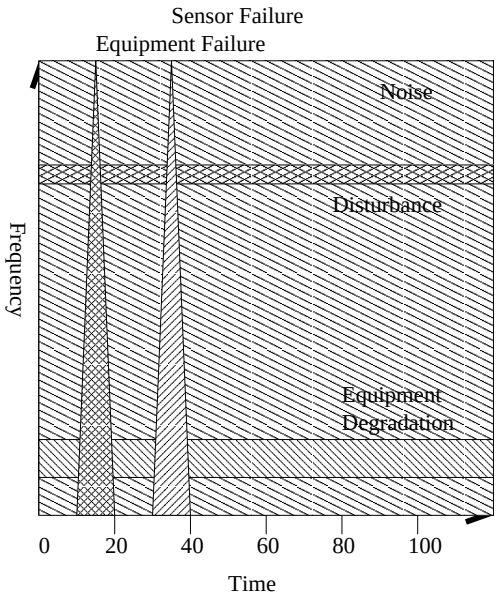
In recent years, wavelets have been popular for analyzing multiscale or autocorrelated measurements due to their ability to compress multiscale features and approximately decorrelate many autocorrelated stochastic processes [1][12]. Thus, wavelet coefficients provide compact information about a signal at different localizations in time and frequency. Also, the wavelet coefficients of many stochastic processes are approximately white since wavelets are approximate eigenfunctions of many mathematical operators [13]. Wavelet-based multiscale methods have been developed for improved solutions of many tasks including data compression, estimation, feature extraction, and filtering [1][14][12][15][16].

Advantages of these applications of multiscale techniques 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 2.2a and 2.2b). 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.

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. [17] proposed a wavelet-domain Hidden Markov Model for univariate statistical signal processing. Swami, Sadler, and co-workers [18][16][19][20][21][22] have presented multiscale methods for step detection and estimation. Other researchers [23][24] have investigated



(a) Scales of Signal Components of a Typical Process



(b) Schematic Time-frequency Representation

Figure 2.2: Multiresolution Analysis of a Typical Process Signal

wavelet-based shockwave detection, mean value jump detection, monitoring applications for 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 [4]. 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.

The benefits of the multiscale representation using wavelets have also been extended to change detection and process monitoring. Methods for Bayesian multiscale change detection require prior knowledge about the nature of the process change [25]. Such methods can perform well if the necessary information is available, but are not popular in industrial monitoring applications due to a lack of the necessary information for the prior, and a lack of familiarity with Bayesian methods [26].

Our approach, called multiscale SPC (MSSPC), consists of decomposing each measured variable to multiple scales by using a selected family of wavelet basis functions. The decomposition permits identification of signal features at various scales as relatively large coefficients in uncorrelated data. This work presents theoretical

analysis and properties of univariate and multivariate MSSPC. The performances of MSSPC and existing methods are compared based on the average run lengths (ARL) for detecting shifts of different sizes. The ARL is determined both empirically by Monte-Carlo simulation, and theoretically for uncorrelated as well as stationary and nonstationary autocorrelated measurements. For autocorrelated univariate measurements, MSSPC is compared with residuals, weighted batch means [10] and MCEWMA [11] charts. The performance of multivariate MSSPC using PCA is compared with steady-state and dynamic PCA. The theoretical analysis and application to industrial data indicate that MSSPC is a good general method for SPC of processes containing features of different sizes and at different scales from uncorrelated or autocorrelated measurements. Furthermore, MSSPC subsumes existing methods such as Shewhart, MA, EWMA, and CUSUM charts, depending on the nature of the selected wavelet. Thus, MSSPC can adapt the data filter and the detection limits according to the nature of the process change, and can specialize to existing methods if necessary. These results indicate that MSSPC is a good method for SPC of measurements with unknown and different types of changes. If the nature of the features representing abnormal operation is known a priori, then existing methods may be tailored to outperform MSSPC.

2.2 Wavelets

A family of wavelet basis functions may be represented as,

$$\psi_{su}(t) = \frac{1}{\sqrt{s}}\psi\left(\frac{t-u}{s}\right) \tag{2.1}$$

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

parameters as, $s = 2^m$, $u = 2^m k$, with m and k being integers, permits the wavelets to be orthonormal [27]. Efficient methods have been developed for decomposing a signal on a family of wavelet basis functions based on convolution with the corresponding filters [14]. Thus, any signal can be decomposed to its contribution at multiple scales as a weighted sum of dyadically discretized orthonormal wavelets.

$$y(t) = \sum_{m=1}^{L} \sum_{k=1}^{N} d_{mk} \psi_{mk}(t) + \sum_{k=1}^{N} a_{Lk} \phi_{Lk}(t) \tag{2.2}$$

where, y is the measurement, d_{mk} are the detailed wavelet coefficients at the scale m, a_{Lk} are the scaled signal coefficients at the coarsest scale L. Orthonormal wavelets have found application in many data analysis and modeling tasks due to their computational efficiency, ability to compress deterministic features in a small number of relatively large wavelet coefficients, and to approximately diagonalize a variety of mathematical operators. These properties are illustrated in Figures 2.3 and 2.4. Figure 2.3 shows the wavelet decomposition of an example process previously presented in Figure 2.1(bottom). The fine scale features are captured by the large wavelet coefficients in Figures 2.3b, c, and d, and the remaining coarse scale features by the last scaled signal in Figure 2.3e. Thresholding the wavelet coefficients by eliminating small coefficients and reconstructing the signal results in a signal with less random variation, as shown in Figure 2.3f. Thus, the stochastic noise in the signal may be reduced by thresholding the coefficients at each scale to retain only the large coefficients. The properties of such a wavelet thresholding approach have been studied extensively [1].

The decorrelation ability of wavelets is depicted in Figure 2.4. The original signal, a nonstationary stochastic process, and its autocorrelation function are shown in Figure 2.4a. The wavelet coefficients and corresponding autocorrelation functions

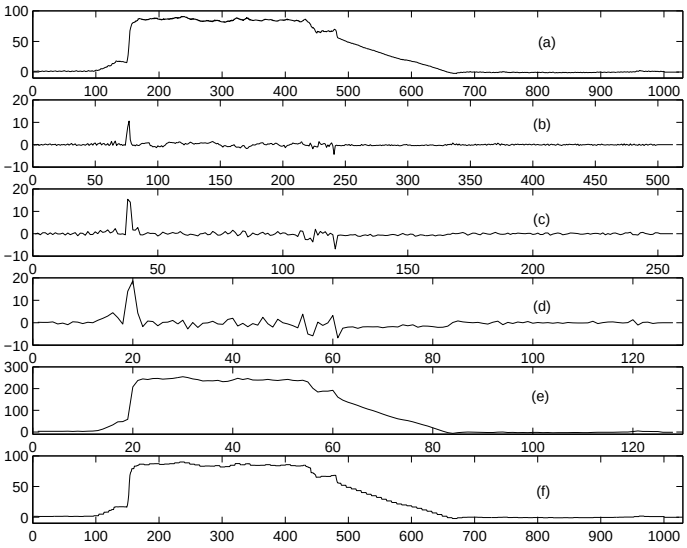


Figure 2.3: Extraction of deterministic changes by wavelet decomposition. (a) original signal, (b) wavelet coefficients at $m = 1$, (c) wavelet coefficients at $m = 2$, (d) wavelet coefficients at $m = 3$, (e) last scaled signal coefficients at $m = 3$, (f) reconstructed signal after wavelet thresholding [1]

shown in Figures 2.4b, c, and d indicate that the coefficients are approximately uncorrelated. The last scaled signal in Figure 2.4e contains some residual correlation, which may be further reduced by decomposing to coarser scales. The variance of the wavelet coefficients at each scale varies with the power spectrum of the original signal. Thus, the variance of the wavelet coefficients and the last scaled signal in Figures 2.4b, c, d, e, increases with decreasing frequency. These properties form the basis of the multiscale SPC method described in this work.

The wavelet decomposition in Figures 2.3 and 2.4 downsamples the coefficients at coarser scales due to the dyadically discretized translation parameter, as shown in Figure 2.5a. A disadvantage of using dyadically discretized wavelets is that every measurement is decomposed after a time delay, with this delay increasing at coarser

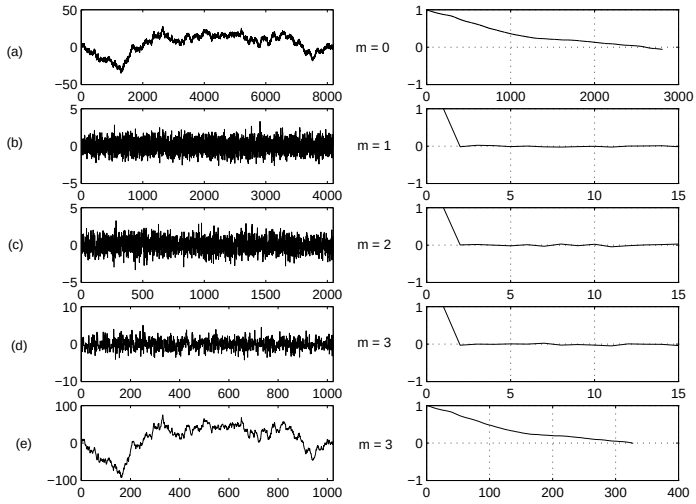


Figure 2.4: Approximate Decorrelation due to Dyadic Wavelet Transform

scales. Thus, the fifth measurement, x_5 is decomposed to d_{16} after x_6 is obtained, and to d_{28} after measuring x_8. Thus, if a mean shift occurs at x_5, it will not be detected at scale $m = 1$ without a delay of one sample, and at scale $m = 2$ without a delay of three samples.

This time delay can be eliminated by using wavelets with a uniformly discretized translation parameter, $u = k$. This results in wavelet decomposition without down-sampling, as shown in Figure 2.5b. The wavelets are no longer orthonormal to each other. Consequently, the decorrelation ability is lost, but the compression ability is retained. These wavelets are convenient for pattern recognition and on-line multiscale methods [15].

x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8
	d_{12}		d_{14}		d_{16}		d_{18}
			d_{24}				d_{28}
							d_{38}

(a)

x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8
	d_{12}	d_{13}	d_{14}	d_{15}	d_{16}	d_{17}	d_{18}
			d_{24}	d_{25}	d_{26}	d_{27}	d_{28}
							d_{38}

(b)

Figure 2.5: Wavelet decomposition with (a) Dyadic, and (b) Integer discretization of translation parameter.

2.3 Multiscale Statistical Process Control

2.3.1 Methodology

Figure 2.6 shows a schematic diagram of the MSSPC approach for online anomaly detection. Given the SPC parameters (such as detection thresholds) and the number of scales L, the following approach allows us to construct the SPC diagnostic models that constitute the MSSPC architecture. Let P be the number of process variables in a multivariate process. All the constituent diagnostic models of the MSSPC scheme model 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 phase, the following steps synthesize normal diagnostic models 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 2.6 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. The SPC diagnostic modeling algorithm of choice (such as PCA, PLS, ART, etc) is independently applied to each of these training matrices. Let the resulting diagnostic models in the wavelet domain be represented as $SPC\mathbf{D}_m$, $m = 1, \ldots, L$, and $SPC\mathbf{A}_L$, respectively. We thus have $L + 1$ SPCs that constitute the *Scale Selection Layer* of wavelet-domain detectors. For example, Figure 2.6 shows a Scale Selection Layer composed of $SPC\mathbf{D}_1$, $SPC\mathbf{D}_2$, $SPC\mathbf{D}_3$, $SPC\mathbf{D}_4$, and $SPC\mathbf{A}_4$, which represent diagnostic models of wavelet coefficients of normal data at the respective scales.

A crucial feature of the MSSPC 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 diagnostic modeling independently to each of these reconstructed training matrices to obtain diagnostic models in a signal space filtered to retain the selected combination of scales. These 2^{L+1} SPC diagnostic models, $SPC\hat{\mathbf{Y}}_i$, $i = 1, \ldots, 2^{L+1}$, constitute the *Diagnosis Layer* of detectors. In Figure 2.6, 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 2.6 is composed of 32 SPC diagnostic models, each of which represents a model 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 time-domain SPC. Hence the time-domain detector is a special case of the multiscale hierarchy presented in this work.

Online Testing

Having trained the Scale Selection Layer and Diagnosis Layer SPCs, 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 MSSPC approach.

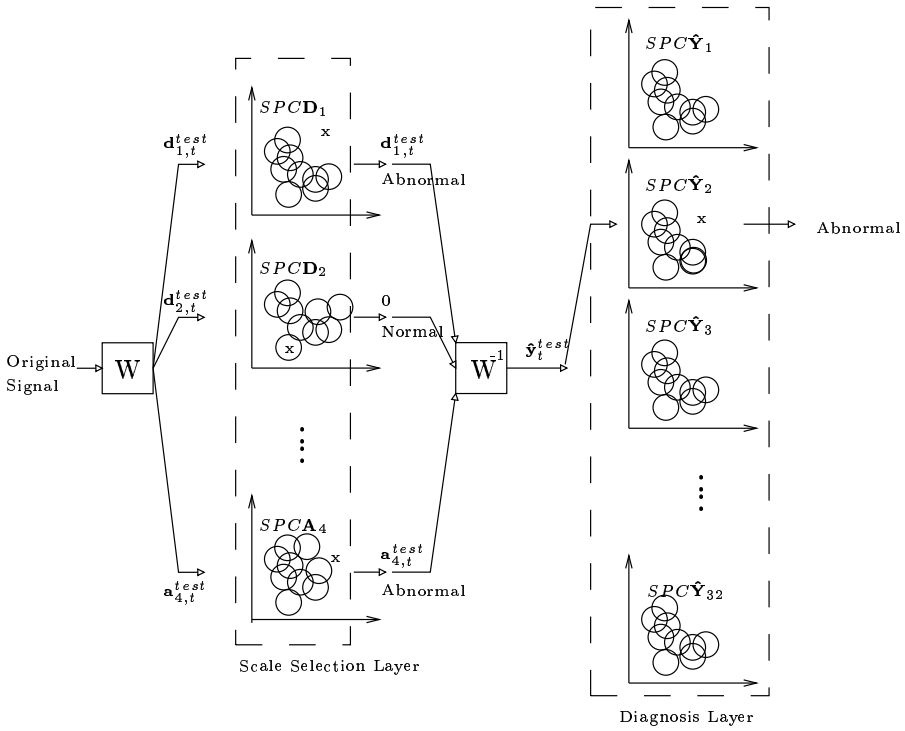


Figure 2.6: The MSSPC Architecture for Robust Fault Diagnosis

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 2.6 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 $SPC\mathbf{D}_m$ of the Scale Selection Layer. Similarly, construct the vector $\mathbf{a}_{L,t}^{test}$ to be presented to the detector $SPC\mathbf{A}_L$.

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 diagnostic models is above a pre-set detection threshold. Only if the network $SPC\mathbf{D}_m$ provides an "abnormal" diagnosis, the coefficients $d_{m,t,p}^{test}$, $p = 1, \ldots, P$, are retained for reconstruction. Similarly, only if the network $SPC\mathbf{A}_L$ provides an "abnormal" diagnosis, the coefficients $a_{L,t,p}^{test}$ are retained for reconstruction. For example, in Figure 2.6, the d_2 coefficient vector was deemed "normal" by $SPC\mathbf{D}_2$. 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} $SPC\hat{\mathbf{Y}}$ Diagnosis Layer detectors corresponding to the combination of scales selected for reconstruction. For instance, the chosen Diagnosis Layer network in Figure 2.6 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 MSSPC is of the order of $L+2$ times the computation for the time-domain detector. The worst-case storage requirement for MSSPC is of the order of $L + 1 + 2^{L+1}$ times the storage requirement for time-domain SPC.

2.3.2 Illustration

Figure 2.7 illustrates the potential of wavelet decomposition for process fault detection. As stated earlier (Figure 2.2a and 2.2b), 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 2.7 shows successive approximations of the signal from Figure 2.2a 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 $\mathbf{d}_4[3]$.

The MSSPC methodology and its features are illustrated by the univariate example shown in Figures 2.8 and 2.9. These figures use the Haar wavelet. The data

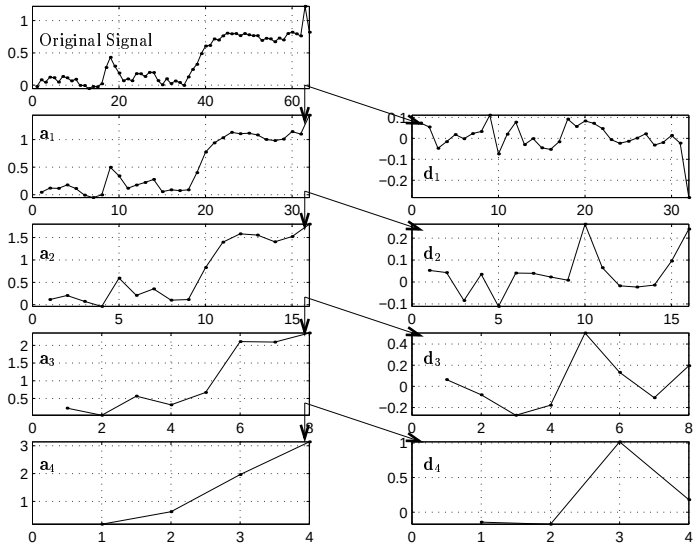


Figure 2.7: Representation of Process Signals at Successive Levels of Approximation

representing normal operation are independent, identically distributed (IID) Gaussian with unit variance. The detection limits at each scale are determined from the normal data. In this case, the detection limits at each scale are equal for dyadically discretized wavelets, since the coefficients are also uncorrelated with approximately unit variance. For uniformly discretized wavelets, the limits need to be adjusted to account for the lack of downsampling and autocorrelation in the coefficients. Abnormal operation is indicated in Figure 2.8a by a shift of magnitude 5 between samples 30 to 60. The results for the Shewhart chart and MSSPC for this case are shown in Figure 2.8. MSSPC results in a filtered signal that mainly contains the feature representing abnormal operation. Furthermore, the MSSPC detection limits change according to the nature of the signal, and the scales at which the features are present. The detailed steps in MSSPC at four time steps are shown in Figure 2.9. When

the shift first occurs at t=30, it is detected only at the finest scale, as shown in Figure 2.9b for t=30. The signal reconstructed with the coefficients outside the limits and the limit at the current time (t=30) are shown in Figure 2.9f, indicating that MSSPC detects the shift. The values of the limit and reconstructed signal at t=30 in Figure 2.9f form the corresponding points in Figure 2.8b. The plots for t=34 depict the behavior of MSSPC after the shift has persisted for some time. Now, the shift is detected by the wavelet and scaling function coefficients at the coarsest scale in Figures 2.9d and e. The detection limit at t=34 is depicted by the circle in Figure 2.9f, and is determined based on the variance of the normal data at the selected scales. As the shift persists, it is detected only by the last scaled signal. The behavior of MSSPC when the process returns to normal operation is shown by the plots at t=61 and t=64. At t=61, the last scaled signal continues to violate the detection limit, but the return to normal is picked up at the finest scale, as shown in Figure 2.9b. The reconstructed signal and limit at t=61 confirm that the process has returned to normal operation. As the return to normal persists, the last scaled signal continues to be outside the limit, but wavelet coefficients at other scales help in keeping the reconstructed signal within its limits. Finally, the last scaled signal also stops violating the detection limit. The figures for t=61 and t=64 clearly demonstrate the need for the final wavelet reconstruction and monitoring step, as discussed in the previous paragraph.

2.3.3 Integer versus Dyadic Discretization

As discussed in Section 2.2, the wavelet translation parameter may be discretized dyadically or uniformly. The type of discretization may be selected according to

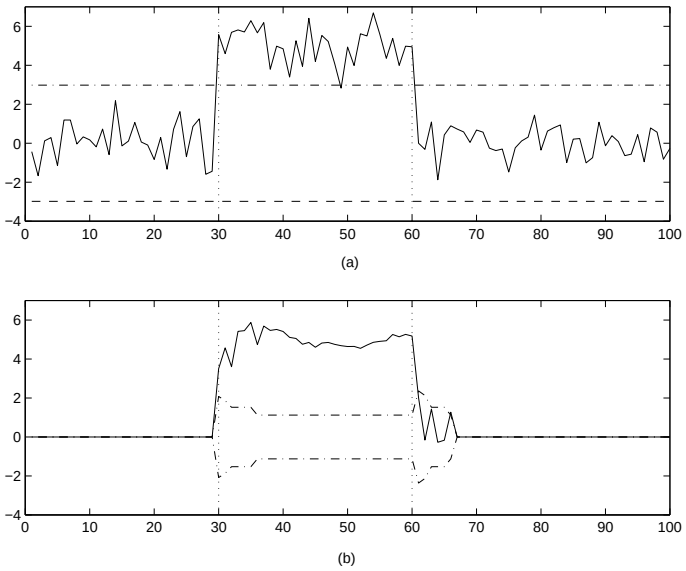


Figure 2.8: (a) Shewhart chart, (b) MSSPC. Detailed figures for obtaining each point on the reconstructed signal and limits are shown in Figure 2.9.

the nature of the measured data, and the objective of the SPC problem. Dyadic discretization permits the use of orthonormal wavelets and approximate decorrelation of autocorrelated measurements, but introduces a time delay in detecting changes due to a lag in the computation of the wavelet coefficients. Uniform discretization does not downsample the wavelet coefficients and permits truly on-line SPC, but the decorrelation ability and orthonormality are lost. These properties and the average run length analysis in this work indicate that, in general, wavelets with uniform discretization are best for on-line SPC of uncorrelated or moderately autocorrelated measurements, particularly for detecting large shifts. If the measurements are highly correlated or non-stationary, it is best to use dyadically discretized wavelets. For small shifts, the performance with dyadic or integer discretization is quite similar.

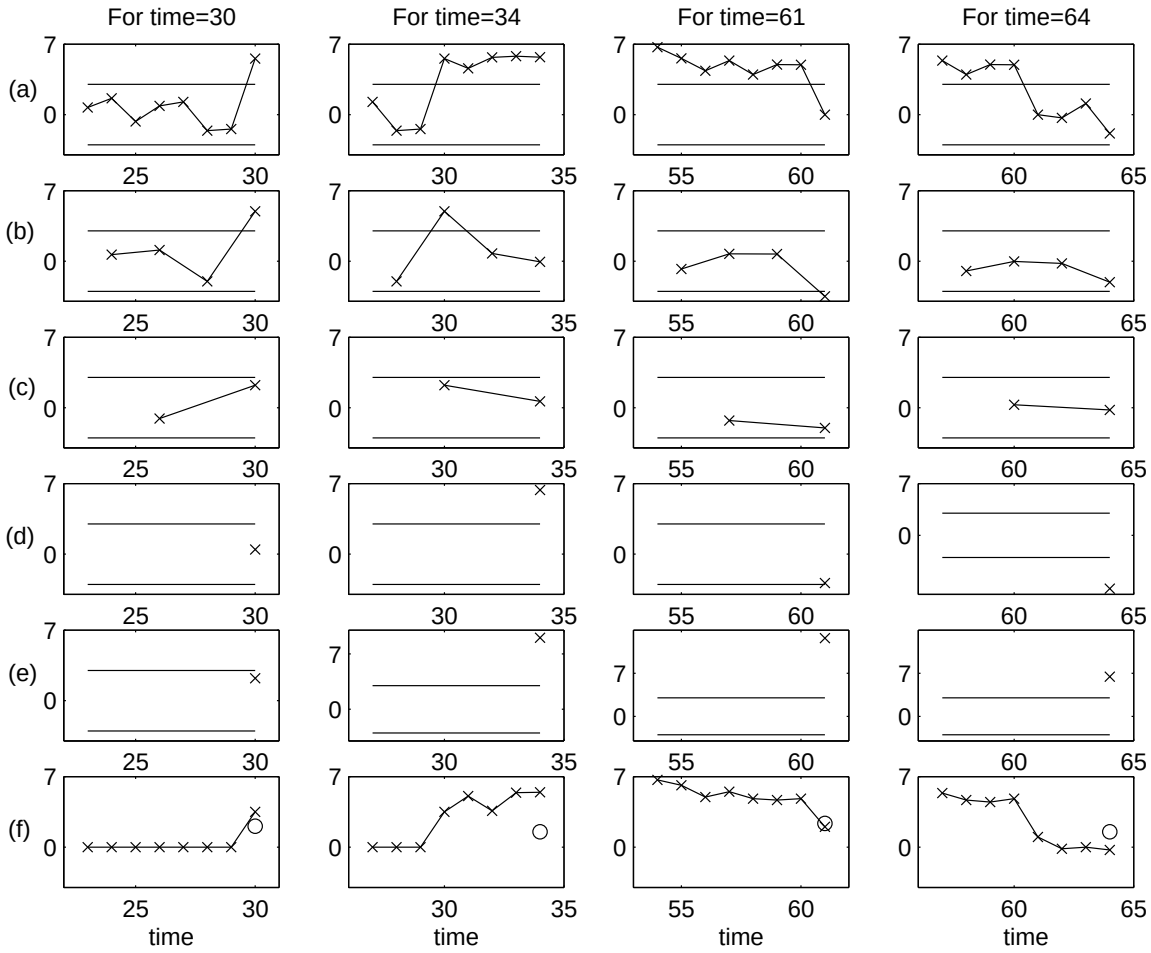


Figure 2.9: Illustration of MSSPC methodology. (a) original signal used for monitoring measurement at selected time, (b), (c), (d) wavelet coefficients at $m = 1, 2, 3$, (e) coefficients for last scaled signal, $m = 3$, (f) reconstructed signal ($\times$) and detection limit (o) corresponding to selected time, and plotted on Figure 2.8.

For *dyadic* discretization, if the measurements representing normal process operation are uncorrelated and Gaussian, the coefficients at each scale will also be uncorrelated and Gaussian with almost equal variance. If the normal data are autocorrelated, then the coefficients of an orthonormal wavelet decomposition at each scale will be approximately uncorrelated Gaussian, with the variance proportional to the power spectrum of the measurements in the corresponding frequency band. Due to these properties, the limits at each scale for MSSPC with dyadically discretized wavelets can be determined directly from the wavelet decomposition or power spectrum of normal data.

For *integer* discretization, the variance of the coefficients at each scale is still proportional to the power spectrum of the measured data, but the detection limits need to be adjusted to account for the fact that $O(N)$ coefficients in the original signal yield $O(N \times (L+1))$ coefficients for a decomposition depth of L, and for the autocorrelation between the measurements. The increase in the number of coefficients in the decomposed signal requires the limit at each scale to be larger than that for the original data to maintain equal false alarm rates. Different approaches may be used to determine the detection limits for the wavelet and scaling function coefficients and the reconstructed signal to obtain the same false alarm rate. For example, the wavelet and scaling function coefficients may be subjected to smaller confidence limits to be more sensitive to changes in the signal, followed by a higher confidence limit for the reconstructed signal. Alternatively, the confidence limits for the coefficients may be increased according to Equation 2.3 to account for the increase in the total number of coefficients [4].

$$C_L = 100 - \frac{1}{L+1}(100 - C) \tag{2.3}$$

Variable C is the desired confidence limit for the reconstructed signal, and C_L is the confidence limit that must be used for the coefficients at each scale for a decomposition of depth L. The autocorrelation in the coefficients requires further adjustment of the detection limits to obtain the desired in-control run lengths. This adjustment is similar to that used by SPC methods that filter the measurements such as, MA, EWMA, and CUSUM [28]. For the examples in this work, this adjusted detection limit is determined empirically.

2.3.4 Relation with Existing Methods

The characteristics of MSSPC can be appreciated by comparing the filters used by MSSPC with those used by existing SPC methods. The filters used by Shewhart, EWMA, MA and CUSUM charts, shown in Figure 2.10[29], indicate that these methods differ in the scale at which they filter the measurements, and the nature of the filter. Shewhart, MA, and CUSUM charts filter the data at increasingly coarse scales. The scale of the EWMA filter is determined by the value of the filter parameter, and that of the MA chart by the length of the filter window. These methods represent the measurements at a single fixed scale since the size of the filter does not vary, making them best for detecting changes that occur at a single scale. Consequently, these methods are best for detecting changes that span only one scale or range of temporal and frequency localization, since the filter does not adapt to the scale of the signal features. Thus, Shewhart charts are best for detecting large and localized changes, while CUSUM charts are best for detecting smaller coarse changes. Techniques based on combining multiple charts aim to improve the performance of control

charts over a broader range of scales. For example, combined Shewhart-CUSUM charts [9] represent the measurements at two scales, the finest and coarsest.

The filters available to MSSPC are of many different shapes, and include the filters used by existing methods. The filters available to MSSPC with Haar wavelets for a decomposition depth of $L = 3$ are shown in Figure 2.11. The filters are labeled by a 4-bit binary number to indicate the coefficients retained for reconstruction. For example, the combination 0100 corresponds to reconstruction after retaining only the wavelet coefficient at $m = 2$ and the combination 1101 corresponds to the selection of wavelet coefficients at scales $m = 1, 2$ and the scaling coefficient $\mathbf{a}_3$. Comparison with Figure 2.10 reveals that these MSSPC filters subsume the filters used by Shewhart, MA and CUSUM charts. For example, the combination 1111 corresponds to a Shewhart chart, the combination 0011 corresponds to an MA filter of window size 4, while the combination 0001 to a CUSUM chart for a signal of length 8. The use of smoother boundary corrected wavelets, such as Daubechies wavelets, approximates an EWMA filter as one of the possible cases [15]. Thus, MSSPC can automatically specialize to many existing methods depending on the nature of the measurements and abnormal features. The example in Figure 2.9 shows that when the shift is first detected at scale $t = 30$, the filter used to detect this change is 1000 in Figure 2.11. This filter adapts to 0011 at $t = 34$, which is a mean filter of length 4, and to 1001 at $t = 61$.

MSSPC with dyadic discretization is also related to the approach of decorrelating the measurements by taking their weighted or unweighted batch means [10]. Batch means control charts select the window size according to the nature of the autocorrelation, thus requiring a longer window with increasing correlation. Since the windows

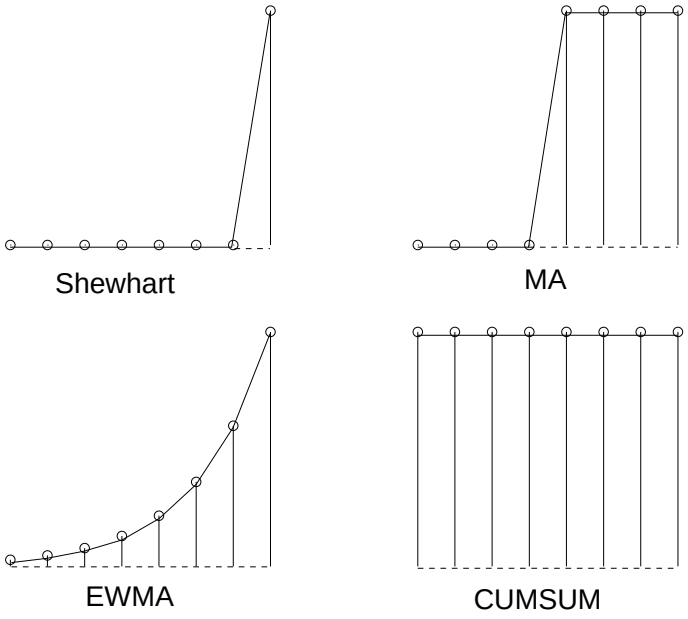


Figure 2.10: Commonly Used SPC Filters

do not overlap, this approach cannot detect a shift sooner than the window length. Unweighted batch means with a window of dyadic length is equivalent to the last scaled signal from a Haar decomposition at a depth equal to the base two logarithm of the batch means window length. Thus, MSSPC using Haar wavelets can subsume batch means of dyadic lengths. Unlike batch means, MSSPC also has the wavelet coefficients at finer scales available for detecting the shift sooner than a batch means control chart.

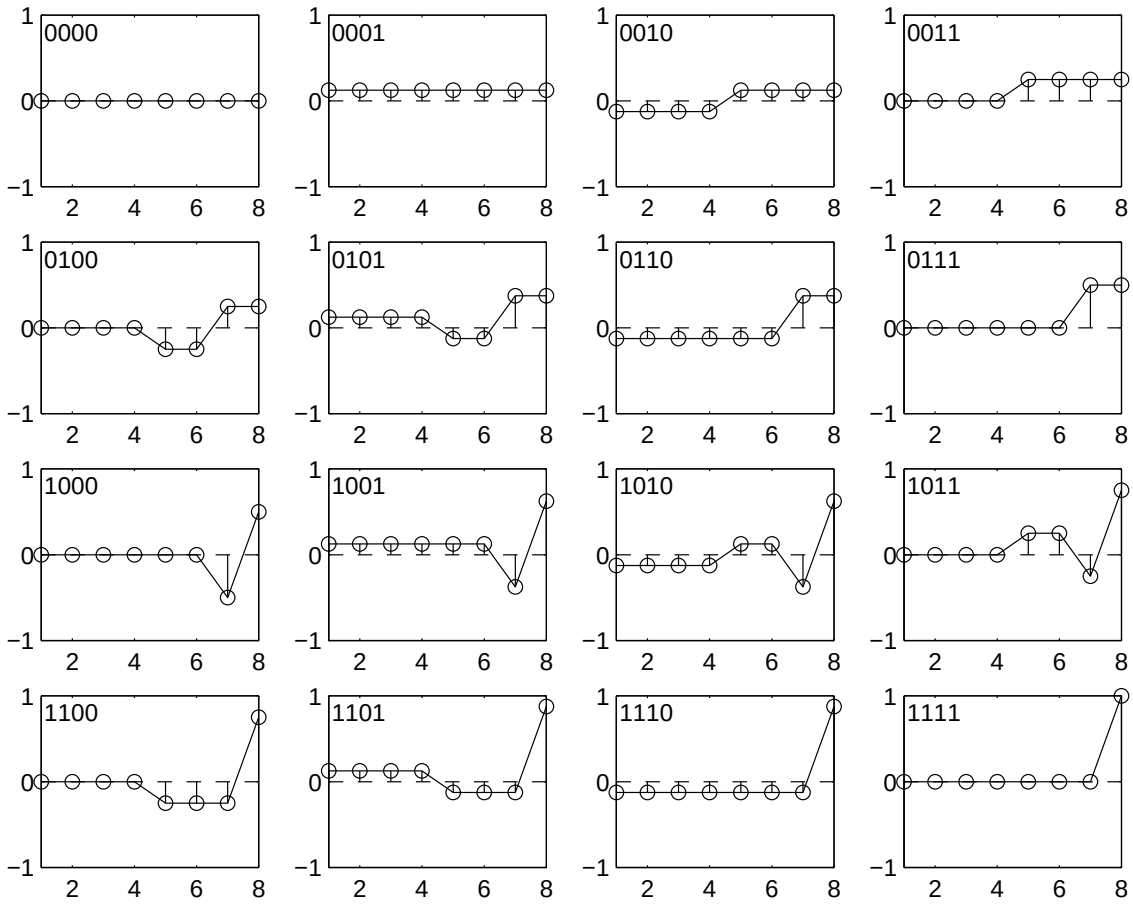
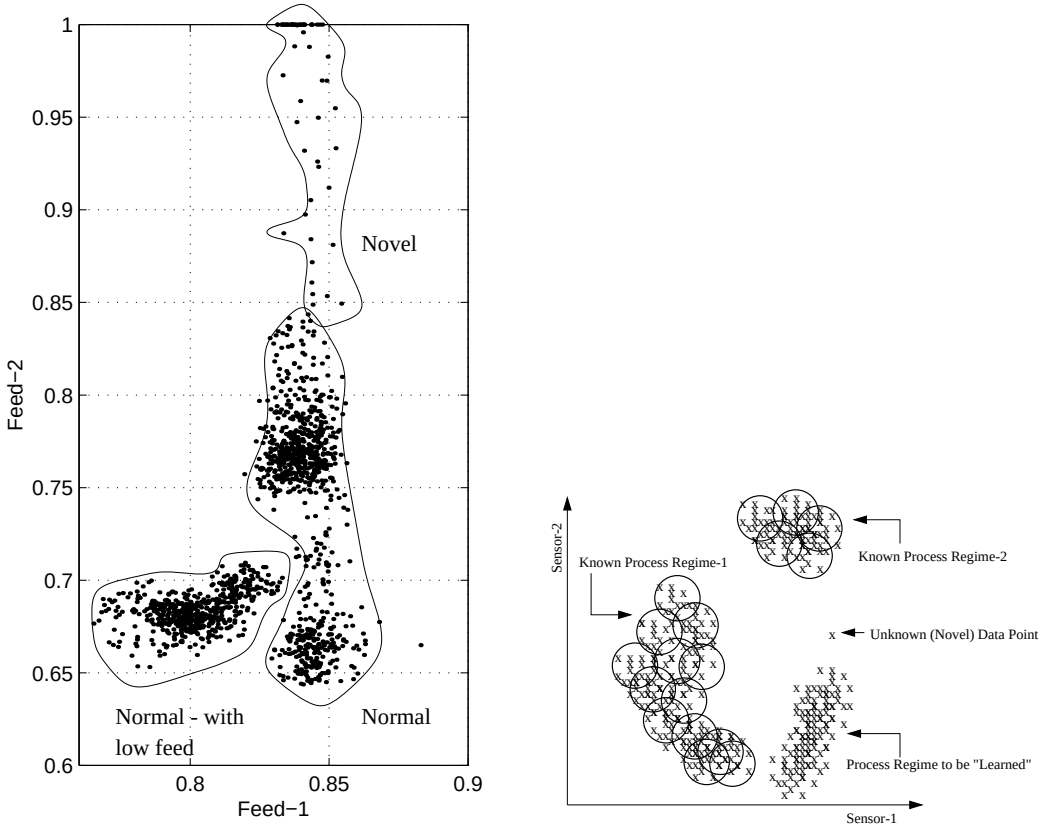


Figure 2.11: Filters available in MSSPC using Haar wavelets for $L = 3$

2.4 ART for Diagnosis

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 several possible paths depending on a large number of known and 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 2.12a). 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, offer several important characteristics for diagnosis of industrial processes [30][31]. Specifically, the ART family of networks [32][33][34][35][36] includes some of the few clustering algorithms that explicitly address the issue of stable adaptation and incremental learning with changing process behavior. This dissertation has made use of ART-2 for nonlinear diagnostic modeling in wavelet space because of these desirable properties of ART family of algorithms. This section presents a brief introduction for ART-2 based diagnosis. For a more detailed discussion on this topic, the reader is referred to the Ph.D. dissertation by Whiteley [37].

ART-2 is an unsupervised clustering mechanism proposed by Carpenter and Grossberg (Figure 2.12b) [32]. 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



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

Figure 2.12: Fault Detection for Industrial Processes

and system maintenance benefits. Later developments in the ART family of algorithms, such as ARTMAP and Fuzzy ARTMAP [33][34][35][36], 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 threshold 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., [31]). It has been shown to be able to work consistently well over a wide range of simulated as well as real-life process situations [38][39][30].

Typical real world processes under a continuous transition 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 that 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 [40][41][42], Hopke and co-workers [43], as well as Wang and co-workers [44], in addition to the previous work by our group at the Ohio State University [38] (Figure 2.12b).

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. [45] 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 [46].

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

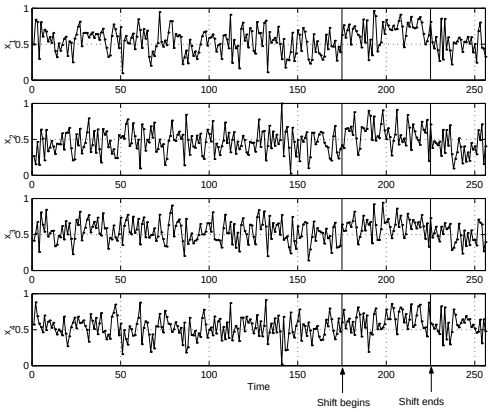
a multivariate, linearly correlated, noisy simulated process shown in Figure 2.13a. Abnormal operation was simulated as a mean shift added to all four variables from time-steps 176 through 225. Only normal data were used for training, so that the abnormal data were expected to be detected as an unknown event. Due to noise, however, we can see that normal and abnormal operations were not clearly separated. An ART-2 network was trained with independently generated normal data and was subjected to the test data. At each time step, the ART-2 similarity measure between the current four-dimensional data vector and stored cluster prototypes of normal data formed the basis for anomaly detection. For the given test data, the ART-2 similarity measures versus time are shown (Figure 2.13b-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. Such missed points can be misleading to an automated correction action and are undesirable. ART-2 diagnosis for such a noisy mapping was, thus, not robust. There was one false flag.

Several ART and ARTMAP variants have been proposed in the past to tackle this issue. The PROBART network proposed by Marriot and Harrison [46] 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 [47] was shown to approach Bayes optimal classification rates. The work by Srinivasa [48] proposed a PROBART variant that improved its generalization ability in the context of high noise. Gaussian ARTMAP by Williamson [49] combined a Gaussian

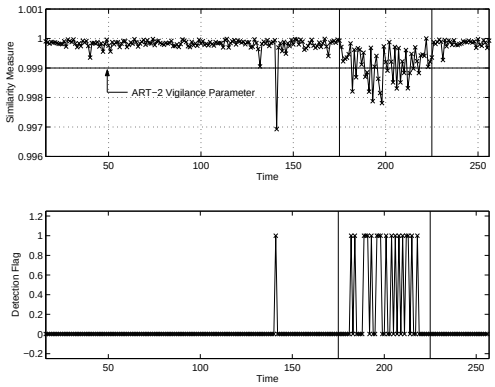
classifier and an ARTMAP network by appropriately changing the definitions of ART *choice* and *match* functions. Recently, Wang and co-workers [44] have proposed the use of wavelet feature extractors in place of the original data preprocessing and feature enhancement units within ART-2.

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. 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.

The use of several types of noise reduction filters, including wavelet-based filters, presents itself as a potential solution to the above noise vulnerability. However, 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. 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.



(a) A Simulated Noisy Data-set



(b) ART-2 Diagnosis Performance in a Noisy Environment

Figure 2.13: An Example to Illustrate the Noise Sensitivity of the ART-2 Detector

CHAPTER 3

MSSPC AND MSPCA - THEORETICAL ANALYSIS AND PROPERTIES

Data from most industrial processes contain contributions at multiple scales in time and frequency. In contrast, most existing univariate and multivariate SPC methods are best for detecting events at only one scale. As stated previously, Shewhart charts are best for detecting localized and large changes, while EWMA and CUSUM charts are best for detecting small changes at a coarse scale. A multiscale approach for SPC that can adapt to the scale of the relevant signal features has been developed based on wavelet analysis [4]. This dissertation presents the theoretical analysis of univariate and multivariate multiscale SPC (MSSPC), and compares its properties with existing SPC methods based on their average run length (ARL) analysis. This comparison shows that existing SPC methods are best for situations where the scale of the signal features that represent abnormal operation is known beforehand. If the nature of the abnormal features cannot be predicted a priori, MSSPC provides better average performance. MSSPC also performs well for monitoring autocorrelated measurements due to the ability of dyadic wavelets to approximately decorrelate most stochastic processes. MSSPC with dyadic discretization is appropriate for SPC of

highly autocorrelated or nonstationary stochastic processes. If the normal measurements are uncorrelated or contain only mild autocorrelation, it is better to use MSSPC with uniformly discretized wavelets. Many existing methods such as, MA, EWMA, CUSUM, Shewhart and batch means charts are shown to be special cases of MSSPC. Thus, MSSPC can specialize to these methods depending on the nature of the signal features. The ARL for MSSPC is determined via Monte-Carlo simulation as well as derived theoretically. The properties of MSSPC are also illustrated by application to univariate and multivariate SPC problems from a large scale petrochemical process.

3.1 Average Run Length Analysis of Univariate MSSPC

We have compared the performance of univariate MSSPC with existing univariate SPC methods for detecting a mean shift in uncorrelated and autocorrelated measurements. The average run length (ARL) [28] is the average number of samples required to detect a shift, and is determined in this dissertation both theoretically and by Monte-Carlo simulation. The theoretical derivation of ARL for wavelets with integer and dyadic discretization is presented in the Appendix. For Monte-Carlo ARL, different instances of test data are generated by simulating a shift of a given magnitude at a particular time. Time to detect this shift is recorded for each instance. This detection delay for each instance is referred to as its run length and an average of these run lengths over a large number of simulations is the Monte-Carlo ARL for the shift magnitude under consideration. The ARL value for a non-zero mean shift thus represents the detection probability. When plotted against the magnitude of the shift, the ARL curve is expected to be non-increasing and typically converges to the location of the mean shift as the magnitude of shift tends to infinity. When the magnitude of

shift is zero, the corresponding ARL value indicates the probability of false alarms, and is referred to as the in-control run length. For the same in-control run length, it is desirable to have lowest possible ARL values for non-zero mean shifts. For the examples in this dissertation, the mean shift is located at the first measurement. The Monte-Carlo simulation results are generated from at least one thousand realizations.

3.1.1 A Uncorrelated Gaussian Process

Suppose that the measurements are univariate IID Gaussian with unit variance,

$$x(t) = N(0, 1) \tag{3.1}$$

where $N(0, 1)$ is a Gaussian random number with zero mean and unit variance. The performance of Shewhart and MA charts is compared with MSSPC with dyadic or integer discretization. In each case, the parameters are selected to maintain approximately equal in-control run lengths. MA charts are selected for comparison with MSSPC using Haar wavelets, since both methods use filters of the same shape.

For *dyadically* discretized wavelets, the ARL curves obtained from theory and simulation for different depths of decomposition are shown in Figure 3.1. The ARL for a Shewhart chart and a moving average chart with window of size 16 are also shown in Figure 3.1 for comparison. The theoretical run lengths are determined based on the assumption that the downsampled coefficients are uncorrelated and Gaussian with variance equal to that of the original measurements. The close match between the theoretical and simulated curves confirms the validity of these assumptions. Detailed derivation of the theoretical approach is provided in Appendix A.1.

With increasing depth of the wavelet decomposition, the ability of MSSPC to detect large shifts deteriorates due to the increasing time delay in obtaining the

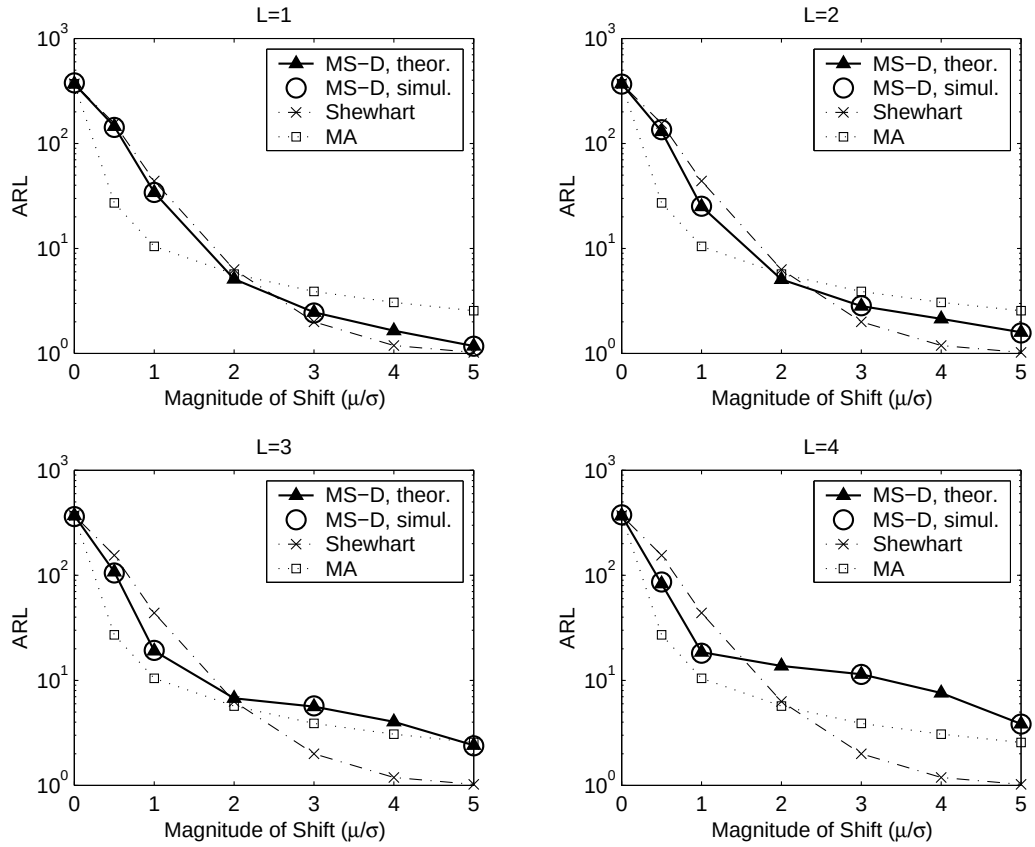


Figure 3.1: ARL of MSSPC-Dyadic simulated, MSSPC-Dyadic theoretical, Shewhart, and MA charts applied to univariate IID Gaussian Process for different depths of decomposition.

wavelet and scaling function coefficients at coarser scales. The ability to detect small shifts improves at greater depths due to greater separation between the stochastic variation and deterministic mean shift at coarser scales. Depending on the selected depth of decomposition, the performance of MSSPC tends to be better than that of Shewhart charts for small shifts, and better than MA charts for large shifts. For higher depths ($L = 3, 4$), the delay introduced due to dyadic downsampling adversely affects the detection performance for large shifts.

The ARL plots for *integer* discretization shown in Figure 3.2 indicate better performance than that with dyadic discretization for detecting large shifts (Figure 3.1). As expected, the improvement is particularly significant for larger depths of decomposition. For example, for $L = 4$, MSSPC-Integer ARLs match the MA ARLs for small shifts and the Shewhart ARLs for larger shifts. This observation indicates that MSSPC-Integer successfully selected the finer scales for larger shifts and the coarser scales for smaller shifts, *without any previous knowledge of the magnitude of shift.* Its performance is thus robust overall for a range of shifts.

The theoretically computed ARL for MSSPC with integer discretization is shown in Figure 3.3 for depth of $L = 1$, and indicates a good match with the empirical values. The theoretical analysis of Appendix A.1 cannot be accurately applied to MSSPC with integer discretization since the assumption of uncorrelated coefficients is violated. However, this correlation can be modeled as a Markov chain, and a theoretical method for ARL may be derived as shown in Appendix A.2. This approach uses numerical integration to compute the probability of detection in terms of the magnitude of the shift and the number of time steps passed since the introduction of the shift. These probabilities are then used to derive the value of the average run length. Detailed derivation of ARL values for a two-scale wavelet decomposition with Haar wavelets for this signal is provided in Appendix A.2. The procedure is general and can be extended to multiple scales and other wavelets. The computation cost for numerical integration, however, increases exponentially with the depth of wavelet decomposition.

Since the number of coefficients from the uniformly discretized wavelet decomposition is more than the number of samples in the original signal, the confidence

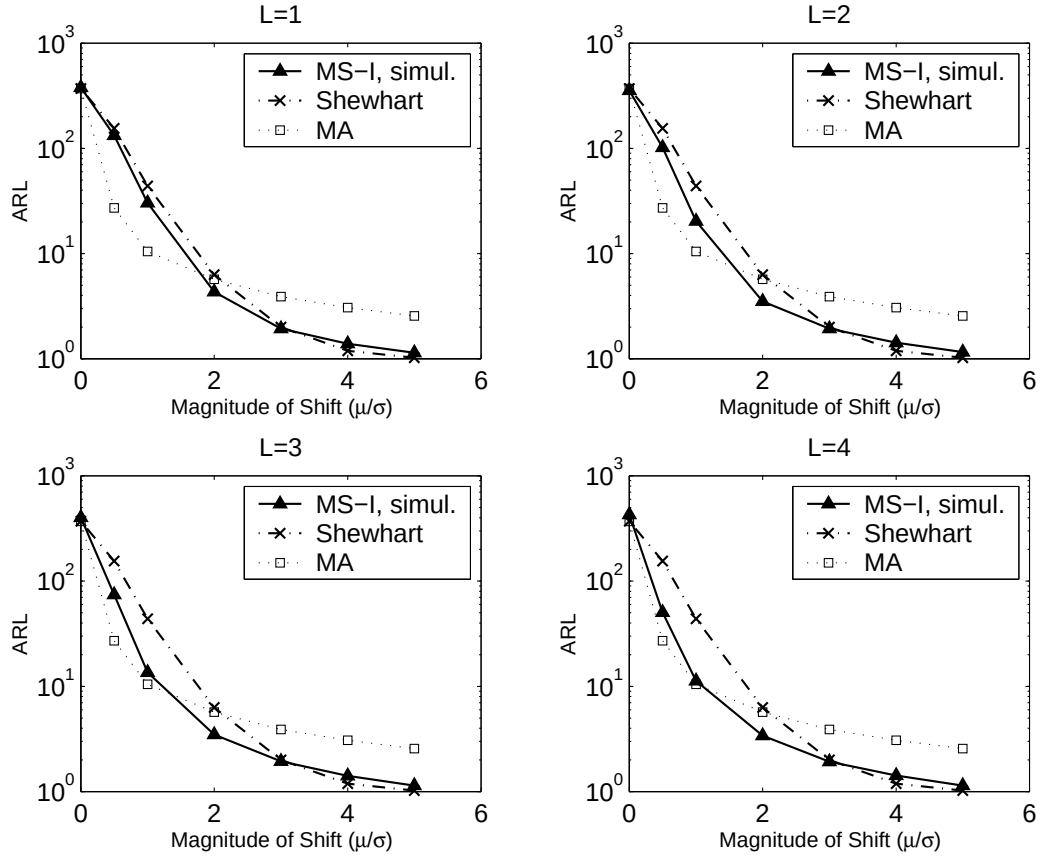


Figure 3.2: ARL of MSSPC-Integer simulated, Shewhart, and MA charts applied to univariate IID Gaussian Process for different depths of decomposition.

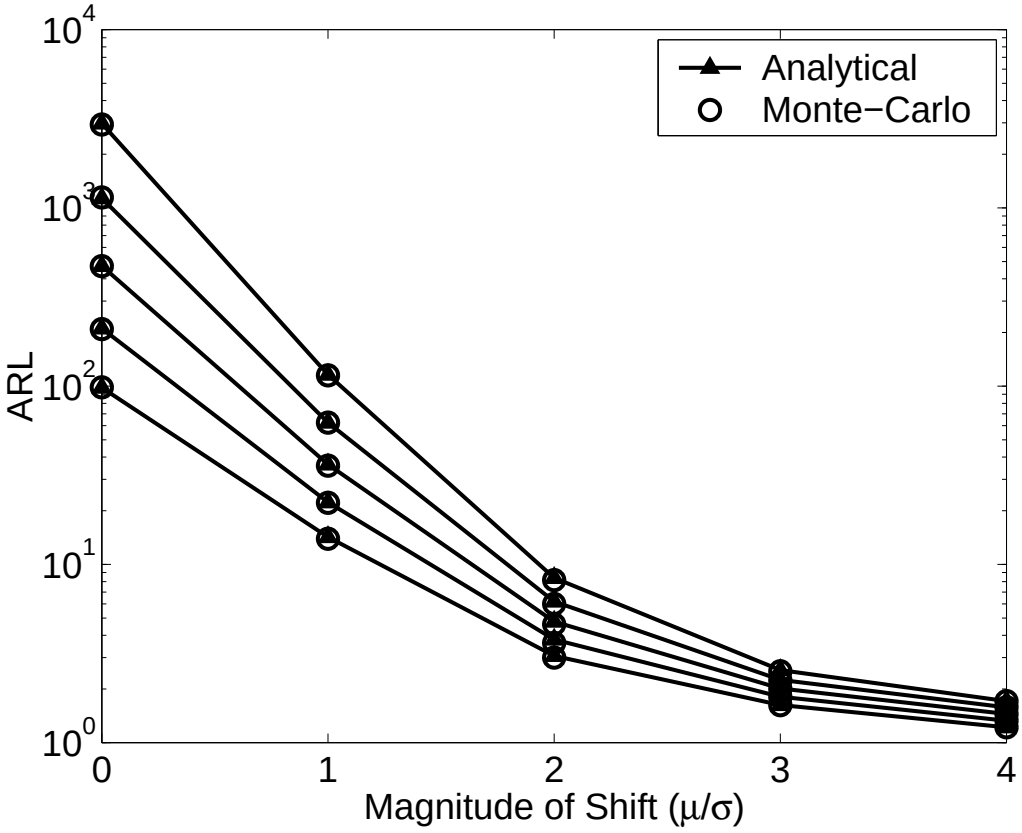


Figure 3.3: Comparison of theoretical and simulated ARL values for different detection thresholds with MSSPC-Integer applied to IID Gaussian Process

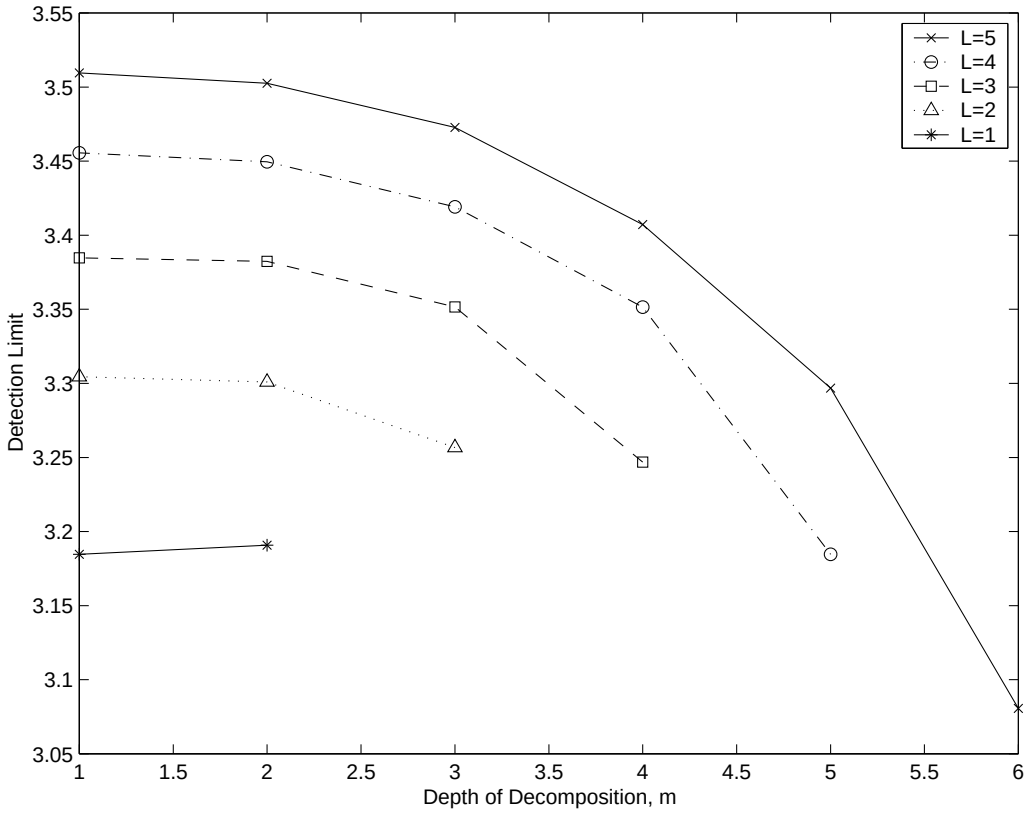


Figure 3.4: Adjusted detection limits to obtain confidence, C_L computed by Equation 2.3, for different decomposition depths. Limit at largest m for each curve is for the last scaled signal.

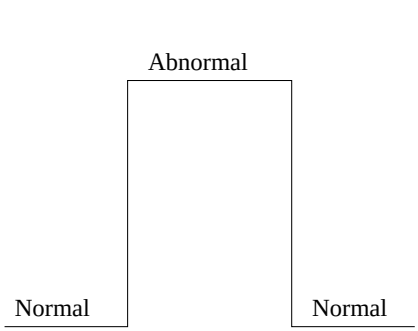
limit at each scale is increased by using Equation 2.3. The lack of downsampling also results in autocorrelated coefficients at each scale, requiring further adjustment in the limits to obtain the desired in-control run lengths. The adjusted limits are determined by simulating the run length for different limits at each scale. The limit corresponding to the desired run length is then determined by interpolation between the simulated values. The resulting limits at each scale for different decomposition depths are shown in Figure 3.4.

Comparison of the ARL curves for MSSPC-dyadic, MSSPC-integer, MA, and Shewhart charts in Figures 3.1 and 3.2 indicates that no single method performs best for all shift sizes. Thus, if the objective of SPC is to detect only small shifts, it is best to use a MA control chart. Similarly, if the objective is to detect only large shifts, it is best to use a Shewhart chart. Other methods such as CUSCORE charts [50], can be tailored to detect specific changes whose nature is known beforehand. In most practical problems, it is impossible to predict the type or extent of change due to abnormal operation. For such problems, MSSPC is more appropriate due to its ability to adapt for detecting a wide variety of shifts and signal features.

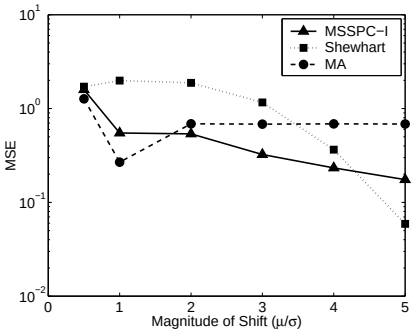
This general nature of MSSPC is illustrated in Figure 3.5 which shows the mean-square error (MSE) for classifying a mean shift of different sizes. The shift persists for 50 samples, and is surrounded on both sides by 50 samples each of normal data, as shown in Figure 3.5a. Similar to the ARL analysis, the MSE analysis from Figure 3.5b shows that the average MSE values for all shifts in Figure 3.5 are smallest for MSSPC-I over the range of mean shifts considered. Thus the MSSPC approach is seen to be robust over a wide range of shift magnitudes.

3.1.2 A Stationary Autocorrelated Process

Many methods have been devised for SPC of autocorrelated measurements [28]. If a time-series model of the measurements is available, it can be used to decorrelate the measurements. The residuals will be uncorrelated, and may be monitored by existing SPC methods [51]. Since time-series models may not be readily available, and are often not practical to use, other methods that do not require an explicit time-series model of the measurements have been developed for decorrelating the measurements.



(a) Nature of Change



(b) Mean Squared Error in Classification

Figure 3.5: Mean-square error for classification by MSSPC-Integer, Shewhart, and MA charts for IID Gaussian process with mean shift. Total MSE for all shifts is 7.1722×10^{-3} for Shewhart, 4.2808×10^{-3} for MA, and 3.4099×10^{-3} for MSSPC-I.

These methods include batch means control charts [10] and moving center line EWMA (MCEWMA) [11]. Batch means control charts decorrelate the data by taking the average of the measurements in non-overlapping windows. The window size is selected such that the means in each window are uncorrelated. Thus, more correlation will require a longer window size. Weighted batch means (WBM) determine the weights based on knowledge of the autocorrelation, while unweighted batch means use equal weights regardless of the type of correlation. A significant limitation of this approach is that it cannot detect a shift sooner than the window length. In addition, the window size may have to be determined empirically. The MSSPC approach with dyadic discretization takes a weighted mean of the measurements, but does better than the batch means approach due to the multiscale nature. MCEWMA fits an EWMA to the measurements to minimize the one-step ahead prediction error, and is ideally suited for decorrelating IMA stochastic processes. It has also been applied to

other types of autocorrelated processes [11]. As illustrated by the examples in this section, MCEWMA does not perform as well as MSSPC for other types of stochastic processes.

The ARLs for the following AR(1) process,

$$x(t) = 0.5x(t-1) + \epsilon(t) \tag{3.2}$$

for different depths are shown in Figure 3.6 based on Monte-Carlo simulation and the derivation presented in Appendix A.1. The match between the analytical and Monte-Carlo results is not as good as that for MSSPC of uncorrelated data shown in Figure 3.1. This is because the assumption of uncorrelated measurements at each scale is not as valid. Furthermore, the match between theory and experiment deteriorates with increasing depth, since the assumption of uncorrelated coefficients is less accurate at coarser scales. Figure 3.6 also shows the ARL of SPC residuals and weighted batch means charts [10]. These results indicate that WBM does well only for small shifts, while residual charts work well only for very large shifts. Across a wide range of shifts, MSSPC with dyadic discretization works quite well, particularly for small L.

Figure 3.7 depicts the ARL for an AR(1) process given by,

$$x(t) = 0.9x(t-1) + \epsilon(t) \tag{3.3}$$

The high degree of autocorrelation in this stochastic process makes it more difficult to detect shifts. This figure compares the ARL of MSSPC-D with that of WBM, and MCEWMA control charts. WBM does well for detecting small shifts, but deteriorates quickly for larger shifts. MCEWMA does better than WBM only for the largest shift. MSSPC-D does worse than WBM for small shifts, but does significantly better

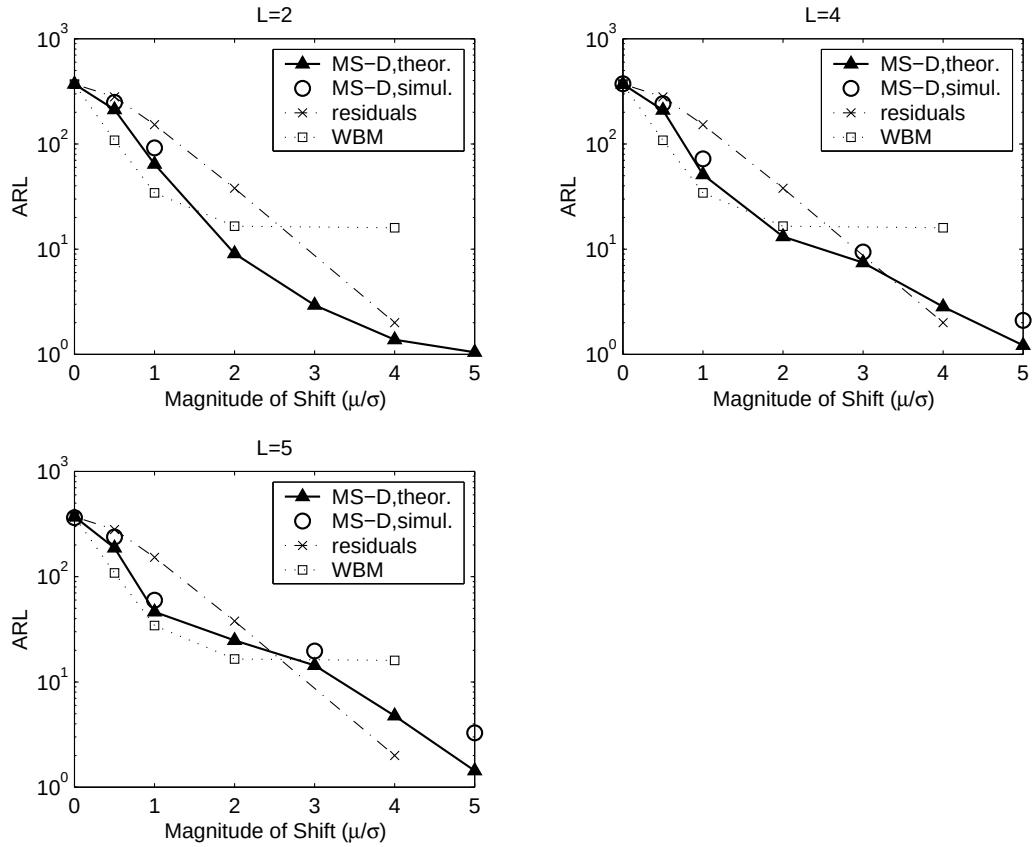


Figure 3.6: ARL curves for MSSPC-Dyadic simulated, MSSPC-Dyadic theoretical, residuals, and weighted batch means charts for AR(1) process given by Equation 3.3

than the other two approaches for larger shift sizes. The time delay due to downsampling is made up to a certain extent by the easier detection of shifts in uncorrelated coefficients.

3.1.3 A Non-stationary Process

Nonstationary stochastic processes present special challenges for SPC, since their mean tends to change over time. The ARL performance of MSSPC and MCEWMA

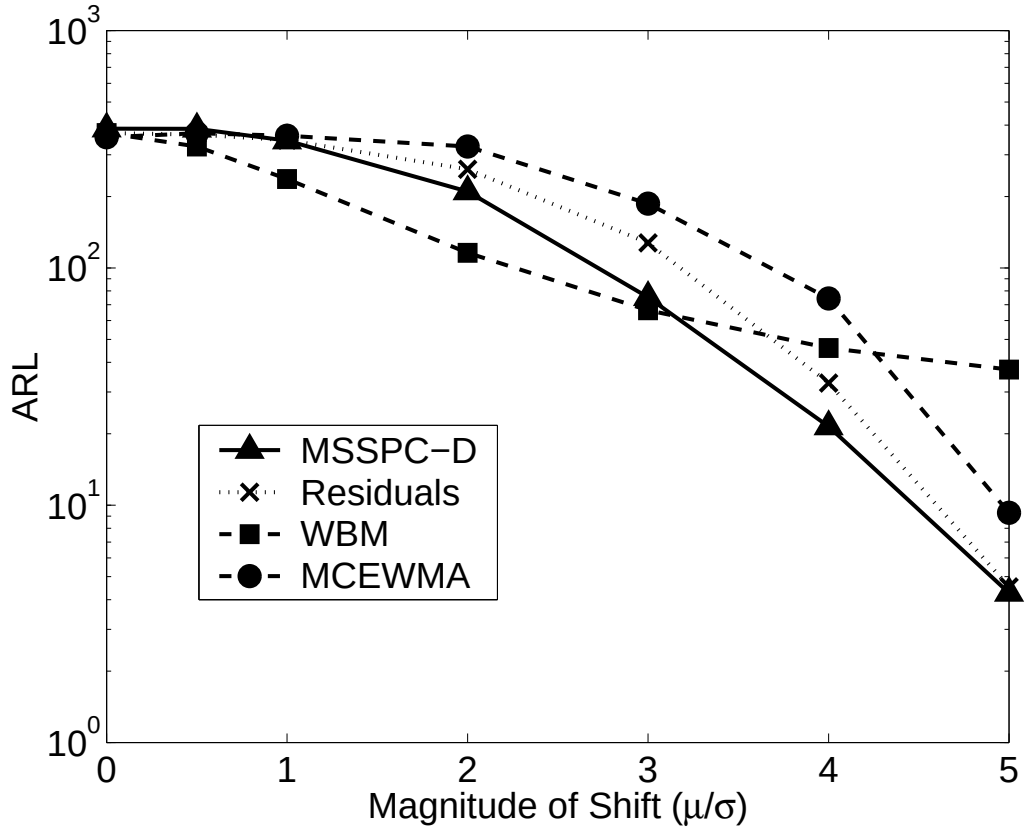


Figure 3.7: ARL curves for MSSPC-Dyadic, WBM, and MCEWMA charts for an AR(1) process given by Equation 3.3

are compared in Figure 3.8. In this case, the stochastic process is IMA(1,1) given by,

$$x(t) = x(t-1) + \epsilon(t) - 0.5\epsilon(t-1) \tag{3.4}$$

which can be modeled optimally by EWMA. Using wavelets without downsampling is not feasible for SPC of such nonstationary measurements, since the high autocorrelation in the non-downsampled wavelet coefficients increases the rate of false alarms for the same fault detection ability.

Figure 3.8 shows that for small shifts, the performance of MCEWMA and MSSPC-D are equivalent, but MCEWMA performs better for detecting large shifts. The superiority of MCEWMA for this type of stochastic process is due to it being the optimal approach for decorrelating an IMA(1,1) time series. A mean shift in an IMA(1,1) process appears as a spike of very small duration in the decorrelated residuals [52]. The MSSPC-D approach may easily miss this spike, even when it is large, due to the downsampling of the wavelet coefficients. Furthermore, the mean shift cannot be detected in the coefficients of the last scaled signal due to the residual autocorrelation and extremely large detection limits. The extremely localized nature of the spike indicates that a Shewhart chart on uncorrelated data should work best, which is essentially the approach used by MCEWMA.

3.2 Average Run Length Analysis of Multivariate MSSPC

The general framework of MSSPC shown in Figure 2.6 may be used to transform any existing multivariate SPC method to a multiscale approach. This section focuses on multivariate SPC using principal component analysis (PCA). PCA has been popular for process monitoring since it allows extension of the principles of univariate statistical process monitoring to multivariate processes by capturing a large number

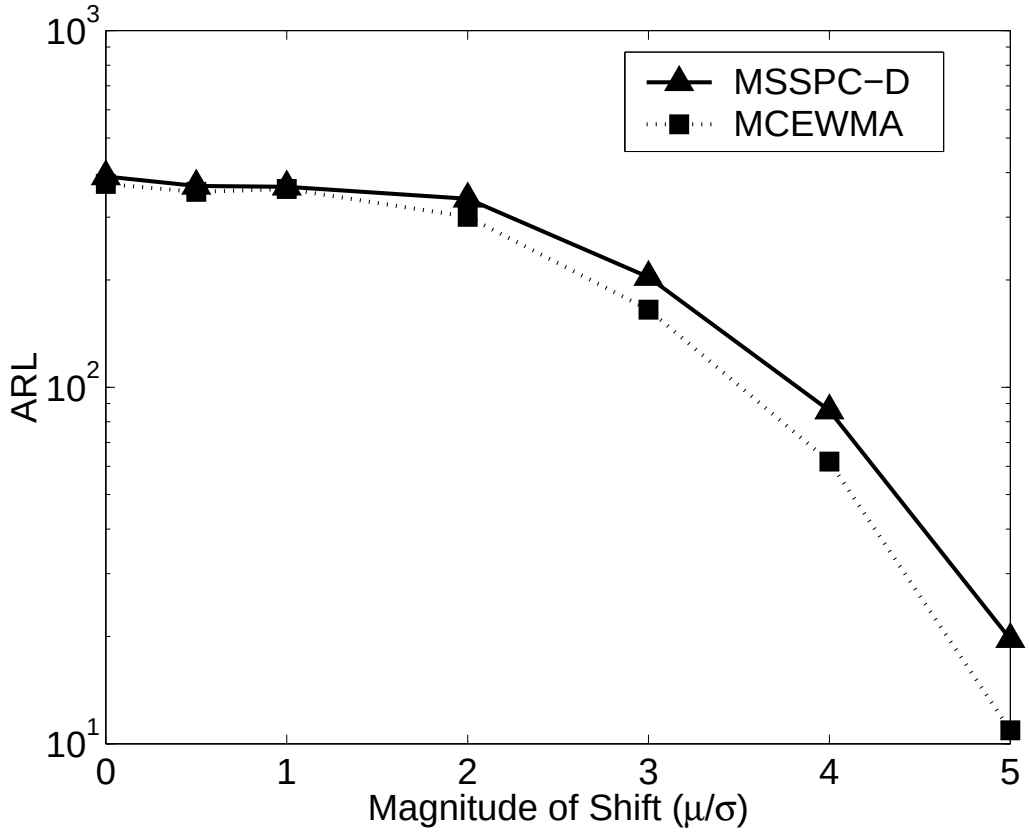


Figure 3.8: ARL of MSSPC-Dyadic and MCEWMA for an IMA(1,1) process given by Equation 3.4

of related measured variables in a small number of uncorrelated principal component scores [53][5]. The T^2 plot monitors the space of selected principal components, whereas the Q plot monitors the residual space. The detection limits are based on the assumption of IID Gaussian data, or are found empirically.

Multivariate monitoring by MSPCA applies monitoring by conventional PCA to the coefficient matrix at each scale. Thus, each column of the data matrix is decomposed by the selected wavelet, and T^2 and Q plots are developed for the coefficients at each scale. Since the wavelet transform is a linear operation, MSPCA preserves the modeling qualities of PCA and does not affect the eigenvectors or eigenvalues of the original matrix.

Many extensions and variations of PCA have been developed for dealing with various practical situations. A common approach for monitoring autocorrelated measurements is by dynamic PCA (DPCA) [6][7]. DPCA decorrelates the measurements by augmenting the data matrix by lagged variables. Thus, PCA of the augmented matrix implicitly finds the time-series model between the lagged and other variables. Multivariate EWMA has also been applied to the selected principal component scores to benefit from the dimensionality reduction and to improve the detection of small shifts [54]. A rigorous ARL analysis of PCA and its many variations is usually not available. This section studies the theoretical properties of PCA, DPCA, and MSPCA based on simulated uncorrelated and autocorrelated data.

3.2.1 Uncorrelated Measurements

This example considers the linear uncorrelated multivariate process modeled as follows [4],

$$x_1(t) = N(0,1) \tag{3.5}$$

$$x_2(t) = N(0,1) \tag{3.6}$$

$$x_3(t) = \frac{x_1(t) + x_2(t)}{\sqrt{2}} \tag{3.7}$$

$$x_4(t) = \frac{x_1(t) - x_2(t)}{\sqrt{2}} \tag{3.8}$$

$$\mathbf{Y}(t) = \begin{bmatrix} x_1(t) \\ x_2(t) \\ x_3(t) \\ x_4(t) \end{bmatrix} + 0.2\epsilon(t) \tag{3.9}$$

The data matrix, $\mathbf{Y}$, for normal operation is generated by Equation 3.9, while that for abnormal operation is generated by adding a mean shift of various sizes to all the variables. Both, PCA and MSPCA select two principal components for the T^2 plots. A run is terminated if there is violation of the limits in either of the T^2 or Q plots. The ARL curves for both methods shown in Figure 3.9 indicate that the behavior of PCA and MSPCA-I is analogous to that of Shewhart chart and MSSPC-I shown in Figure 3.2. Similar to the univariate uncorrelated case, MSPCA shows improvement over PCA for this multivariate process.

3.2.2 Autocorrelated Measurements

This section presents ARL analysis of the following multivariate time-series model [6],

$$z(t) = \begin{bmatrix} .118 & -.191 \\ .847 & .264 \end{bmatrix} z(t-1) +$$

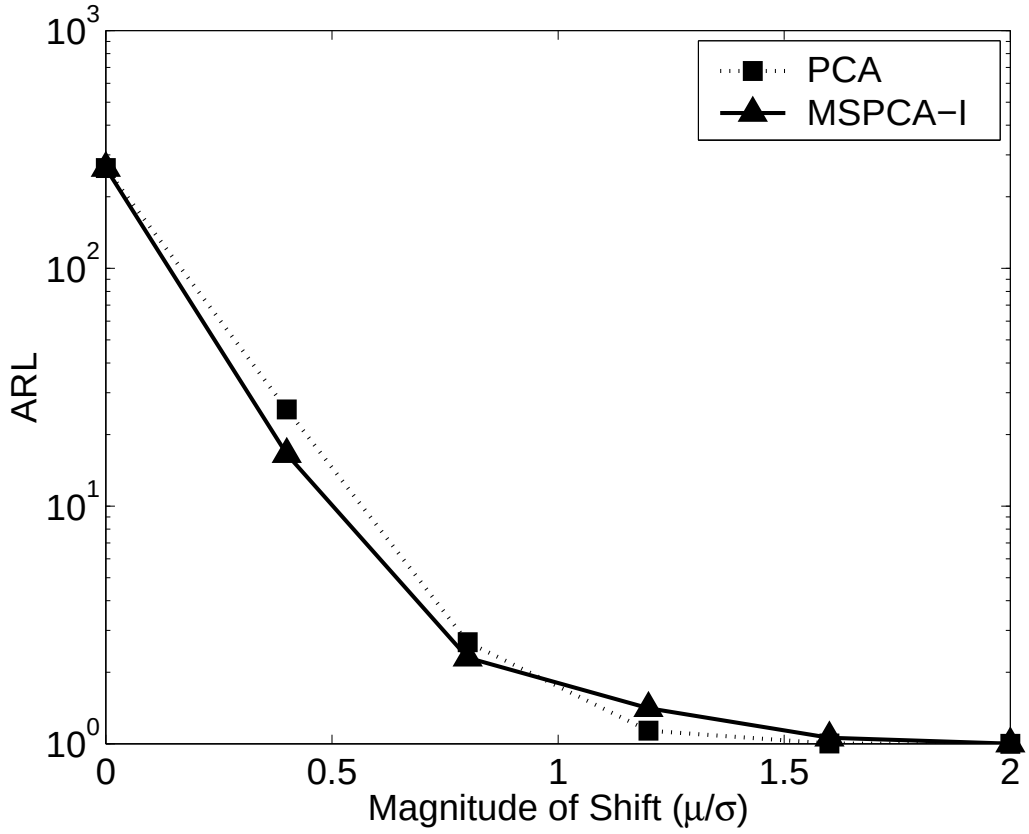


Figure 3.9: ARL of PCA and MSPCA-Integer for a multivariate uncorrelated process given by Equation 3.9

$$\begin{bmatrix} 1 & 2 \\ 3 & -4 \end{bmatrix} u(t-1) \tag{3.10}$$

$$y(t) = z(t) + v(t) \tag{3.11}$$

$$u(t) = \begin{bmatrix} .811 & -.226 \\ .477 & .415 \end{bmatrix} u(t-1) + \begin{bmatrix} .193 & .689 \\ -.320 & -.749 \end{bmatrix} w(t-1) \tag{3.12}$$

The data matrix for steady-state PCA is $[u(t)\, y(t)]$, while that for dynamic PCA is $[u(t)\, u(t-1)\, y(t)\, y(t-1)]$. Following Ku et al., two principal components are selected for steady-state PCA, and seven for DPCA. Data representing abnormal operation is generated by introducing a shift in u. The ARL analysis for this process using MSPCA-D is shown in Figures 3.10 and using MSPCA-I in Figure 3.11. Both figures include ARL plots for PCA and DPCA. For all methods, a run is terminated when either the T^2 or Q plot violates its detection limit.

These figures indicate that DPCA does perform better than PCA, particularly for detecting large shifts in the mean. MSSPC with dyadic discretization using both PCA or DPCA does well for detecting small shifts, but not as well as PCA or DPCA for detecting large shifts. Its performance for detecting large shifts deteriorates due to the delay introduced by downsampling. These results are comparable to those for univariate MSSPC by Shewhart chart and MSSPC-D shown in Figure 3.1. In contrast, MSSPC with integer discretization based on PCA and DPCA does better for all shifts than their single-scale counterparts. The best performance is provided by MSDPCA-I since it benefits from the decorrelation by DPCA and quicker detection by the multiscale approach.

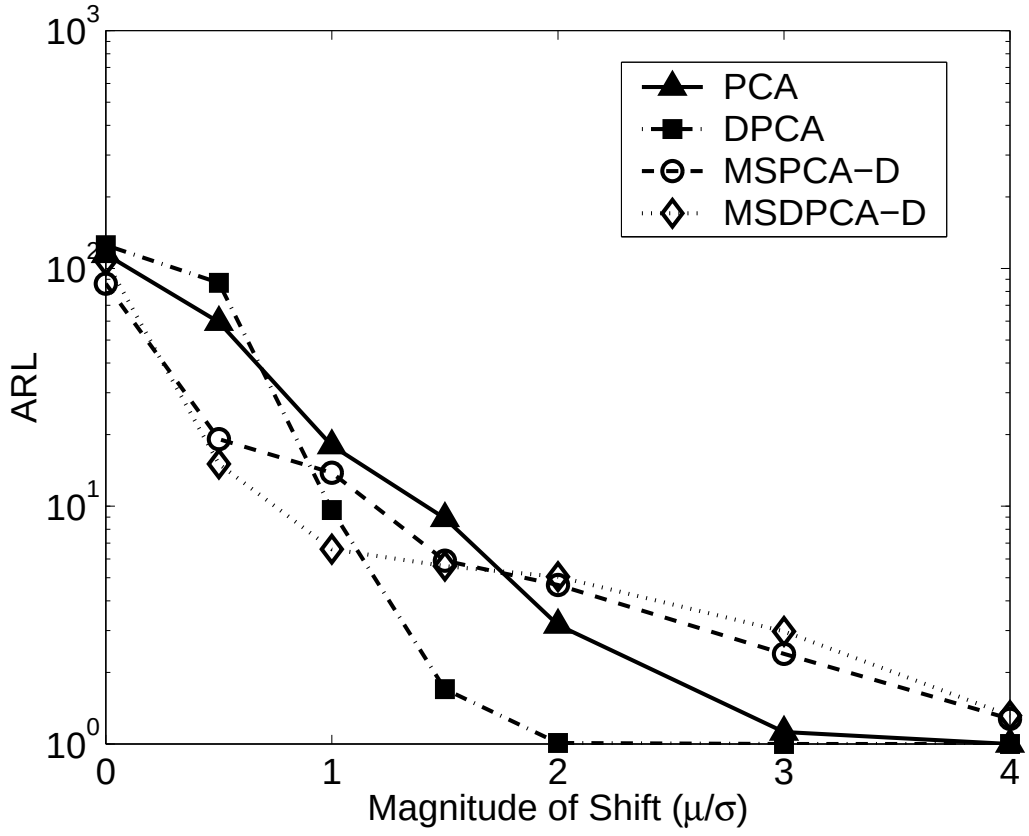


Figure 3.10: ARL of PCA, DPCA, MSPCA-Dyadic and MSDPCA-Dyadic for a multivariate correlated time series given by Equations 3.10-3.12

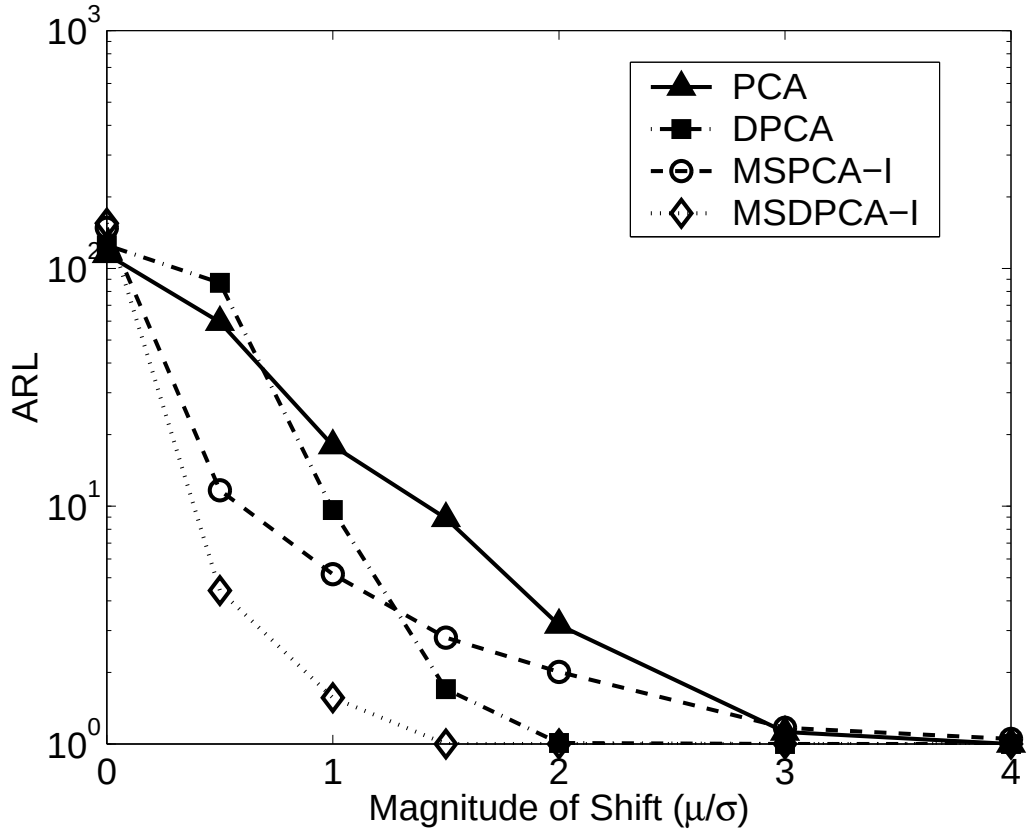


Figure 3.11: ARL of PCA, DPCA, MSPCA-Integer and MSDPCA-Integer for a multivariate correlated time series given by Equations 3.10-3.12

3.3 Industrial Case Studies

The SPC methods studied in this dissertation are applied here to univariate and multivariate data obtained from a petrochemical process. The data are provided by the Abnormal Situation Management Consortium.

3.3.1 Univariate Processes

The univariate measurements analyzed in this section represent a malfunction in a charge gas drier unit and a sensor malfunction due to oil accumulation. Both events exhibit very small levels of noise relative to the magnitudes of the events under consideration. This facilitates localizing the onset and end times for faulty behavior by operator annotation and manual inspection of data. Unlike the previous simulated examples, this industrial data-set contains different types of changes in the measurements. As illustrated by these examples, actual operating data usually contain changes of different sizes and shapes, making them ideally suited for MSSPC.

Charge Gas Drier Cooling

Measurements under normal and drier cooling conditions are shown in Figure 2.1. Since the normal data has far less variation than the abnormal measurements, it is possible to identify the onset and end of the drier cooling event at 100 and 666 time units respectively, as shown in Figure 3.12. The thresholds for the Shewhart, MA, and MSSPC charts were determined empirically to obtain approximately the same rate of false alarms of 1% on the normal data. The large sudden change at the start of the abnormal operation is ideally suited to detection by a Shewhart chart. As shown in Figure 3.13, a Shewhart chart performs slightly better than MSSPC which does better

than MA for identifying the starting point of the fault. In contrast, the slow return to normal operation is more difficult for the Shewhart chart to detect. As shown in Figure 3.13, MSSPC is the best method for detecting the return to normal operation, with the Shewhart chart providing the worst performance. Since the return is a slow change occurring at a coarse scale, it is difficult to detect only from the magnitude of the measurements, but is easier to detect based on the slope of the signal. MSSPC is particularly good for detecting such changes since the wavelet coefficients correspond to the local slope of the measurements, and capture the behavior at multiple scales. Consequently, MSSPC is the only method that detects abnormal operation between samples at 655 and 666.

Sensor Malfunction due to Oil Accumulation.

This example analyzes a sensor failure due to oil accumulation. The difference between the faulty sensor and a coupled redundant sensor is shown in Figure 3.14. The abnormal event is indicated by a change in the frequency and amplitude of the measurements. The small oscillations make it difficult to detect the beginning of this fault by any of the methods used in this dissertation. Once again, the rate of false alarms was maintained equal for all methods to approximately 1%. As illustrated in Figure 3.15, the MA chart results in the worst performance for detecting the onset and end of the event. MSSPC performs slightly better than Shewhart charts.

3.3.2 A Multivariate Process

PCA and MSPCA were used to detect a disturbance in furnace feeds. Normalized sensor readings for 10 furnace feed sensors used as test data are presented in Figure 3.16. The event began at approximately 108^{th} time-step. The detection performance

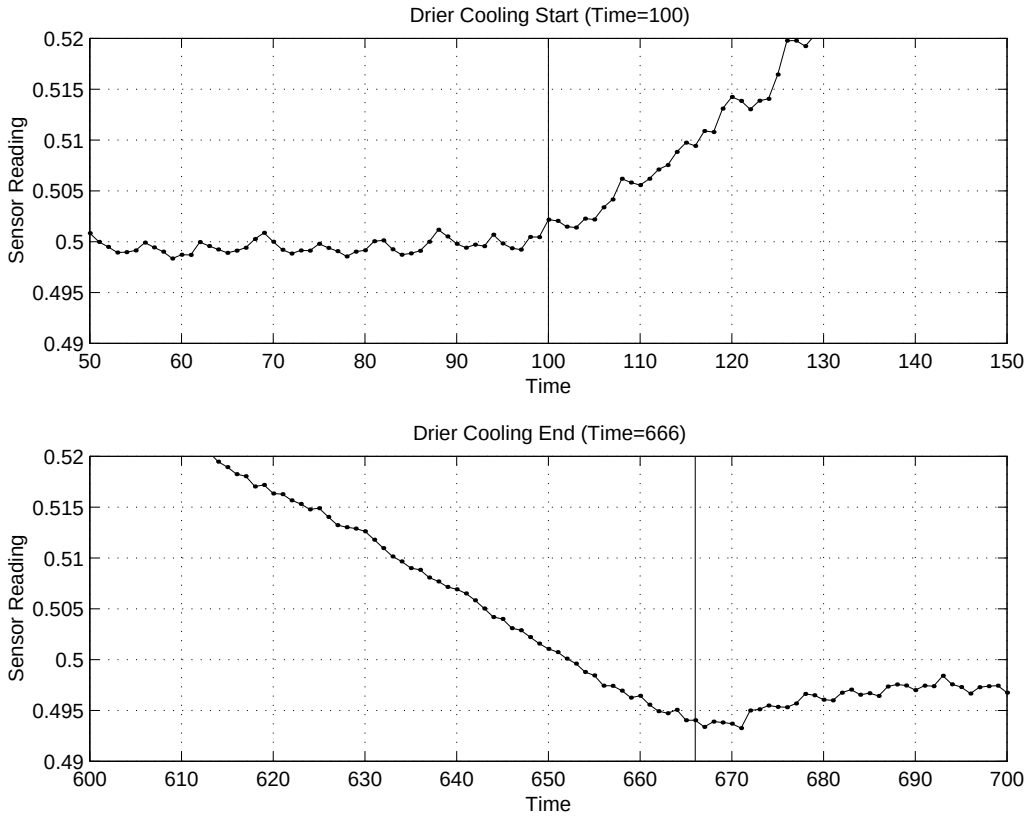


Figure 3.12: Onset and End of the Drier Cooling Event

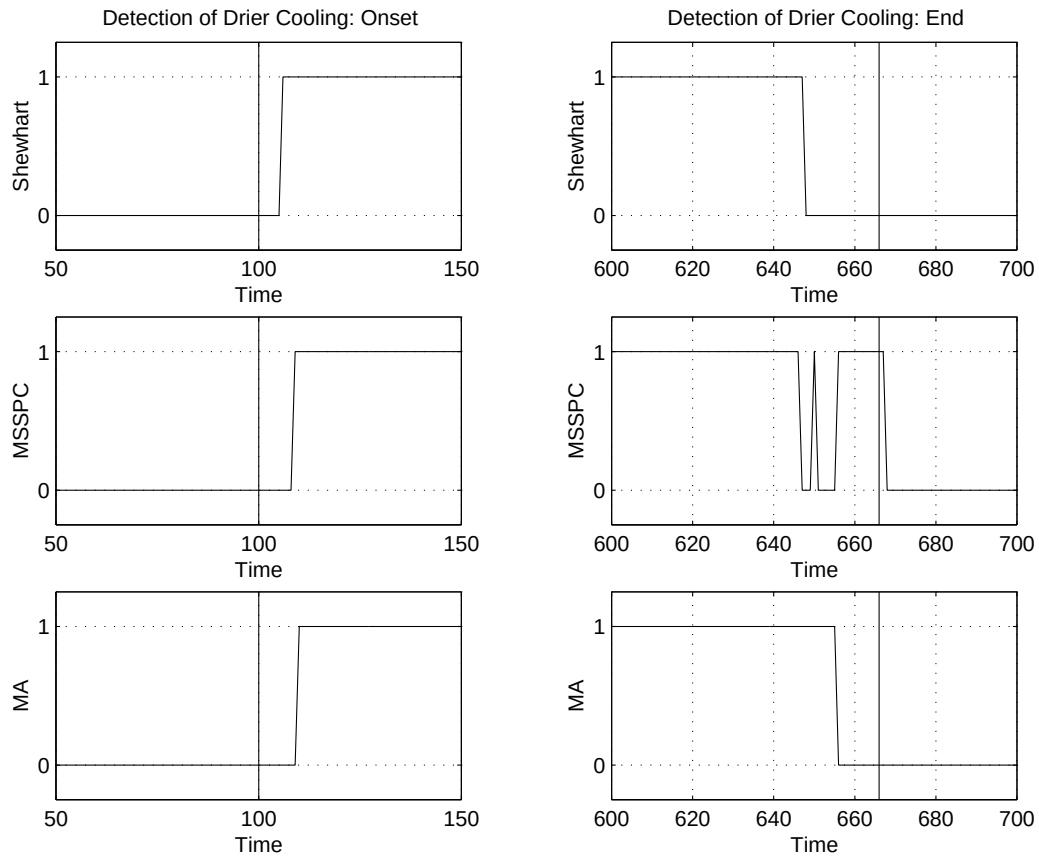


Figure 3.13: Detection of the start and end of Drier Cooling Event by Shewhart, MSSPC-Integer, and MA charts

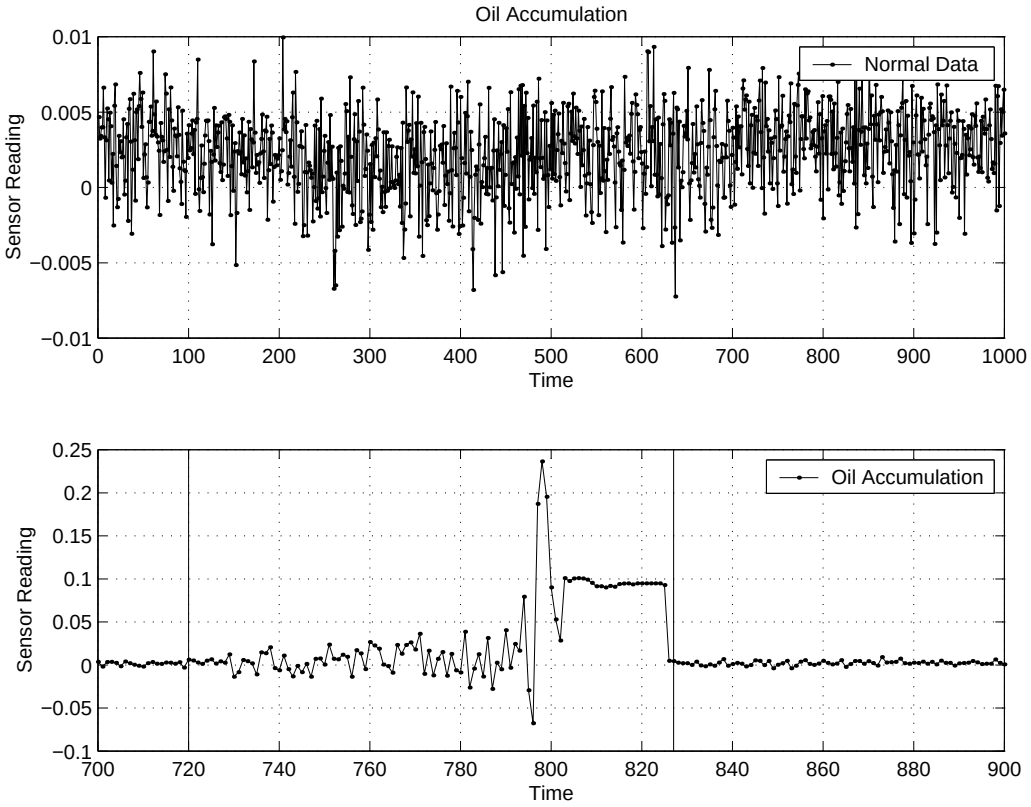


Figure 3.14: Data for Normal Operation and the Oil Accumulation Event

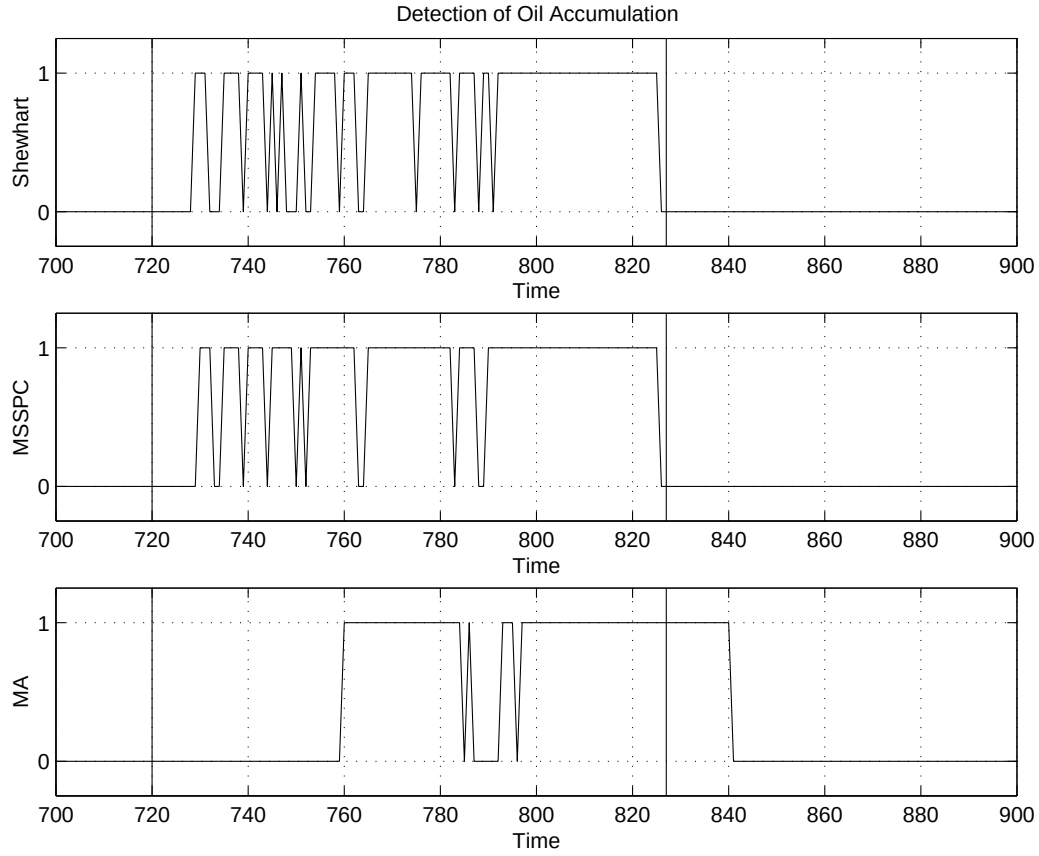


Figure 3.15: Performance of Shewhart, MSSPC-Integer, and MA charts for detection of Oil Accumulation Event

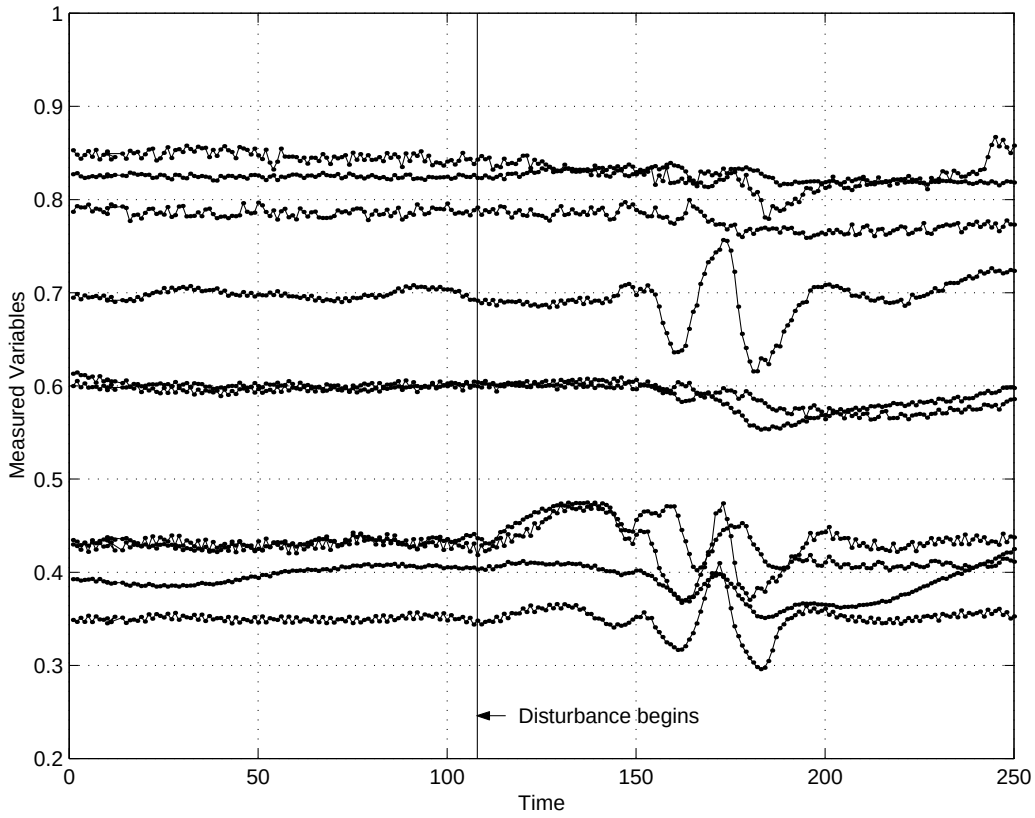


Figure 3.16: Test Data for Furnace Feed Disturbance Detection

presented in Figure 3.17 shows that MSPCA detects the event slightly earlier than PCA and continues to detect it with its Q value well above the detection limits. On the other hand, PCA misses a few genuine alarms at around time-step 150 in both, T^2 and Q plots, while MSPCA does not in its Q plot.

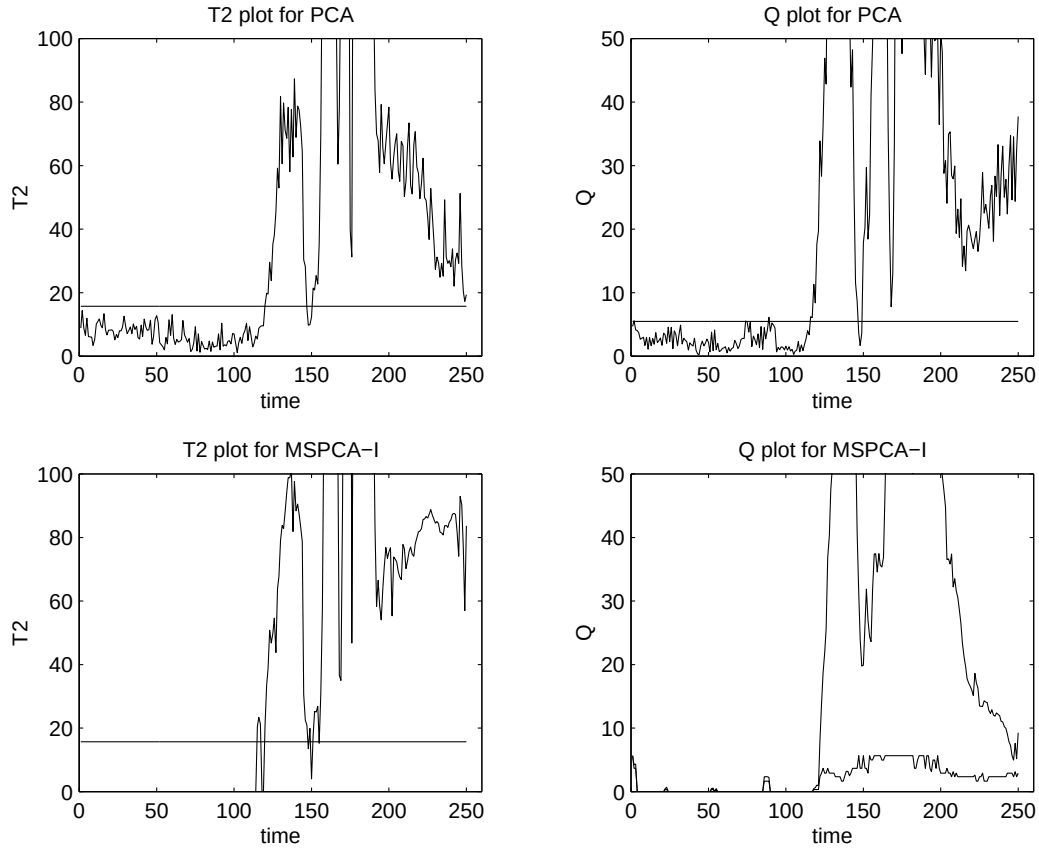


Figure 3.17: Performance of T^2 and Q plots of PCA and MSPCA-Integer for monitoring multivariate industrial data

CHAPTER 4

MSART - CLUSTERING IN WAVELET DOMAIN

We have developed a method for process fault detection based on the integration of multiscale signal representation and scale-specific clustering-based diagnosis. Previous chapter demonstrated the utility of our multiscale detection scheme applied to linear projection-based methods such as PCA and Dynamic PCA. This chapter 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 MSART framework exploits the 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

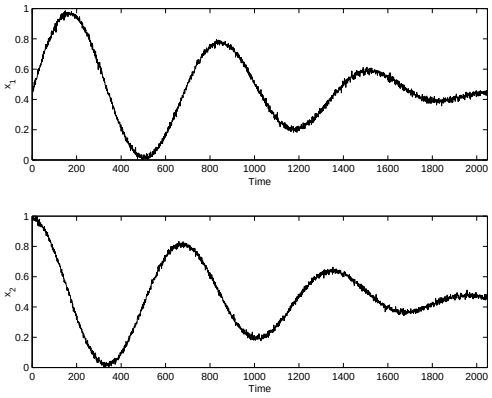
industrial case studies. Our results indicate that MSART-2, as compared to ART-2, is a general approach that may be preferable for problems where it is necessary to detect all changes drawn from processes of various statistical characteristics.

4.1 Illustration of the MSART-2 Algorithm

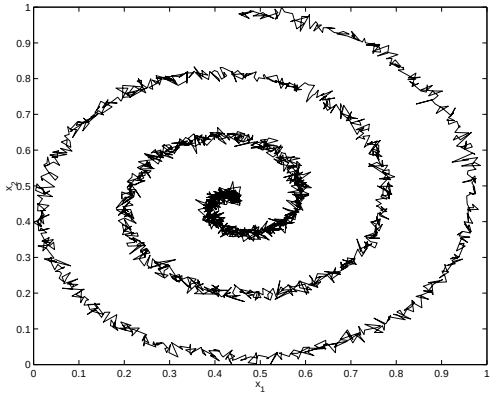
To illustrate the advantages of our approach in more detail, we present three case studies. To facilitate a visual representation, let us limit ourselves to two variables, although the method is general and can be applied to data with any number of variables. The three cases differ in terms of noise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

The experiments discussed in this chapter 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 4.1a 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 4.1b 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



(a) Normal Data for Cases One and Two



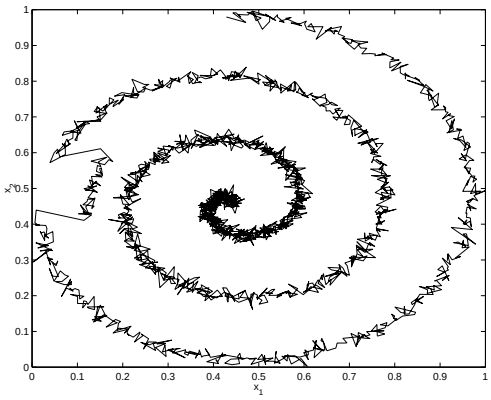
(b) Variable Correlations Under Normal Conditions

Figure 4.1: A Bi-variate Process for Illustration of the MSART-2 Approach

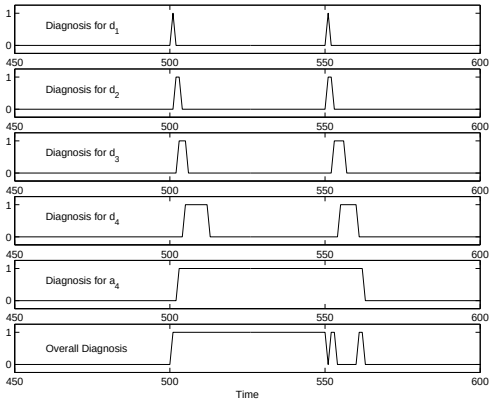
of noise, followed by resumption of normal behavior. The test signals were subjected to online diagnoses by applying (1) an ART-2 detector, (2) a moving average (MA) filter followed by an ART-2 detector (referred to as ART-2+MA), and (3) an MSART-2 detector. Comparative analyses brought out the strengths and weaknesses of the current approach with respect to the basic ART-2 based detection/diagnosis.

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

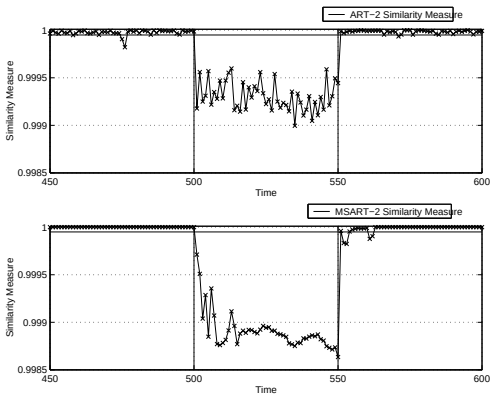
Figure 4.2a shows the test data used for diagnosis in this section. As can be seen as the left side of the outermost arm of the spiral, a shift was introduced to simulate abnormal behavior from time-step 501 to time-step 550. The number of scales, L, was chosen to be 4.



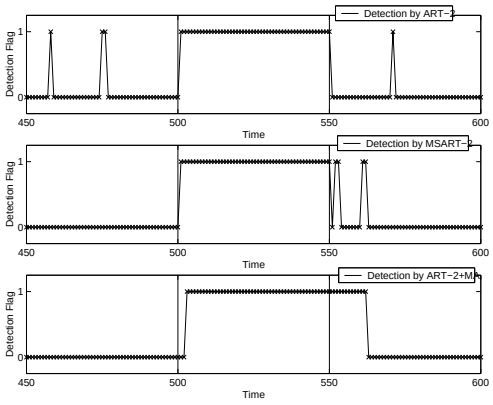
(a) Simulated Abnormal Condition for Case One



(b) Diagnoses of Decomposed Signals



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



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

Figure 4.2: Comparative Performance for Test Case One

Scale Selection

The detection flags of the decomposed signals provide an insight into the mechanism of scale selection in the MSART-2 architecture. Figure 4.2b 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 4.2b, a detection 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 4.2b: bottom-most graph), illustrates the effect of simultaneous selection of multiple scales. Figure 4.2b 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 a 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 4.2b: 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 4.2b, 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 4.2b), 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-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 4.2c and 4.2d. Figure 4.2c 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 4.2c).

Figure 4.2d 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.

4.1.2 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 4.3a). 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 4.3b through 4.3d 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 4.3c shows the similarity measures for abnormal operation for ART-2 and MSART-2. When compared to Case One (Figure 4.2c, 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 4.3c, top graph). Similar reduction in the level of separation is seen with MSART-2 as well (Figures 4.2c and 4.3c, 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, a close match of the test and normal data affected the similarity measure.

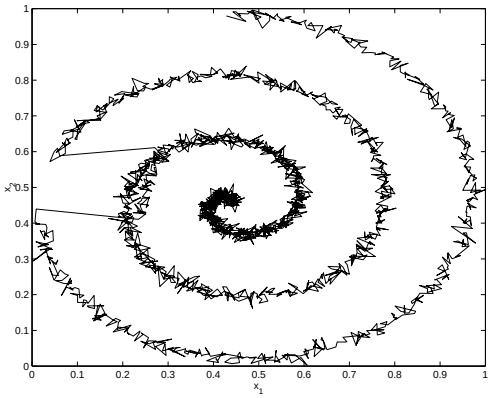
As can be seen from the diagnoses reported in Figure 4.3d, ART-2 did not detect the fault consistently because of the smaller distinction between normal and abnormal data with respect to the level of normal noise. Use of the MA filter alleviated

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.

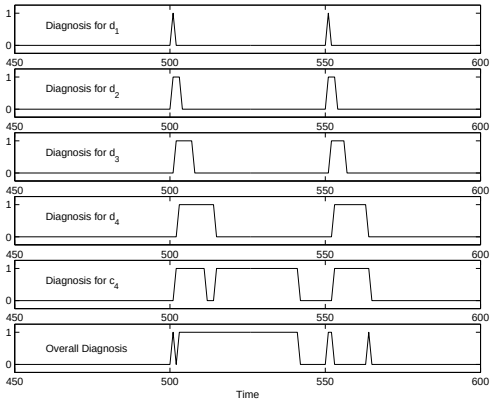
4.1.3 Case Three: A High-noise Process

The purpose of the third case-study was to study the effect of multiscale architecture on anomaly detection in the presence of noisy signal. The training and testing data used for this case are provided in Figures 4.4 and 4.5a. 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.

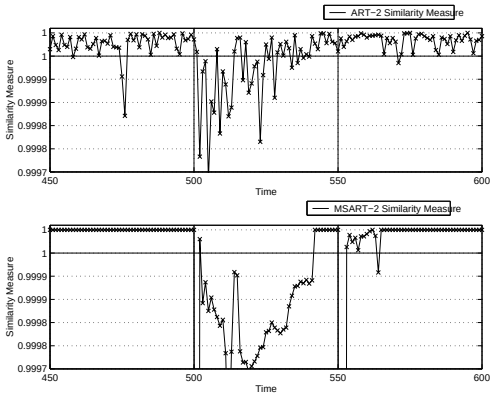
In contrast with the earlier cases, the high noise in this case hampered the detection of the transient phases in finer scales. Analysis of the decomposed scales (Figure 4.5b) shows that the two finest ART-2 detectors (d_1 and d_2) did not detect the shift at all, unlike the earlier cases. The overall 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 4.5c and 4.5d compare ART-2 and MSART-2 detectors for this test case. The similarity measure plot (Figure 4.5c, top graph) shows that ART-2 was unable to separate the normal and abnormal process operation. Thus, the ART-2 detector



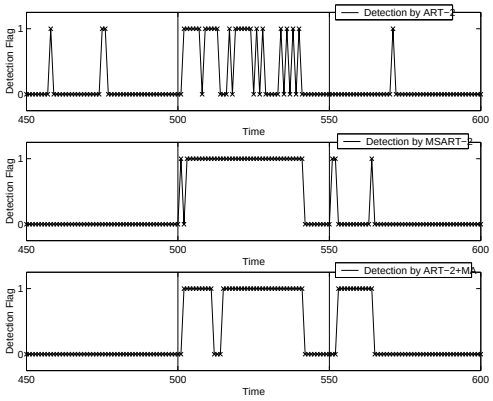
(a) Simulated Abnormal Condition for Case Two



(b) Diagnoses of Decomposed Signals



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



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

Figure 4.3: Comparative Performance for Test Case Two

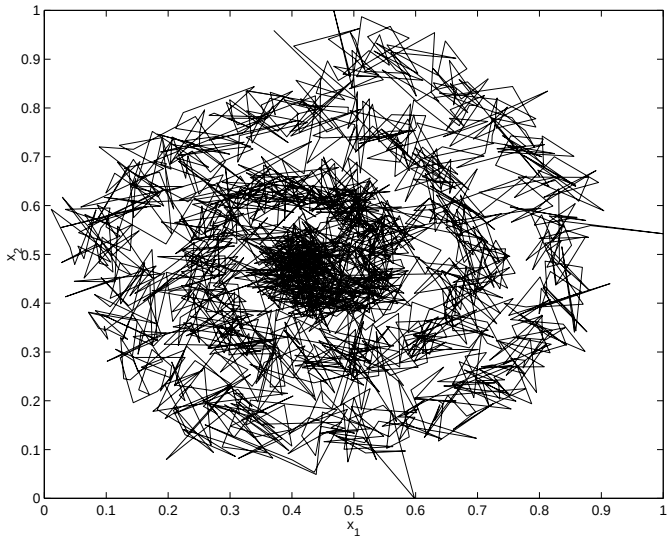
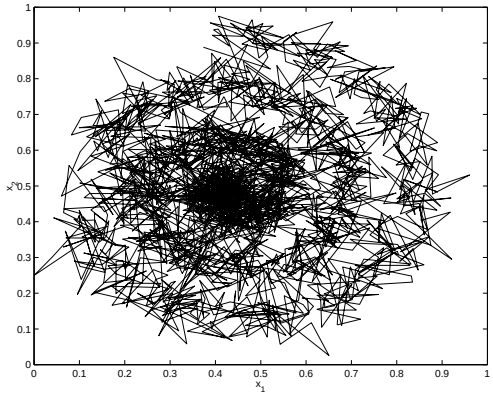


Figure 4.4: Variable Correlations Under Normal Conditions for Case Three

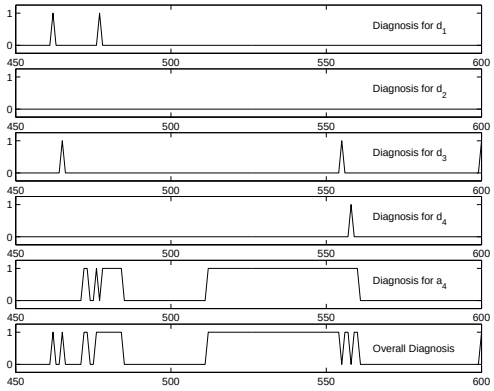
led to many missed and false alarms (Figure 4.5d, top graph). The detection flag for MSART-2 closely resembled that for ART-2+MA. Both detectors detected the transition away from normal and resumption of normal operation at a lag approximately equal to the width of the filter used (16 in this case).

Because of the large amount of 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.

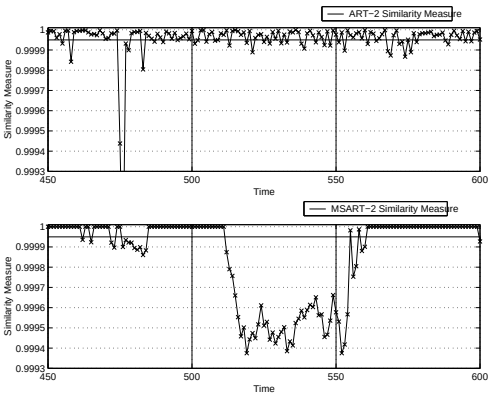
These three representative cases illustrate that the multiscale approach is a generic approach that works well on wide variety of situations. On the other hand, single-scale methods such as ART-2 with and without moving average filter work best only



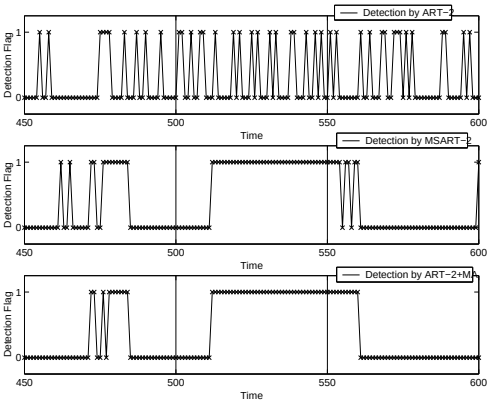
(a) Simulated Abnormal Condition for Case Three



(b) Diagnoses of Decomposed Signals



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



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

Figure 4.5: Comparative Performance for Case Three

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).

4.2 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 on the basis of the ARL curves.

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 (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.

4.2.1 A Univariate Process

We now consider the following simple univariate process model:

$$x(t) = N(0, 1) \tag{4.1}$$

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 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 in Figure 4.6. 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 since the shift is easily separable from the inherent noise in the mapping.

4.2.2 A Multivariate, Linearly Correlated Process

Consider the following linear multi-variate process:

$$x_1(t) = N(0,1) \tag{4.2}$$

$$x_2(t) = N(0,1) \tag{4.3}$$

$$x_3(t) = \frac{x_1(t) + x_2(t)}{\sqrt{2}} \tag{4.4}$$

$$x_4(t) = \frac{x_1(t) - x_2(t)}{\sqrt{2}} \tag{4.5}$$

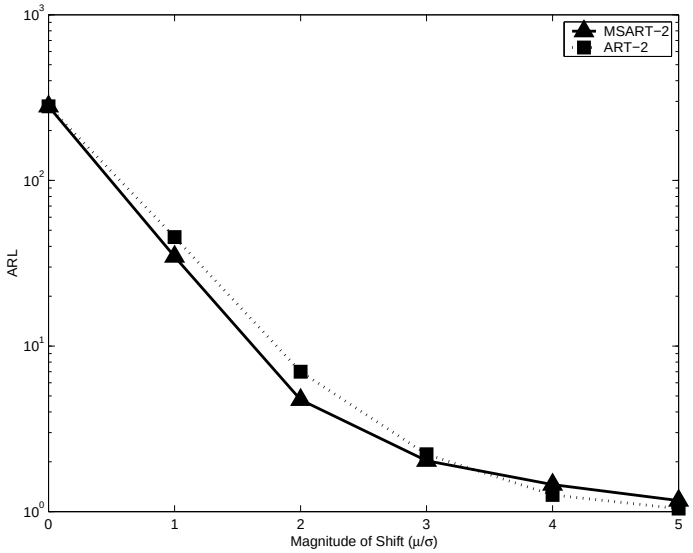


Figure 4.6: Comparison of ART-2 and MSART-2 Performances based on ARL for a Univariate Process

$$y_i(t) \quad = \quad x_i(t) + \epsilon_i(t) \tag{4.6}$$

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 4.7. Again, we observe that MSART-2 outperformed ART-2 for a wide range of shifts. Shifts of a given magnitude are applied

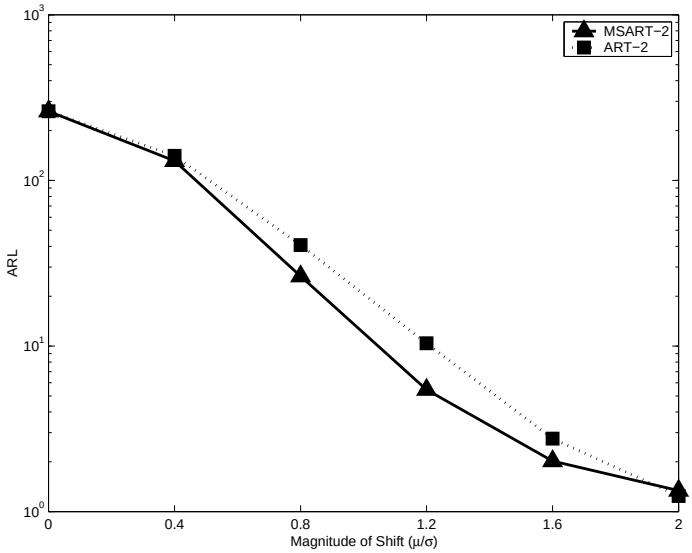


Figure 4.7: Comparison of ART-2 and MSART-2 Performances based on ARL for a Linear Multivariate Process

across all process variables, and hence shifts are detected earlier (lower run-lengths) when compared to the univariate process (Figure 4.6). We observe that MSART-2 performs better than ART-2, except for large shifts when abnormal operation is well-separated from normal operation.

4.2.3 A Multivariate, Nonlinearly Correlated Process

We now present the ARL results for a non-linear spiral process similar to the one used for Section 4.1.

$$r(t) = r(t-1) - 0.001 \tag{4.7}$$

$$\theta(t) = \theta(t-1) + 2 * \pi * 0.006 \tag{4.8}$$

$$x_1(t) = r(t) * cos(\theta(t)) \tag{4.9}$$

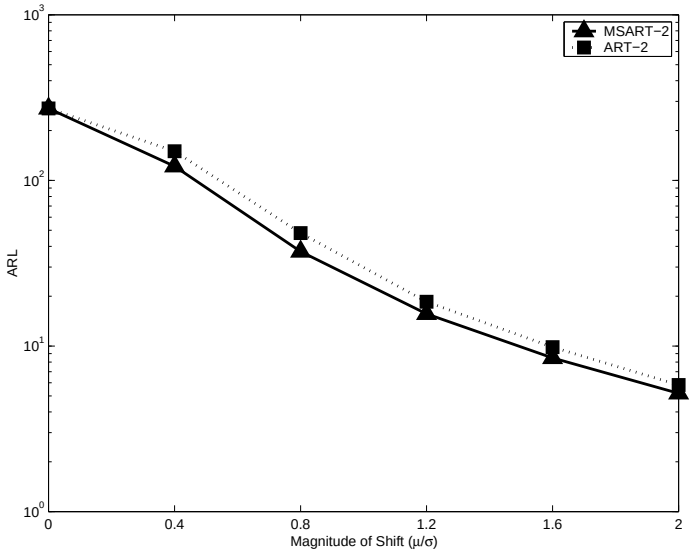


Figure 4.8: Comparison of ART-2 and MSART-2 Performances based on ARL for a Nonlinear Multivariate Process

$$x_2(t) = r(t) * sin(\theta(t)) \tag{4.10}$$

$$y_i(t) = x_i(t) + \epsilon_i(t) \tag{4.11}$$

The ARL results presented in Figure 4.8 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 4.7, 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.

Similar to the ARL analyses from the previous chapter, the ARL curves presented in this section confirm the utility of MSART-2 over ART-2 as established in Section *. By exploiting wavelet-domain clusters, we see that MSART-2 can detect small shifts with smaller detection delays when compared with ART-2 for three example process models of varying statistical characteristics.

4.3 Industrial Case Studies

As claimed earlier, deviations from normality in real processes can be slow or fast. In addition, they may differ in the level of noise, and in a random and/or deterministic nature of the change. We go back to the two representative process changes from a a real large-scale petrochemical process that were discussed in the context of univariate MSSPC in the previous chapter. For each example case, ART-2, MSART-2, and ART-2+MA were trained with the same training data and same training parameters. Similar to the illustration from Section 4.1, 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.

4.3.1 Example 1: Drier Cooling

The normal and test data for this example were presented earlier as Figures 2.1 and 3.12 in the previous chapter. 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 4.9.

The test data from Figure 3.12 were subjected to ART-2, MSART-2, and ART-2+MA detectors trained on the normal data from Figure 2.1. 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.

4.3.2 Example 2: Sensor Malfunction due to Oil Accumulation

In the context of univariate MSSPC, this example process was discussed earlier in the earlier chapter as Figure 3.14. 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.

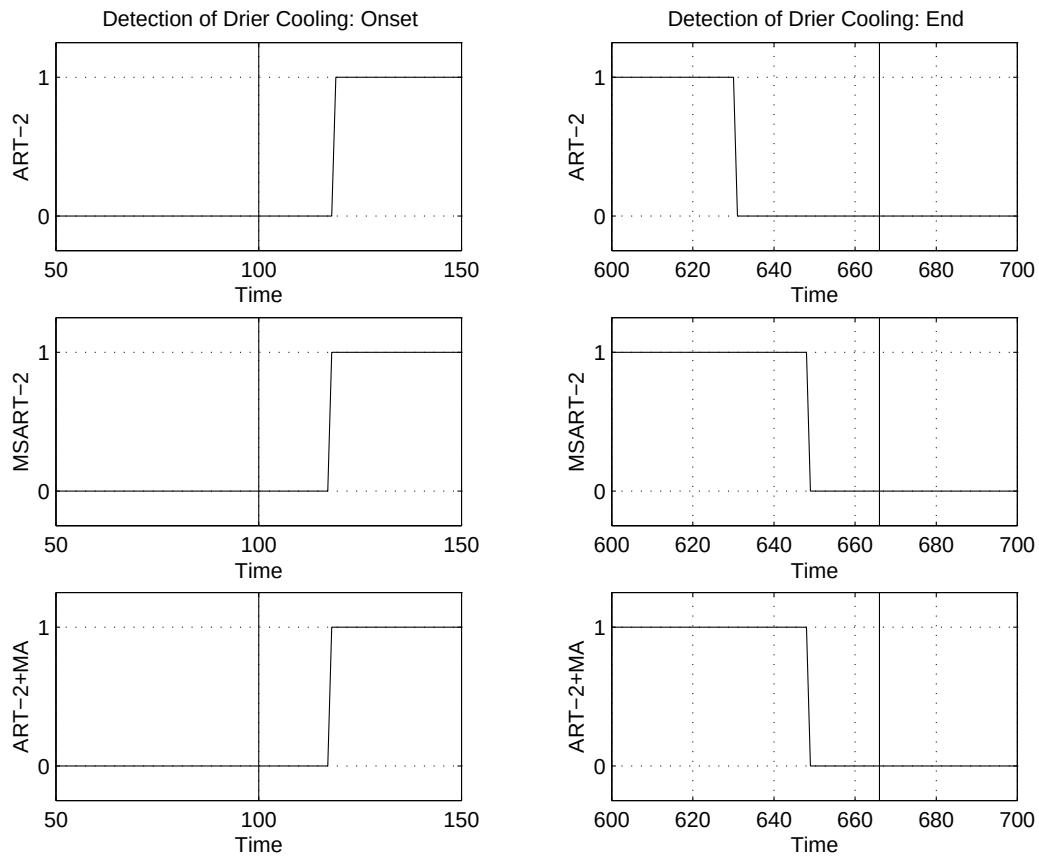


Figure 4.9: Comparative Detection Performance for the Drier Cooling Event

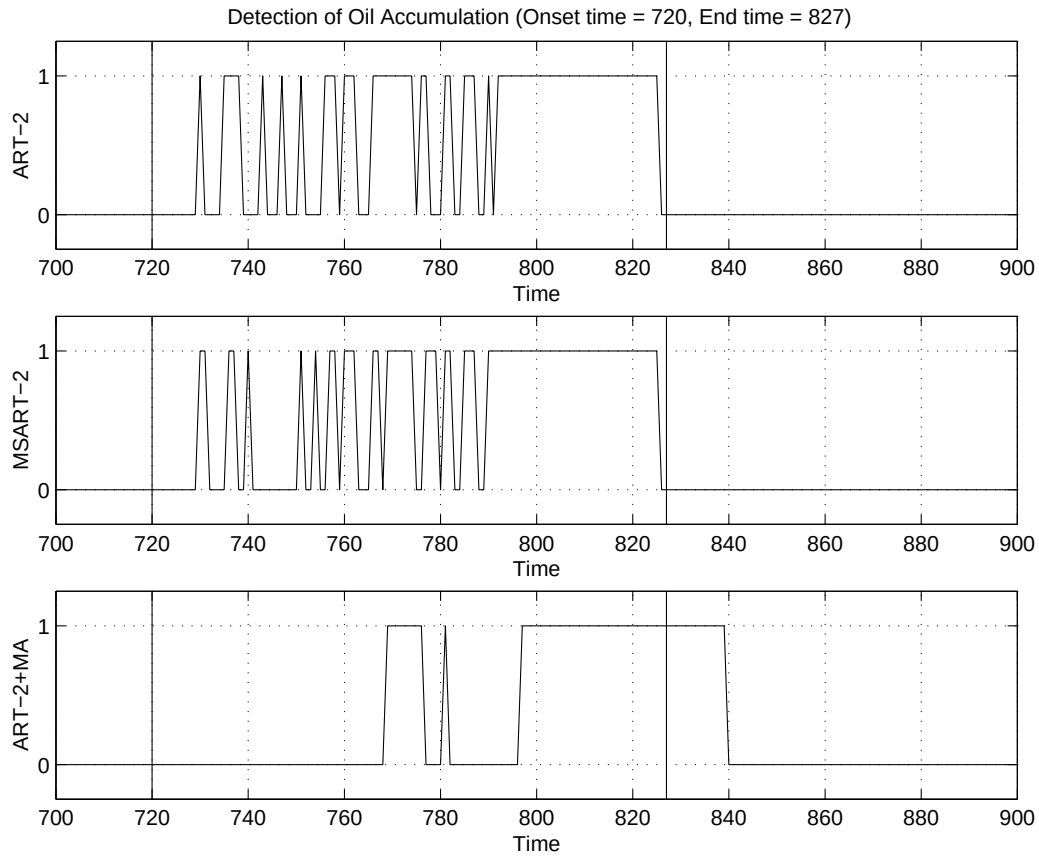


Figure 4.10: Detection of Oil Accumulation (Sensor Failure)

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. 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 4.10). Similar to figure 4.9, we observe that MSART-2 conforms to the scale of the change under consideration and mimics the best performance for the event at hand.

CHAPTER 5

INDUSTRIAL VALIDATION AND COMPARISON

Results from the earlier chapters illustrated the utility of the proposed algorithms primarily with the help of simulated process models and/or specific detection limits. The analysis presented in this chapter displays the capability of our algorithms when compared with uniscale methods on univariate/multivariate industrial data over a broad range of model fidelities and false alarm rates. This analysis thus provides a detailed comparison of the performances of these methods on a broad range of detection operating regimes. In some cases, we also compare the performance of the human operator with that of the MSART algorithm.

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

5.1 Case Studies

5.1.1 A Valve Leak Malfunction

Figure 5.1 shows the sensor readings during an abnormal process operation caused by a leaking valve. This event manifests itself primarily in three non-redundant sensors, trends of which are shown. As can be clearly observed, different sensors respond to the same root cause with different scales and delays due to differences in the underlying physical quantities being measured and also due to the process control scheme in place.

The performance curve from Figure 5.2 shows the average missed alarm rate plotted against the average false alarm rate. Different false alarm rates were achieved by changing the respective threshold parameters for the detection algorithms in question, namely PCA, MSPCA, ART, or MSART. For each false alarm rate, the corresponding average missed alarm rate was calculated for each abnormal event and plotted for each algorithm. It is obviously desirable to have the lowest possible missed alarm rate for a given false alarm rate.

It can be seen that PCA outperforms ART for the detection of this event. This is expected since the process variables involved in this event are linearly correlated, stationary, and approximately Gaussian. The proposed multiscale versions of these algorithms, MSPCA and MSART, outperform the respective single-scale versions. Due to the linear correlation between wavelet coefficients, corresponding to a linear correlation between process variables, MSPCA results in fewer false alarms than MSART.

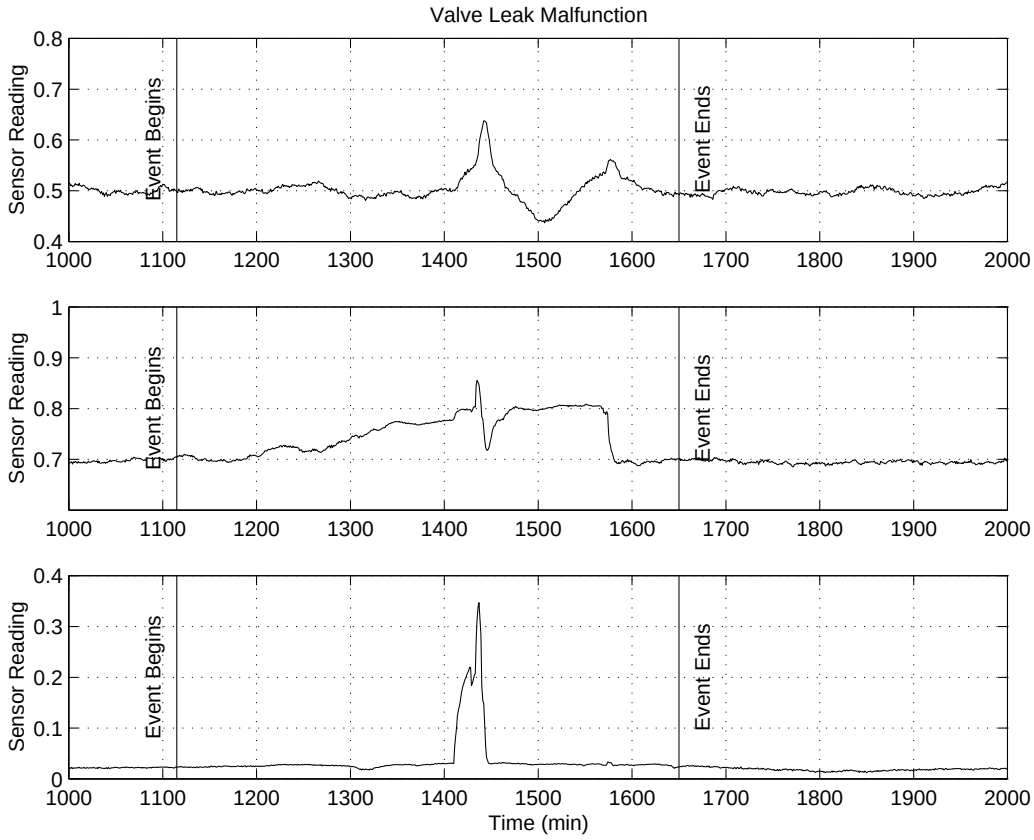


Figure 5.1: Sensor Data for the ValveLeak Event

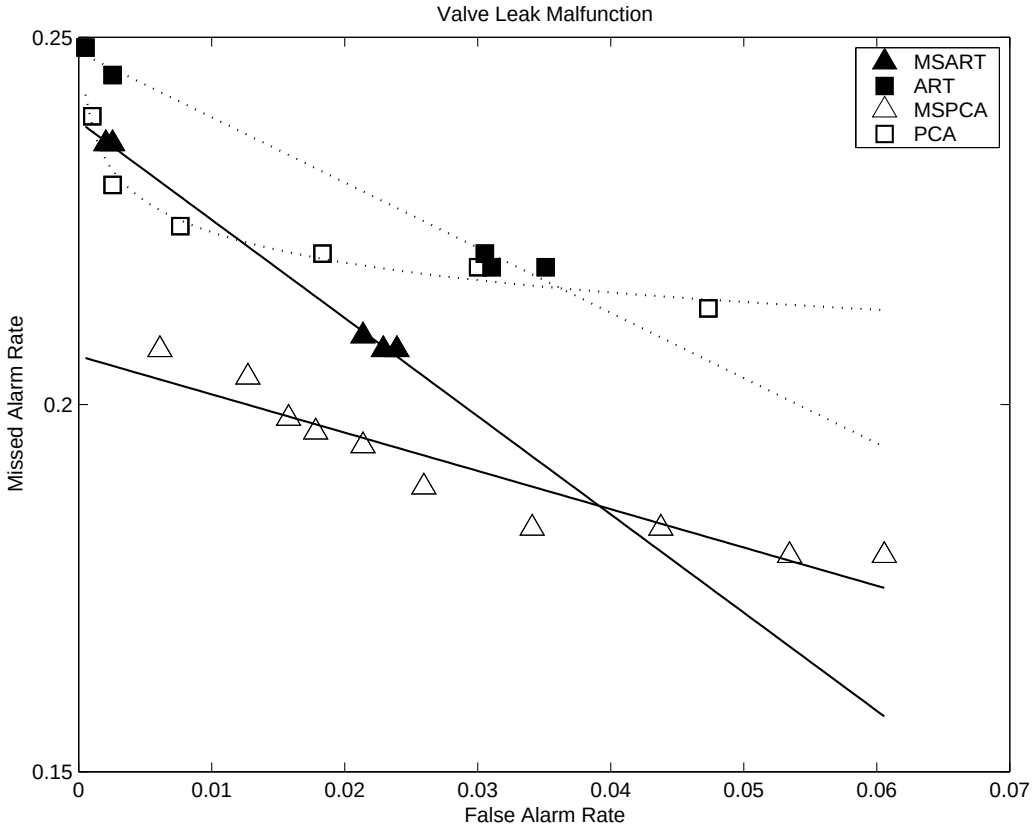


Figure 5.2: Comparison of Performance for the ValveLeak Event

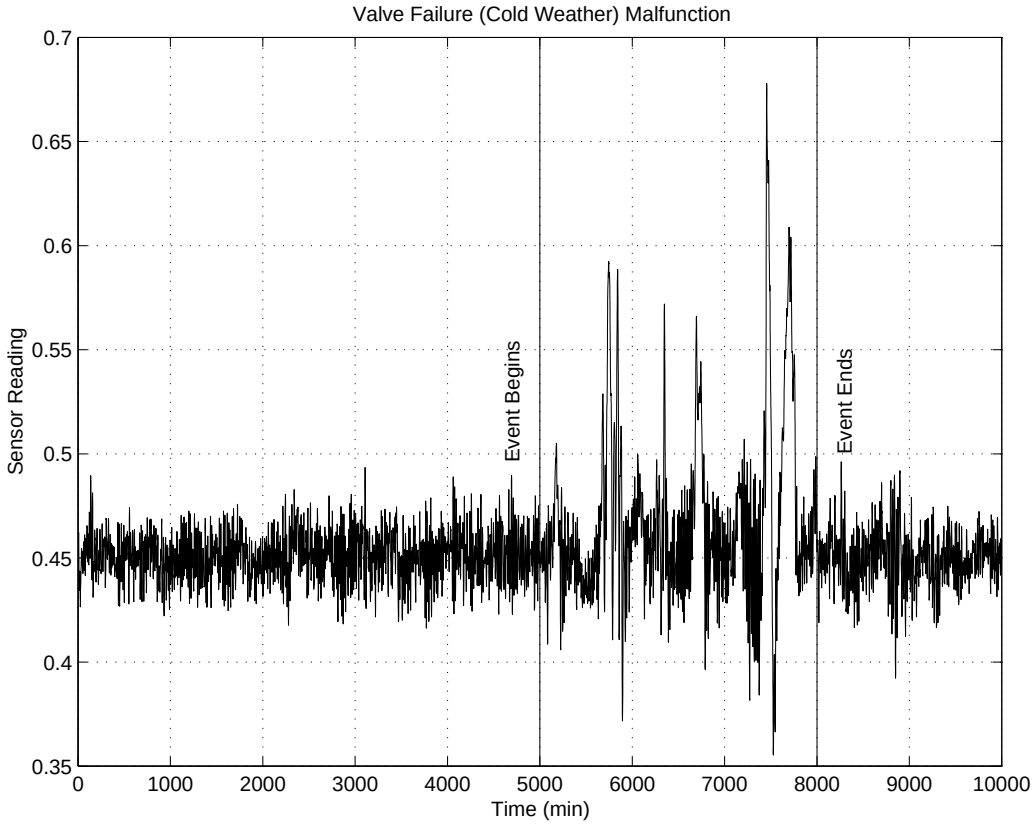


Figure 5.3: Sensor Data for the ColdWeather Event

5.1.2 A Cold Weather Malfunction

Figure 5.3 shows our next industrial example, which involves a valve failure due to an unexpectedly lower ambient temperature. This is a univariate example which approximately obeys the assumptions of stationarity and unimodal Gaussian distribution. PCA is again seen to miss fewer alarms when compared to ART as seen in Figure 5.2. Similarly, MSPCA performs better than MSART. As was the case of the previous example, the multiscale versions of PCA and ART outperformed the corresponding uniscale detectors.

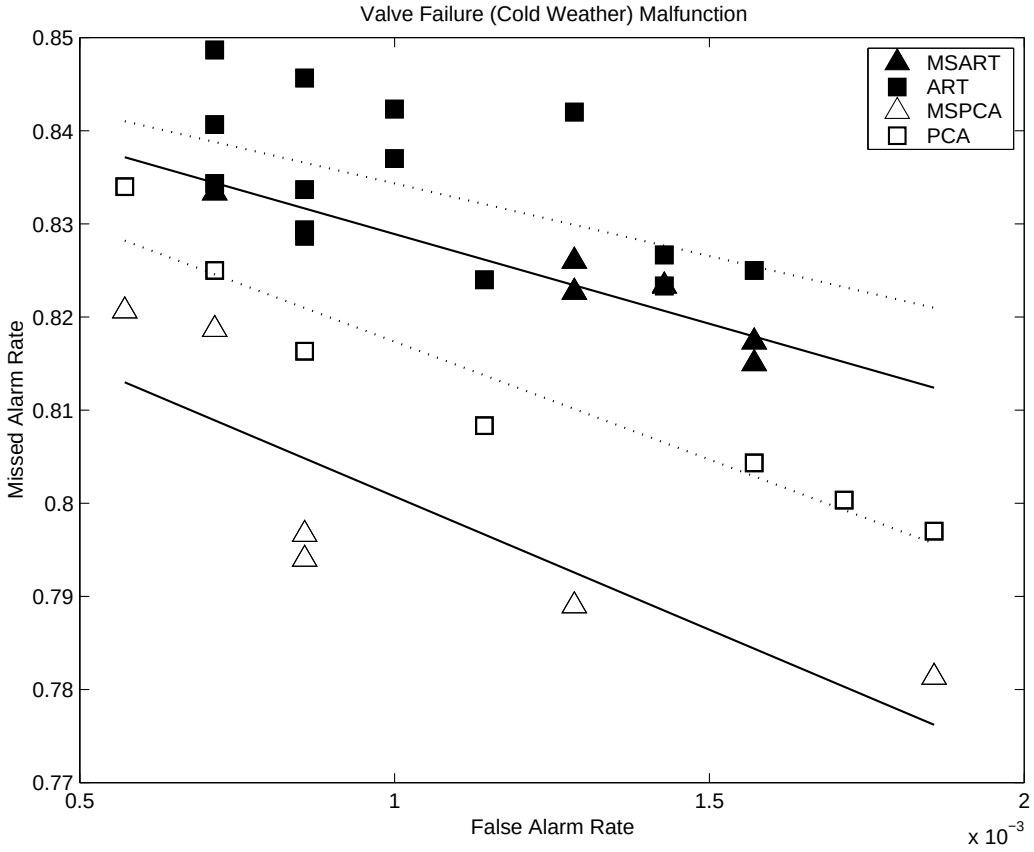


Figure 5.4: Comparison of Performance for the ColdWeather Event

5.1.3 A Change in Furnace Feed Event

Our third example corresponds to a "normal but unusual" process operation associated with a change in furnace feed. It was typically observed multiple times every day and lasted a few hours. Figure 5.5 shows several instances of this event. Although a high-level description of the sensor trend is seen to be similar for all instances, the event clearly showed characteristics for which multiscale analysis would be valuable, since the event lasted for varying intervals for different instances and rose to different magnitudes with differing approach and reset rates.

The performance curves for this event are shown in Figure 5.6. Unlike previous two examples, we observe that ART performs better than PCA, possibly due to deviations from ideal assumptions such as stationarity and Gaussian distribution. These deviations were seen to affect MSPCA more strongly than PCA, perhaps due to the strong auto-correlation in the data. However, MSART continues to outperform ART. The nonlinear modeling capability of ART makes it insensitive to non-stationarity and non-Gaussian behavior, and these benefits appear to hold in the wavelet domain as well.

5.1.4 A Feed Malfunction

The data presented in Figure 5.7 show a distinctly nonlinear and multi-modal correlation in the bivariate sensor space. The process is seen to exist in three disjoint normal regimes. As a result, detector based on ART is expected to perform better than PCA for this process. The Figure 5.8 shows this assertion to be true. Although

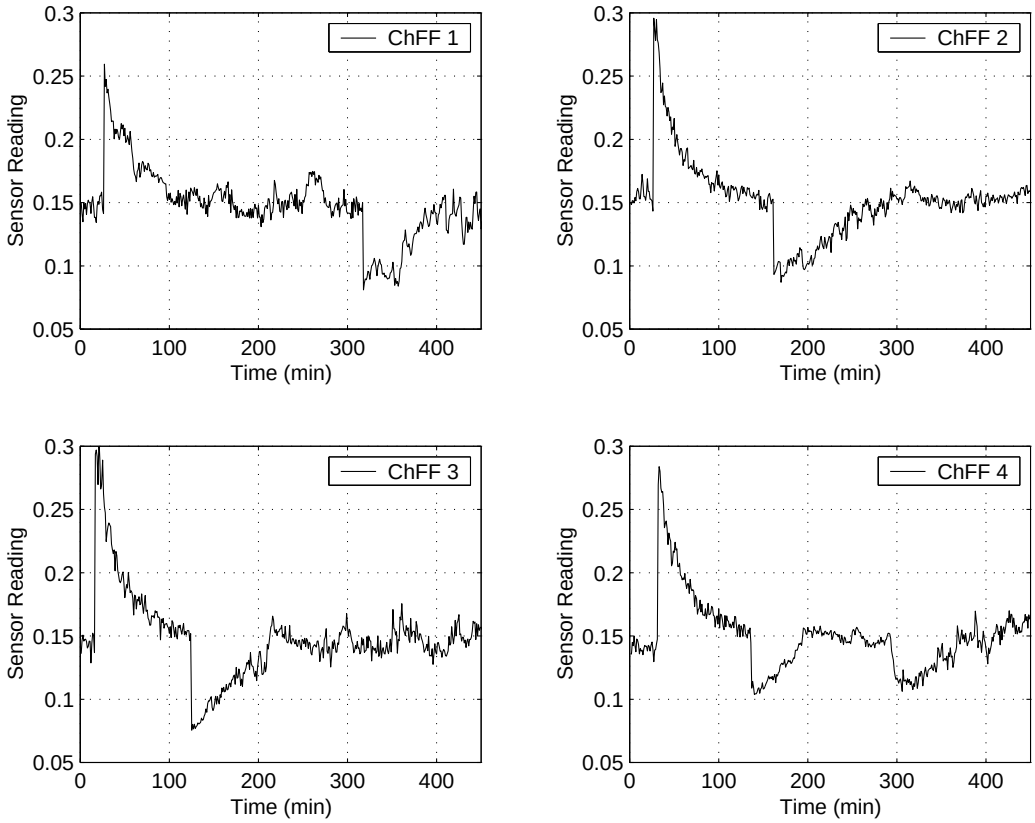


Figure 5.5: Sensor Data for the Change in Furnace Feed Event

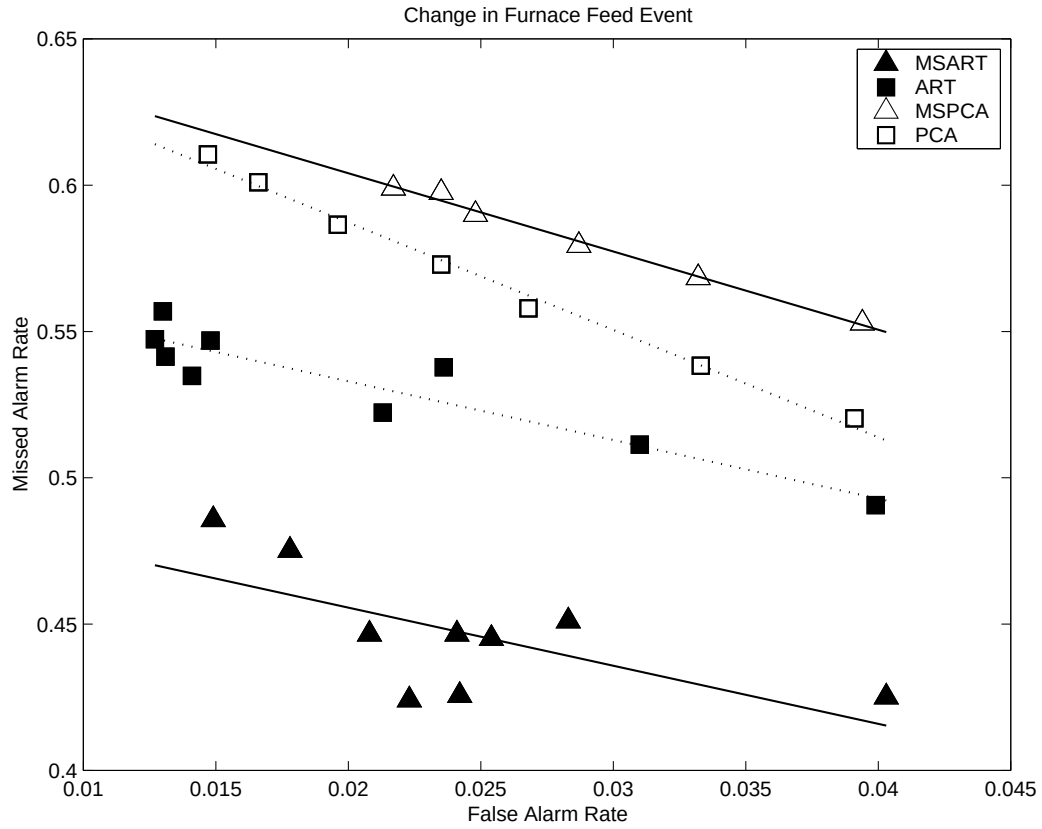


Figure 5.6: Comparison of Performance for the Change in Furnace Feed Event

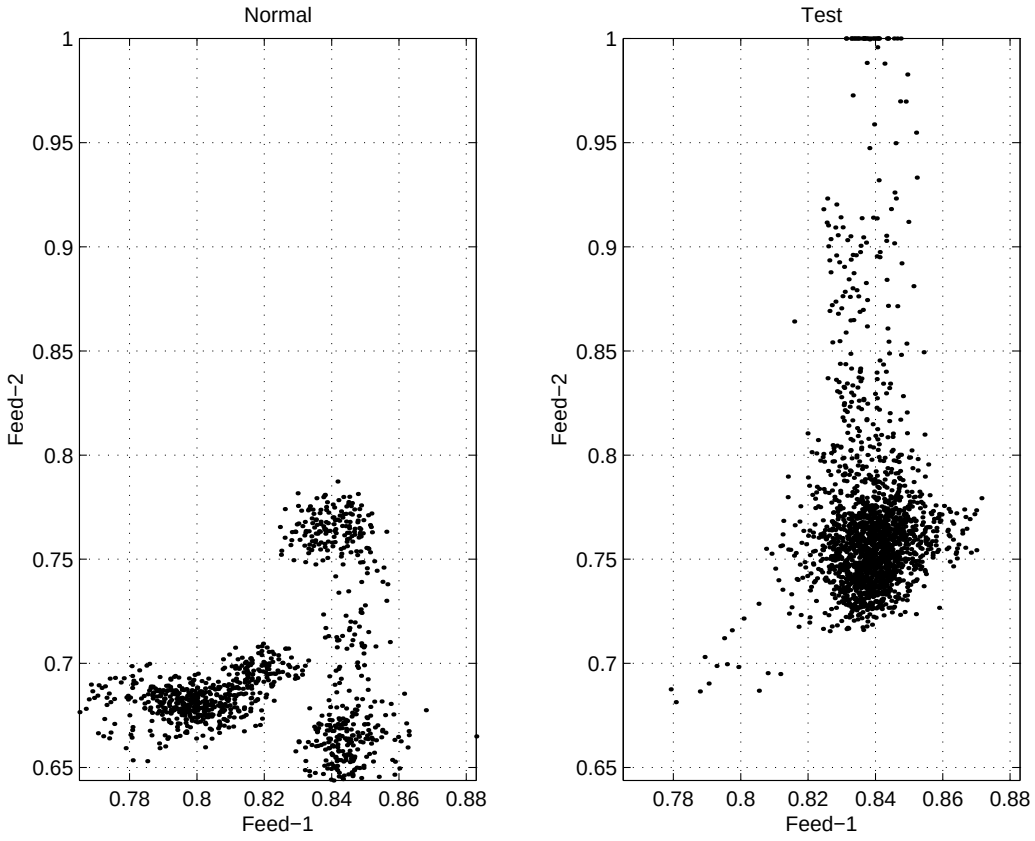


Figure 5.7: Sensor Data for the FeedMalfunction Event

the multiscale versions of these algorithms are seen to perform better, the improvement is more significant in the case of ART due to the nonlinear nature of the variable correlation.

5.2 Comparison with Human Operator

To provide a perspective of detection performance of the algorithms proposed in this work, we have provided two examples where the detection delay of a human operator is compared with that of our MSART algorithm. Figure 5.9 shows the

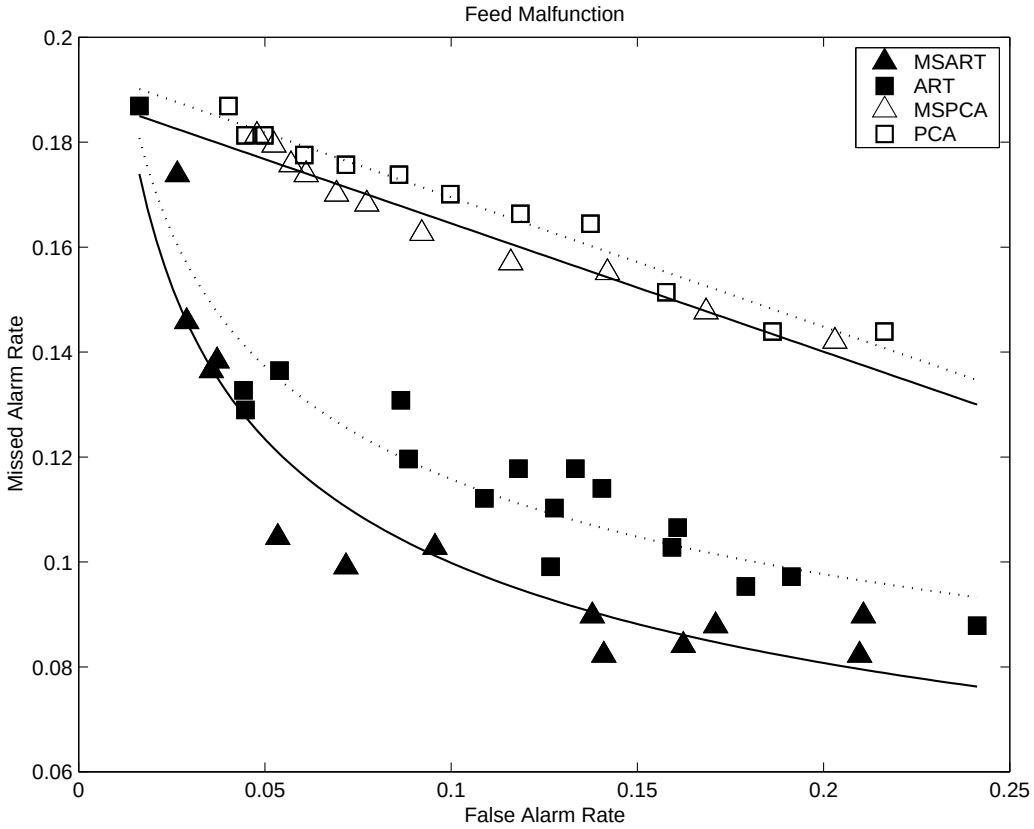


Figure 5.8: Comparison of Performance for the FeedMalfunction Event

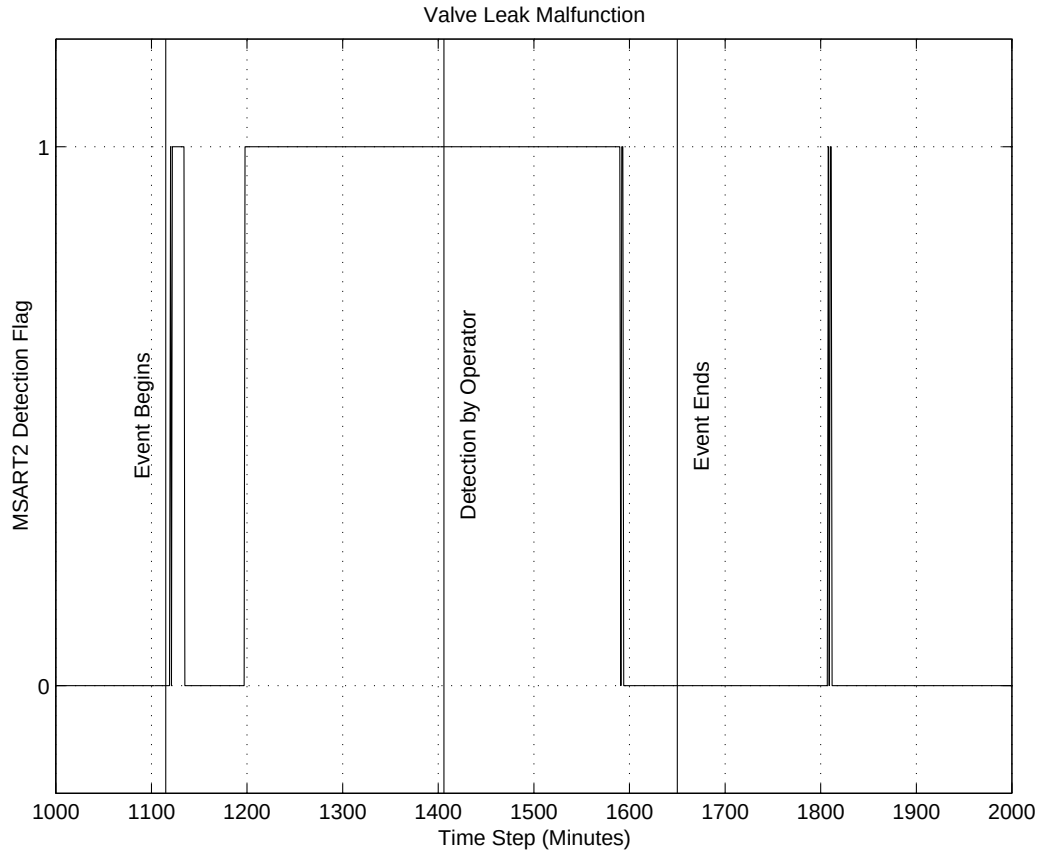


Figure 5.9: Operator Detection of the ValveLeak Event

detection flag of the MSART algorithm for the valve leak event (Figure 5.1) with 0 being normal and 1 being abnormal diagnosis. It can be seen that the MSART algorithm detected the abnormal event hours in advance of the human operator.

Another example is provided with the help of figures 5.10 and 5.11. This bivariate event was caused by a sensor "acting up", i.e., providing erroneous reading. It can be seen that the MSART algorithm is a few minutes faster than the human operator, which is a significant fraction of the event duration.

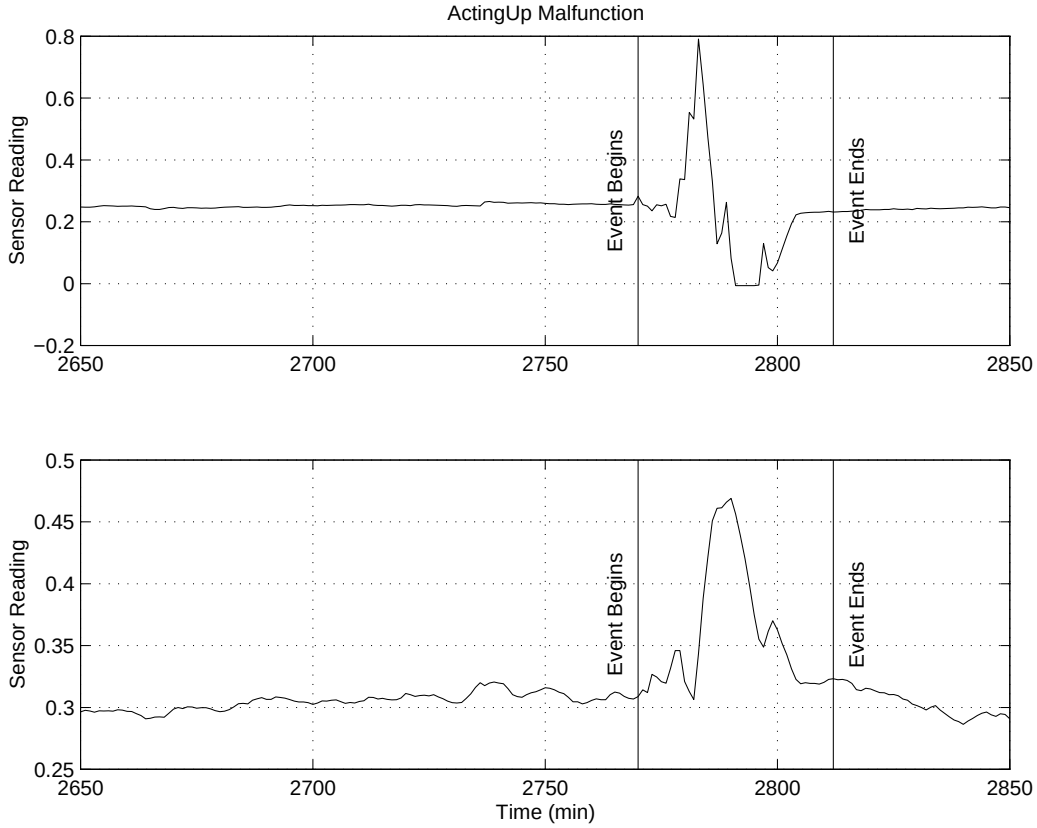


Figure 5.10: Sensor Data for the ActingUp Event

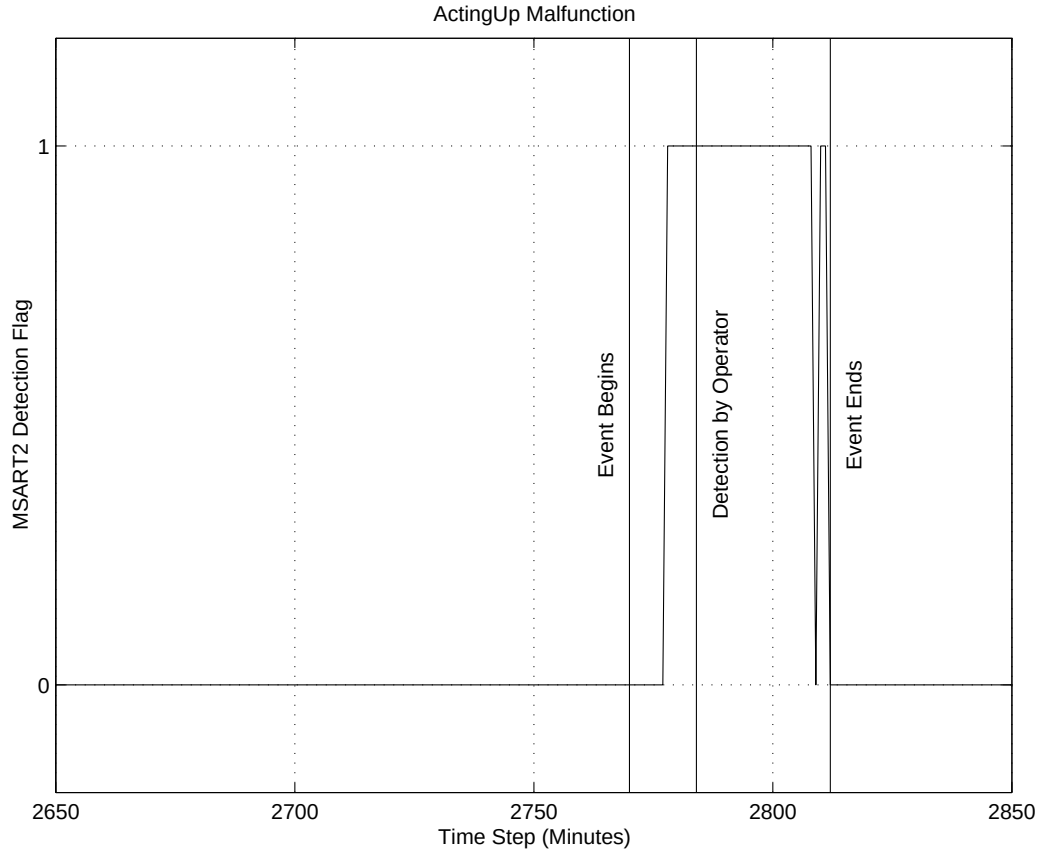


Figure 5.11: Operator Detection of the ActingUp Event

CHAPTER 6

LESSONS LEARNED FROM INDUSTRIAL IMPLEMENTATION

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

First of all, even with the best of efforts by the engineer, the sheer scale of the process often caused annotations to be inconsistent. An event listed as "unusual" for one day, for instance, may be omitted entirely for the next day. Hence, the data used for training had to be very carefully screened before use. Often, the information available to the operator came from sources outside of the 509 sensors and hence similar events ended up having different annotations. Although these issues adversely

affected all SPCs being compared, the ART-based detectors were found to be specifically sensitive to contradicting annotations and overlapping class definitions. During incremental training, if new data overlapped in the sensor space with previous clusters that belonged to a different class, the previous clusters were often completely overwritten, thus generating wrong diagnosis for future cases.

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

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

Since the sensor signals were being sampled per minute, the detector algorithms were expected to diagnose each snapshot of sensor data within a small fraction of a minute. However, the diagnosis operation in ART is an exhaustive search process. The time taken for a decision grows exponentially with the complexity of the search space, which is, in turn, a function of the dimensionality of the space in terms of the number of process variables and the nature of cross-variable relationships. However, as is the case with many large-scale problems, the relationships across process variables were mostly localized and related to the spatial position of the sensor in a process unit. To be able to manage scale and related issues such as speed, it was necessary to decompose the task of diagnosis into smaller subsets which were solved with smaller ART and PCA detectors. The decomposition was generated based on operator knowledge and reduced the associated structural and computational complexity to a large extent. These *focused* SPCs resulted in a better diagnostic performance as well as were easier to train. In addition, the ART detector could adequately classify between different abnormal behaviors for input spaces with low dimensionality and less number of classes to distinguish between. However, it led to false alarms when the detector was trained with many different events. This problem was also solved with this focused detection approach.

Lastly, there was an expectation that the detector algorithm provide a list of sensors that contributed the most to the detector decision. A knowledge of the contributing sensors enables the operator to track the root cause of the abnormality. For PCA and MSPCA, a mechanism for calculating the contribution charts [55] has been developed. For ART and MSART, similar calculations based on the directions in the sensor space that most contribute to the difference between actual and expected

sensor readings lead to a list of three most contributing sensors. Although *ad hoc*, this method often provided an accurate list of contributing sensors.

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

CHAPTER 7

CONCLUSION

The multiscale approach based on wavelets studied in this dissertation is shown to be ideal for SPC of measurements that contain features with different localizations in time or frequency. Examples of such features include deterministic changes with different temporal durations, and stochastic variation with changing intensity over frequency or time. MSSPC exploits the ability of wavelets to compress deterministic features at all scales to a small number of relatively large coefficients. It adapts to features at different scales by focusing only on those scales that contain coefficients outside the detection limits at each scale. The values of the detection limits at each scale depend on the nature of the correlation of measurements representing normal operation. MSSPC adapts both, the nature of the filter applied to the measurements as well as the detection limits for a selected confidence limit, according to the scale of the abnormal feature. Thus, MSSPC can specialize to many existing SPC methods including Shewhart, MA, CUSUM, and EWMA charts depending on the scale of the abnormal feature and the selected wavelet.

Two variations of MSSPC are studied in this dissertation that investigate two different ways of discretization of wavelet translation parameter. If the translation parameter is discretized dyadically, then the coefficients are downsampled at each

scale. This permits the use of orthonormal wavelets, and approximate decorrelation of autocorrelated measurements, but introduces a time delay in obtaining the coefficients. If the translation parameter is discretized uniformly, then there is no downsampling of the coefficients. This permits on-line decomposition without delay, but the wavelets are no longer orthonormal, and the coefficients remain autocorrelated. ARL analysis shows that MSSPC with dyadic discretization performs well for detection of shifts from highly autocorrelated or nonstationary measurements since the benefits of decorrelation outweigh the delay due to downsampling. If the measurements are uncorrelated or only moderately correlated, the benefits of decorrelation are not as significant, and it is better to decrease the time delay in detecting a change. In such situations, it is better to use MSSPC uniform discretization of the wavelet translation parameter.

Comparison of the average run lengths with existing methods for univariate and multivariate SPC indicates that MSSPC does not perform better than methods specifically designed to detect certain types of changes. For example, Shewhart charts are best for detecting large shifts, while MA or EWMA charts can be designed to detect shifts of a certain size. Other methods such as CUSCORE [50] can be tailor-made for detecting changes that are known a priori. These methods are not adaptive or general, and do very well only for the type of change they are designed to detect. In contrast, MSSPC is a general method that consistently does well in detecting a changes in different types of measurements and with a broad range of characteristics. Since the nature of change is unknown a priori for most industrial abnormal or operational situations, MSSPC seems to be an ideal choice for a general purpose

approach. Furthermore, MSSPC can easily deal with autocorrelated measurements and multivariate problems.

Previous work by Whiteley and Davis [30] established ART-2 as a technique 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 presented in this work was shown to significantly enhance the range of applicability of the ART-2-based diagnosis algorithm by reducing its sensitivity to noise without sacrificing on its detection performance. Our approach integrates scale selection and clustering-based diagnosis. Application on both simulated as well as industrial case studies show that our MSART-2 scheme enhances the noise tolerance of the ART-2-based diagnosis algorithm by Whiteley and Davis.

This work also presented rigorous comparison among the various single and multi-scale detection algorithms on several different industrial case studies with univariate as well as multivariate unusual process patterns. Comparisons were made on the basis of operating curves that documented the rate of missed alarms versus the rate of false alarms. We show that PCA and MSPCA outperform ART-2 and MSART-2 in the presence of certain ideal conditions such as linearity and stationary, Gaussian behavior. For non-linear and/or discontinuous process correlations, ART-2 outperforms PCA whereas MSART-2 performs even better. These case studies show that our algorithms detected actual abnormalities in a complex process plant much sooner than human operators, while maintaining a low false alarm rate. The ART-2 models were compared with other algorithms being developed elsewhere, such as QTA and

MVSE, and compared favorably due to qualities such as the ability to build incrementally, and the ability to model linear as well as nonlinear, possibly disjoint process correlations.

The MSSPC methodology can also be used to transform any single-scale process monitoring method to its multiscale equivalent. In addition to developing multiscale versions of other process monitoring methods, there are many other areas of future work. The tuning parameters in MSSPC include the depth of decomposition and type of wavelet. It is expected that both decisions can be automated. This dissertation has only focused on the use of Haar wavelets, but the approach may easily be used with other types of wavelets. The use of smoother, boundary-corrected wavelets may provide better performance than Haar wavelets due to better feature extraction and decorrelation abilities. Furthermore, extension to wavelet packets may permit MSSPC to automatically select the best family of basis functions from a large library. MSSPC may also be extended to the identification of the root causes of the abnormal operation. Since process operation involves making decisions at different time scales, the multiscale approach may also permit integration of various operation tasks.

APPENDIX A

ANALYTICAL DERIVATION OF ARL CURVE FOR UNIVARIATE MSSPC

Consider an IID Gaussian process X with mean 0 and standard deviation σ. If it undergoes a mean-shift of magnitude δ at time $k = 1$ then the variables $X[k], k = 2, 3, \ldots, \infty$ are IID Gaussian variables with mean δ and standard deviation σ. The probability of not detecting this shift with a Shewhart chart using limits $[-\eta\ \eta]$ is,

$$\begin{aligned} \beta &= P\{-\eta \leq X \leq \eta\} \\ &= \int_{-\eta}^{\eta} f(x-\delta, \sigma)dx \\ &= \Phi(\eta - \delta) - \Phi(-\eta - \delta) \end{aligned} \tag{A.1}$$

where,

$$f(x, \sigma) = \frac{1}{\sigma\sqrt{2\pi}} \exp^{-\frac{1}{2}\left(\frac{x}{\sigma}\right)^2}$$

and Φ is the standard normal cumulative distribution function. The ARL is then [28]:

$$ARL = \sum_{k=1}^{\infty} k\beta^{k-1}(1-\beta) = \frac{1}{1-\beta} \tag{A.2}$$

A.1 ARL for Univariate MSSPC with Dyadic Discretization

This derivation is for MSSPC with Haar wavelets. It assumes that the wavelet and scaling function coefficients from a wavelet decomposition are completely uncorrelated.

Since Haar wavelets and scaling functions have a support of 2^m, at scale m, a mean shift can appear only in the first wavelet coefficient at each scale after the shift. Subsequent wavelet coefficients are not affected by the shift, and are statistically similar to wavelet coefficients for IID Gaussian data. A mean shift of size δ located at the second measurement gets scaled to $\delta/2^{m/2}$ in the first wavelet coefficient at any scale. The first coefficient of the last scaled signal becomes $\delta(2^L - 1)/2^{L/2}$ and the other coefficients become $\delta 2^{L/2}$. This difference between the change in the first and other coefficients is sufficiently small to be ignored in the derivation.

The probability of not detecting the shift in the first wavelet coefficient at scale m may be written as,

$$\beta_m = \Phi\left(\eta - \frac{\delta}{2^{m/2}}\right) - \Phi\left(-\eta - \frac{\delta}{2^{m/2}}\right) \tag{A.3}$$

The probability of not detecting the shift in wavelet coefficients other than the first one is denoted as β_0, and is given by Equation A.1, which is the same as the probability of not detecting a shift in IID Gaussian measurements. The probability of not detecting the shift in the first scaled signal coefficient

For a decomposition of depth, L, the probability of not detecting a shift in any scaled signal coefficient is,

$$\overline{\beta}_L = \Phi\left(\eta - \delta 2^{L/2}\right) - \Phi\left(-\eta - \delta 2^{L/2}\right) \tag{A.4}$$

By using Equation A.4 for all the scaled signal coefficients, the slight difference between the size of the shift in the first coefficient versus others is being ignored.

The ARL for MSSPC-Dyadic with $L = 1$ is determined as follows.

$$\begin{aligned} ARL_1 &= \left(1 - \beta_1\overline{\beta}_1\right) \\ &\quad +3\beta_1\overline{\beta}_1\left(1 - \beta_0\overline{\beta}_1\right) \\ &\quad +5\beta_1\beta_0\overline{\beta}_1^2\left(1 - \beta_0\overline{\beta}_1\right) \\ &\quad +\ldots \end{aligned} \tag{A.5}$$

The first term in Equation A.5 is the probability of detecting the shift in the first wavelet or scaling function coefficient. The second term represents the probability of detecting the shift in the second wavelet or scaling function coefficient knowing that there was no detection in the first wavelet or scaling function coefficients, and so on. Equation A.5 can be written as,

$$\begin{aligned} ARL_1 &= (1 - \beta_1\overline{\beta}_1) + (1 - \beta_0\overline{\beta}_1)\beta_1\overline{\beta}_1\sum_{k=1}^{\infty}(2k+1)(\beta_0\overline{\beta}_1)^{k-1} \\ &= 1 + \frac{2\beta_1\overline{\beta}_1}{1 - \beta_0\overline{\beta}_1} \end{aligned} \tag{A.6}$$

Similarly, for $L = 2$, the ARL may be written by referring to two scales in Figure 2.5a as,

$$\begin{aligned} ARL_2 &= (1 - \beta_1) \\ &\quad +3\beta_1(1 - \beta_0\beta_2\overline{\beta}_2) \\ &\quad +5\beta_1\beta_0\beta_2\overline{\beta}_2(1 - \beta_0) \\ &\quad +7\beta_1\beta_0^2\beta_2\overline{\beta}_2(1 - \beta_0^2\overline{\beta}_2) \\ &\quad +9\beta_1\beta_0^4\beta_2\overline{\beta}_2^2(1 - \beta_0) \\ &\quad +11\beta_1\beta_0^5\beta_2\overline{\beta}_2^2(1 - \beta_0^2\overline{\beta}_2) + \ldots \end{aligned} \tag{A.7}$$

which may be simplified to,

$$ARL_2 = 1 + 2\beta_1 + \frac{2\beta_0\beta_1\beta_2\overline{\beta}_2(1 + \beta_0)}{1 - \beta_0^3\overline{\beta}_2} \tag{A.8}$$

Similarly, the ARL for $L = 3$ is,

$$ARL_3 = 1 + 2\beta_1 + 2\beta_0\beta_1\beta_2 + 2\beta_0^2\beta_1\beta_2 + \frac{2\beta_0^4\beta_1\beta_2\beta_3\overline{\beta}_3(1 + \beta_0 + \beta_0^3 + \beta_0^4)}{1 - \beta_0^7\overline{\beta}_3} \tag{A.9}$$

In general, the equation for the ARL for MSSPC-Dyadic at any depth of decomposition may be written from the wavelet decomposition diagram for that depth. Figure A.1 shows the probability of not detecting the shift for each wavelet and last scaled signal coefficient for $L = 1$ and $L = 3$. The first term in ARL_L in Equations A.6, A.8, and A.9 is always 1. Each of the next 2^{L-1} terms is obtained by traversing the grid in time until the first $\overline{\beta}_L$ is reached. Each term is 2 times the product of all the β's up to the time point. For example, the fourth term in Equation A.9 is corresponds to x_6 in Figure A.1b and is two times the product of β_1, β_2, and the two β_0's at scale $m = 1$. The last term in the ARL equation may be written as,

$$\frac{2\beta_0^{2^L-L-1}\overline{\beta}_L \prod_{i=1}^{L}\beta_i}{1-\beta_0^{2^L-1}\overline{\beta}_L}\left(1+\sum_{k=1}^{2^{L-1}-1}\beta_0^{\sum_{j=1}^{k}\zeta(j)}\right) \tag{A.10}$$

where,

$$\zeta(j)=\begin{cases}1 & \text{if } j \text{ is odd}\\ 1+log_2(j) & \text{if } j \text{ is dyadic}\\ 1+log_2\left(j-2^{\lfloor log_2(j)\rfloor}\right) & \text{if } j \text{ is even but non-dyadic}\end{cases} \tag{A.11}$$

ARL results based on this theoretical approach are compared with those from Monte-Carlo simulation in Figures 3.1 and 3.6 for uncorrelated and autocorrelated measurements. These results show a good match between theoretical and simulated results for uncorrelated measurements. For autocorrelated measurements, the assumption of uncorrelated wavelet coefficients is less accurate, and the match between the theoretical and simulated results deteriorates slightly as shown in Figure 3.6. The theoretical ARL tends to be smaller than the simulated ARL. The reason for this underestimation is discussed in the Section A.2 of this appendix.

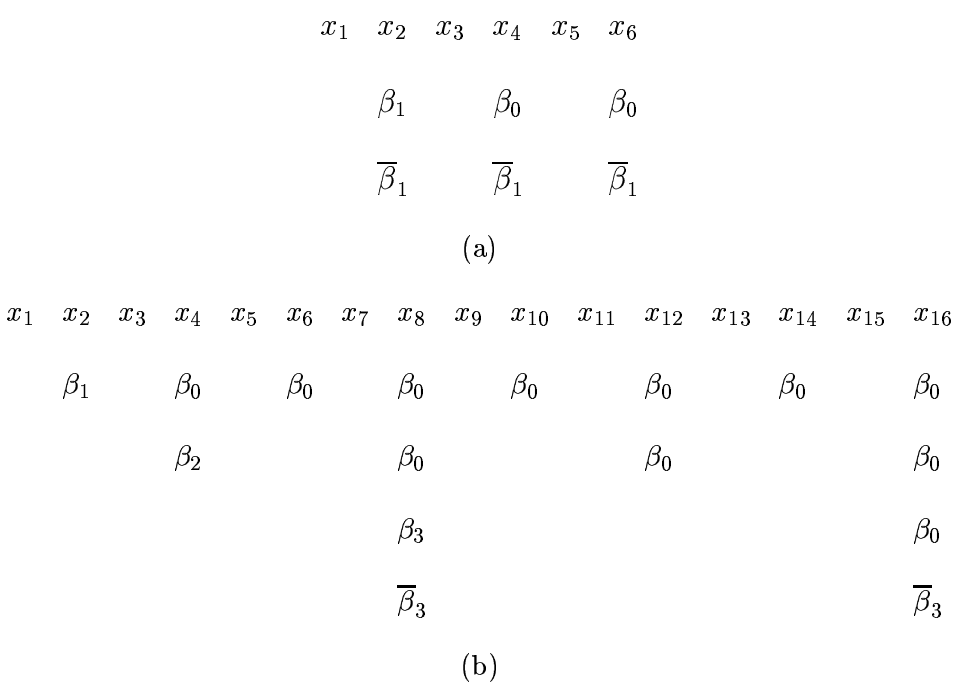


Figure A.1: Dyadic discretization grid for ARL equations. (a) Grid for obtaining Equation A.6 for $L = 1$, (b) Grid for obtaining Equation A.9 for $L = 3$.

A.2 ARL for Univariate MSSPC with Integer Discretization

When integer discretization is used, wavelet coefficients are not downsampled. As in the previous section, consider a zero-mean Gaussian random process $X[k]$ with a mean shift of size δ introduced starting from time-step k. Thus, $X[0]$ is a random variable with mean 0, whereas $X[k], k > 0$ are IID random variables with mean δ.

Let the event that the shift is missed at time $k > 0$ be denoted as E_k. The runlength (RL) is the time-step at which the fault is first detected. Thus, we have,

$$P(RL = k) = P(\overline{E_k}, E_{k-1}, E_{k-2}, \ldots, E_2, E_1) \tag{A.12}$$

The *a priori* probability of the event E_k be $\beta[k]$. Note that the notation β_m used in the previous section denotes the probability of the shift being missed at scale m, where as the notation $\beta[k]$ used here is the *overall* probability of the shift being missed at time k. Given the number of scales chosen for wavelet decomposition, L, this probability is a function of the distributions of the variables $X[k-2^L+1], X[k-2^L+2], \ldots, X[k]$. This implies that $\beta[k]$ is constant for $k \geq 2^L$. Let us choose $L = 1$ for the sake of this derivation. In that case, let p_1 denote $\beta[1]$ and let p_2 denote $\beta[k], k > 1$.

Since the wavelet decomposition without downsampling is not orthogonal, the events E_k are, in general, not independent unlike the previous section. Let us first derive an expression for the ARL assuming independence of these events.

For $k > 1$,

$$\begin{aligned} P(RL = k) &= P(\overline{E_k}, E_{k-1}, E_{k-2}, \ldots, E_2, E_1) \\ &= P(\overline{E_k}) \times P(E_{k-1}) \times P(E_{k-2}) \ldots P(E_2) \times P(E_1) \\ &= (1-p_2)p_2^{k-2}p_1 \end{aligned} \tag{A.13}$$

For $k = 1$, $P(RL = k) = (1 - p_1)$.

The ARL, which is the expected value of the random variable RL, can be derived similar to Equation A.5 as,

$$\begin{aligned} ARL_1 &= \sum_{k=1}^{\infty} k \times P(RL = k) \\ &= (1 - p_1) + \sum_{k=2}^{\infty} k \times (1 - p_2) p_2^{k-2} p_1 \\ &= (1 - p_1) + p_1 \frac{2 - p_2}{1 - p_2} \end{aligned} \tag{A.14}$$

Let $\mathbf{E}$ represent the MSSPC discrete decision function. Since $L = 1$, $\mathbf{E}$ is a function only of $\{X[k-1], X[k]\}$ at time k. If the values of the measurements were a and b at times $k-1$ and k respectively, and the shift was not detected at time k then, we have $\mathbf{E}(X[k-1] = a, X[k] = b) = 1$. The value for $\mathbf{E}(X[k-1] = a, X[k] = b) = 0$ if the shift was detected at time k. Given the values of the detection thresholds for the wavelet coefficients at each scale, as well as the thresholds for the reconstructed signal, $\mathbf{E}$ is a deterministic function in two dimensions. We have,

$$p_1 = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} \mathbf{E}(x_0, x_1) f(x_0, \sigma) f(x_1 - \delta, \sigma) dx_0 dx_1 \tag{A.15}$$

Similarly,

$$p_2 = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} \mathbf{E}(x_1, x_2) f(x_1 - \delta, \sigma) f(x_2 - \delta, \sigma) dx_1 dx_2 \tag{A.16}$$

The terms p_1 and p_2 are computed by numerical integration using the above formulation. Our ongoing work has focused on deriving closed form expressions for the above probabilities, although it is beyond the scope of this work.

The ARL values tabulated in Table A.1 show that the assumption of the events E_k being independent is not valid except for large shifts. The ARL estimates from the above analysis are compared to those generated from Monte Carlo simulations. For each experiment, the runlengths for simulated data were averaged over 100,000 instances. As the shift magnitude increases from small to medium, the signal to noise ratio increases. As a result, the detection probabilities are more correlated. This

is reflected in the increased relative error in ARL estimates. For large shifts, the probability that the shift is detected in the very first step itself grows larger and as a result, the dependence of detection probabilities at successive time-steps is less important. Hence the relative error in ARL estimates decreases for large shifts. It is worth noting that the ARL estimates derived from Equation A.13 are always lower than the actual values. This is due to the fact that, if the variable value at time $k-1$ is unusually high or unusually low, it is likely to trigger detection at time-step $k-1$ as well as time-step k. There exists, thus, a positive correlation between detection probabilities at consecutive time-steps which the Equation A.13 fails to model, and hence the under-estimation of ARL values results. In other words, the Equation A.12 assumes $P(E_k|E_{k-1}) = P(E_k)$, whereas in reality, $P(E_k|E_{k-1}) > P(E_k)$.

To obtain more accurate ARL estimates, let us now model the detection event as a one-step Markov chain. This assumption allows us to model the joint probability distribution as:

$$\begin{aligned} & P(RL = k, k > 2) \\ = \; & P(\overline{E_k}, E_{k-1}, E_{k-2}, \ldots, E_2, E_1) \\ = \; & P(\overline{E_k}|E_{k-1}) \times P(E_{k-1}|E_{k-2}) \times P(E_{k-2}|E_{k-3}) \ldots P(E_2|E_1) \times P(E_1) \\ = \; & (1-p_3)p_3^{k-3}p_4p_1 \end{aligned} \tag{A.17}$$

where $p_3 = P(E_k|E_{k-1}), k > 2$ and $p_4 = P(E_2|E_1)$. The values of p_3 and p_4 are computed by numerical integration.

$$p_4 = \frac{\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} \mathbf{E}(x_0, x_1)\mathbf{E}(x_1, x_2)f(x_0, \sigma)f(x_1 - \delta, \sigma)f(x_2 - \delta, \sigma)dx_0dx_1dx_2}{p_1} \tag{A.18}$$

Similarly,

$$p_3 = \frac{\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} \mathbf{E}(x_1, x_2)\mathbf{E}(x_2, x_3)N(x_1 - \delta, \sigma)N(x_2 - \delta, \sigma)N(x_3 - \delta, \sigma)dx_1dx_2dx_3}{p_2} \tag{A.19}$$

Table A.1: Theoretical ARL Estimation with Assumed Independence of Detection Probabilities (Equation A.13)

Shift Size (δ)	Experimental ARL	Theoretical ARL	Percent Error
0	2937.1	2840.9	3.2756
1	115.05	101.85	11.483
2	8.1691	6.5034	20.390
3	2.5130	2.3066	8.2124
4	1.7087	1.6954	0.7836

Table A.2: Theoretical ARL Estimation with 1-step Markov Model for Detection Probabilities (Equation A.19)

Shift Size (δ)	Experimental ARL	Theoretical ARL	Percent Error
0	2937.1	2975.1	-1.2930
1	115.05	115.95	-0.7775
2	8.1691	8.4056	-2.8961
3	2.5130	2.5553	-1.6826
4	1.7087	1.7101	-0.0819

Also, $P(RL = 2) = (1 - p_4)p_1$ and $P(RL = 1) = 1 - p_1$. We now have the expression:

$$\begin{aligned} ARL &= \textstyle\sum_{k=1}^{\infty} k \times P(RL = k) \\ &= (1 - p_1) + 2(1 - p_4)p_1 + \textstyle\sum_{k=3}^{\infty} k \times (1 - p_3)p_3^{k-3}p_4p_1 \\ &= (1 - p_1) + 2(1 - p_4)p_1 + p_1p_4\tfrac{3-2p_3}{1-p_3} \end{aligned} \tag{A.20}$$

Table A.2 shows the relative error of this ARL estimate. A family of theoretical and experimental ARL curves for different in-control runlengths is shown in Figure 3.3. Tables A.1 and A.2 indicate that by taking into account a 1-step Markov chain dependence, the ARL estimates improve in accuracy. Furthermore, the ARL values are always over-estimated due to the negative correlation between shift detections at time-steps k and $k - 2$, given the detection at time-step $k - 1$. in other

words, the Equation A.16 assumes $P(E_k|E_{k-1}, E_{k-2}) = P(E_k|E_{k-1})$, whereas in reality, $P(E_k|E_{k-1}, E_{k-2}) < P(E_k|E_{k-1})$. This observation indicates that the accuracy of the theoretical ARL estimation can further be improved by modeling with a 2-step Markov chain dependence. Inaccuracies due to a finite simulation sample size and numerical integration step-size are other sources of error.

BIBLIOGRAPHY

[1] D. L. Donoho, I. M. Johnstone, G. Kerkyacharian, and D. Picard, "Wavelet shrinkage: Asymptopia?," *Journal of the Royal Statistical Society- Series B*, vol. 57, pp. 41, 1995.

[2] Ian Nimmo, "Adequately address abnormal situation operations," *Chem. Eng. Prog*, vol. 91, no. 9, pp. 36–45, 1995.

[3] R. Isermann, "Process fault detection based on modeling and estimation methods - a survey," *Automatica*, vol. 20, no. 4, pp. 387, 1984.

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

[5] J. Kresta, J. F. MacGregor, and T. E. Marlin, "Multivariate statistical monitoring of process operating performance," *Canadian Journal of Chemical Engineering*, vol. 69, pp. 35, 1991.

[6] W. Ku, R. H. Storer, and C. Georgakis, "Disturbance detection and isolation by dynamic principal component analysis," *Chemometrics and Intelligent Laboratory Systems*, vol. 30, pp. 179, 1995.

[7] J. F. MacGregor, "Statistical process control of multivariate processes," in *Proceedings of the IFAC ADCHEM*, Kyoto, Japan, 1994.

[8] Western Electric, *Statistical Quality Control Handbook*, Western Electric Corporation, Indianapolis, IN, 1956.

[9] J. M. Lucas, "Combined Shewhart-CUSUM quality control schemes," *Journal of Quality Technology*, vol. 14, no. 2, pp. 51, 1982.

[10] G. C. Runger and T. R. Willemain, "Model-based and model-free control of autocorrelated processes," *Journal of Quality Technology*, vol. 27, no. 4, pp. 283, 1995.

[11] C. M. Mastrangelo and D. C. Montgomery, "SPC with correlated observations for the chemical and process industries," *Quality and Reliability Engineering International*, vol. 11, pp. 79, 1995.

[12] E. Miller and A. S. Willsky, "A multiscale approach to sensor fusion and the solution of linear inverse problems," *Applied and Computational Harmonic Analysis*, vol. 2, no. 2, pp. 127, 1995.

[13] G. Beylkin, R. Coifman, and V. Rokhlin, "Fast wavelet transforms and numerical algorithms i," *Communications on Pure and Applied Mathematics*, vol. XLIV, pp. 141, 1991.

[14] 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.

[15] 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.

[16] 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.

[17] 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.

[18] 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.

[19] 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.

[20] A. Swami, "Cramer-Rao bounds for deterministic signals in additive and multiplicative noise," *Signal Processing*, vol. 53, pp. 231, 1996.

[21] 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.

[22] 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.

[23] 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.

[24] 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.

[25] R. T. Ogden and J. D. Lynch, *Bayesian Inference in Wavelet-Based Methods*, chapter Bayesian Analysis of Change-Point Models, Springer, 1999.

[26] Z. G. Stoumbos, M. R. Reynolds, T. P. Ryan, and W. H. Woodall, "The state of statistical process control as we proceed into the 21st century," *Journal of the American Statistical Association*, vol. 95, no. 451, pp. 992, 2000.

[27] I. Daubechies, "Orthonormal bases of compactly supported wavelets," *Communications of Pure and Applied Mathematics*, vol. XLI, pp. 909, 1988.

[28] D. C. Montgomery, *Introduction to Statistical Quality Control*, John Wiley and Sons, Inc., New York, 1996.

[29] J. S. Hunter, "The exponentially weighted moving average," *Journal of Quality Technology*, vol. 18, no. 4, pp. 203, 1986.

[30] 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.

[31] 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.

[32] 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.

[33] 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.

[34] 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.

[35] 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.

[36] 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.

[37] J. R. Whiteley, *Knowledge-based interpretation of process sensor patterns*, Ph.D. thesis, The Ohio State University, 1991.

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

[39] 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.

[40] 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.

[41] 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.

[42] 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.

[43] 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.

[44] 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.

[45] 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.

[46] 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.

[47] 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.

[48] 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.

[49] 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.

[50] G. Box and J. Ramirez, "Cumulative score charts," *Quality and Reliability Engineering International*, vol. 8, pp. 17, 1992.

[51] T. J. Harris and W. H. Ross, "Statistical process control procedures for correlated observations," *The Canadian Journal of Chemical Engineering*, vol. 69, pp. 48, 1991.

[52] S. A. Vander Wiel, "Monitoring processes that wander using integrated moving average models," *Technometrics*, vol. 38, no. 2, pp. 139, 1996.

[53] J. E. Jackson, "Principal components and factor analysis: Part i- principal components," *Journal of Quality Technology*, vol. 12, no. 4, pp. 201, 1990.

[54] R. Scranton, G. C. Runger, J. B. Keats, and D. C. Montgomery, "Efficient shift detection using multivariate exponentially-weighted moving average control charts and principal components," *Quality and Reliability Engineering International*, vol. 12, pp. 165, 1996.

[55] P. Miller, R. E. Swanson, and C. E. Heckler, "Contribution plots: The missing link in multivariate quality control," *Applied Mathematics and Computer Science*, vol. 8, no. 4, pp. 775, 1998.