NEW CONTROL CHARTS FOR MONITORING UNIVARIATE AUTOCORRELATED PROCESSES AND HIGH-DIMENSIONAL PROFILES

A Thesis
Presented to
The Academic Faculty

By

Joongsup (Jay) Lee

In Partial Fulfillment
Of the Requirements for the Degree
Doctor of Philosophy in the
H. Milton Stewart School of Industrial and Systems Engineering

Georgia Institute of Technology
December 2011

NEW CONTROL CHARTS FOR MONITORING UNIVARIATE AUTOCORRELATED PROCESSES AND HIGH-DIMENSIONAL PROFILES

Approved by:

Dr. Seong-Hee Kim, Coadvisor
Committee Chair
H. Milton Stewart School of
Industrial and Systems Engineering
Georgia Institute of Technology

Dr. James R. Wilson, Coadvisor
Committee Chair
Edward P. Fitts Department of
Industrial and Systems Engineering
North Carolina State University

Dr. Xiaoming Huo
H. Milton Stewart School of
Industrial and Systems Engineering
Georgia Institute of Technology

Dr. Youngmi Hur
Department of Applied Mathematics and
Statistics
Johns Hopkins University

Dr. Jianjun Shi
H. Milton Stewart School of
Industrial and Systems Engineering
Georgia Institute of Technology

Date Approved: August 5, 2011

Oh, Day, if I squander a wavelet of thee,
A mite of my twelve hours' treasure,
The least of thy gazes or glances,
(Be they grants thou art bound to or gifts above measure)
One of thy choices or one of thy chances,
(Be they tasks God imposed thee or freaks at thy pleasure)
—My Day, if I squander such labour or leisure,
Then shame fall on Asolo, mischief on me!

(Robert Browning, *Pippa Passes*)

DEDICATION

To my girls,

Soyoun and Danbie

ACKNOWLEDGEMENTS

I would like to thank my advisors, Dr. Seong-Hee Kim and Dr. James R. Wilson, for their direction, insight, and patience as this thesis was born and matured. Their personal and professional guidance was invaluable and unforgettable.

I would also like to thank the members of my committee, Dr. Xiaoming Huo, Dr. Youngmi Hur, and Dr. Jianjun Shi. Their comments and criticism helped make this work strong and cohesive.

Finally, I cannot express enough gratitude for the support of my two girls, who made all this possible.

TABLE OF CONTENTS

DEDICATION . . . . . **iv**

ACKNOWLEDGEMENTS . . . . . **v**

LIST OF TABLES . . . . . **viii**

LIST OF FIGURES . . . . . **x**

SUMMARY . . . . . **xi**

CHAPTER 1 : INTRODUCTION **1**

CHAPTER 2 : LITERATURE REVIEW **4**

2.1 Statistical Process Control (SPC) for Univariate Processes 4

2.2 SPC for Linear and Nonlinear (Complex) Profiles 6

CHAPTER 3 : DISTRIBUTION-FREE TABULAR CUSUM CHART WITH AUTOMATED VARIANCE ESTIMATION **10**

3.1 Background 12

3.1.1 Motivating examples 13

3.2 Variance estimators 20

3.2.1 Standardized time series (STS) overlapping area estimator 20

3.2.2 "Quick-and-dirty" autoregressive (QDAR) variance estimator . . . 24

3.3 DFTC-VE: Distribution-free tabular CUSUM chart with automated variance estimation 27

3.4 Experiments 28

3.4.1 AR(1) processes 31

3.4.2 EAR(1) processes 34

3.4.3 $M/M/1$ queue waiting times 36

3.5 Conclusion 38

CHAPTER 4 : MONITORING HIGH-DIMENSIONAL PROFILES USING A WAVELET-BASED DISTRIBUTION-FREE TABULAR CUSUM CHART **40**

4.1 Background 40

4.1.1 Motivating Example 42

4.1.2 Wavelet Transform Overview 47
4.2 Wavelet-Based Distribution-Free Tabular CUSUM Profile Monitoring Chart 49
4.2.1 New High-Dimensional Profile Monitoring Chart 49
4.2.2 Dimension Reduction 51
4.2.3 Covariance-Matrix Regularization 54
4.2.4 Batch Size Determination 57
4.3 Experiments 59
4.3.1 Profiles Based on Mallat's Piecewise Smooth Function 60
4.3.2 Laser Range Sensor Data 67
4.4 Conclusion 73
CHAPTER 5 : CONTRIBUTIONS AND FUTURE RESEARCH 75
5.1 Contributions 75
5.2 Future Research 76
APPENDIX A : DERIVATION OF EQUATION (19) 77
APPENDIX B : STANDARD ERRORS FOR ARLS REPORTED IN TABLES 3–5 IN CHAPTER 3 80
APPENDIX C : STANDARD ERRORS FOR ARLS REPORTED IN TABLES 9–14 IN CHAPTER 4 84
REFERENCES 90
VITA 95

LIST OF TABLES

1 Two-sided ARLs for the classical Shewhart and tabular CUSUM charts when they are applied to the AR(1) process defined by Equation (4) with $\varphi_Y = 0.7$ and target $ARL_0 = 10{,}000$ 17

2 Two-sided ARLs for the CUSUM(LR) and DFTC charts when they are applied to the following processes with target $ARL_0 = 10{,}000$: (a) an ARMA(1,1) process defined by Equation (6) with $\varphi_1 = 0.8$, $\theta_1 = 0.16859$, and $\psi = 0.9$; and (b) an EAR(1) process defined by Equation (5) with $\varphi_Y = 0.8$ 18

3 Two-sided ARLs in terms of number of raw observations for an AR(1) process 32

4 Two-sided ARLs in terms of number of raw observations for an EAR(1) process 35

5 Two-sided ARLs in terms of number of raw observations for an $M/M/1$ process 37

6 ARLs delivered by M^* for test process ME_1 46

7 ARLs delivered by M^* for test process ME_2 47

8 Notation summary 50

9 ARLs for error vector with independent standard normal components . . . 63

10 ARLs for error vector with correlated standard normal components 64

11 ARLs for error vector with a general normal distribution 65

12 ARLs for error vector with independent shifted standard exponential components 67

13 ARLs for error vector with correlated shifted standard exponential components 68

14 ARLs for WDFTC and the profile charts (PCs) of Staudhammer [49] 72

15 Standard errors of two-sided ARLs in Table 3 for an AR(1) process 81

16 Standard errors of two-sided ARLs in Table 4 for an EAR(1) process 82

17 Standard errors of two-sided ARLs in Table 5 for an $M/M/1$ process . . . 83

18 Standard errors in Table 9 for independent standard normal components . . 84

19 Standard errors in Table 10 for correlated standard normal components . . . 85

20 Standard errors in Table 11 for correlated general normal components . . . 86

21 Standard errors in Table 12 for independent standard exponential components 87

22 Standard errors in Table 13 for correlated standard exponential components 88

23 Standard errors in Table 14 of WDFTC and profile charts (PC) of Staudhammer [49] 89

LIST OF FIGURES

1 Mallat's piecewise smooth function 44

2 Sample stream of board-thickness measurements 69

3 Common defects in lumber manufacturing 71

SUMMARY

In this thesis, we first investigate the use of automated variance estimators in distribution-free statistical process control (SPC) charts for univariate autocorrelated processes. We introduce two variance estimators—the standardized time series overlapping area estimator and the so-called quick-and-dirty autoregressive estimator—that can be obtained from a training data set and used effectively with distribution-free SPC charts when those charts are applied to processes exhibiting nonnormal responses or correlation between successive responses. In particular, we incorporate the two estimators into DFTC-VE, a new distribution-free tabular CUSUM chart developed for autocorrelated processes; and we compare its performance with other state-of-the-art distribution-free SPC charts. Using either of the two variance estimators, the DFTC-VE outperforms its competitors in terms of both in-control and out-of-control average run lengths when all the competing procedures are tested on the same set of independently sampled realizations of selected autocorrelated processes with normal or nonnormal noise components.

Next, we develop WDFTC, a wavelet-based distribution-free CUSUM chart for detecting shifts in the mean of a high-dimensional profile with noisy components that may exhibit nonnormality, variance heterogeneity, or correlation between profile components. A profile describes the relationship between a selected quality characteristic and an input (design) variable over the experimental region. Exploiting a discrete wavelet transform (DWT) of the mean in-control profile, WDFTC selects a reduced-dimension vector of the associated DWT components from which the mean in-control profile can be approximated with minimal weighted relative reconstruction error. Based on randomly sampled Phase I (in-control) profiles, the covariance matrix of the corresponding reduced-dimension DWT

vectors is estimated using a matrix-regularization method; then the DWT vectors are aggregated (batched) so that the nonoverlapping batch means of the reduced-dimension DWT vectors have manageable covariances. To monitor shifts in the mean profile during Phase II operation, WDFTC computes a Hotelling's T^2–type statistic from successive nonoverlapping batch means and applies a CUSUM procedure to those statistics, where the associated control limits are evaluated analytically from the Phase I data. We compare WDFTC with other state-of-the-art profile-monitoring charts using both normal and nonnormal noise components having homogeneous or heterogenous variances as well as independent or correlated components; and we show that WDFTC performs well, especially for local shifts of small to medium size, in terms of both in-control and out-of-control average run lengths.

CHAPTER 1

INTRODUCTION

The goal of Statistical Process Control (SPC) is to detect quickly the occurrence of shifts in the main performance characteristics of a monitored process so that possible process faults can be investigated and corrective action can be taken before quality seriously deteriorates and many defective units or services are produced. Conventional SPC techniques are effective in reducing process variation and obtaining process improvement; but they have been used primarily in manufacturing and service industries, the traditional application domains of Industrial Engineering and Operations Research (IE/OR). As the applications of IE/OR techniques have expanded to less-traditional areas such as healthcare systems, telecommunications systems, and environmental systems, new SPC methods are needed for effective large-scale monitoring of complex processes exhibiting statistically anomalous behavior. Often there is a need to monitor a large collection of system components simultaneously so that the dimensionality of the vector of target responses could easily run into the hundreds or thousands. Moreover, with the use of automated testing and sensing devices for real-time process monitoring, often the monitored process exhibits substantial nonnormality or a stochastic dependency (correlation) structure that cannot be adequately handled by conventional SPC procedures.

In this thesis, we develop distribution-free SPC methods that can be readily applied in practice to data sets with characteristics that cannot be effectively handled by conventional SPC procedures—in particular, nonnormality, high dimensionality of the response vector, heterogeneity of variance among the components of the response vector, autocorrelation between the observations of each individual component of the response vector at various points in time, and correlation between different components of the response vector at

various points in time.

In Chapter 2, we survey the literature on the following topics in SPC:

1. Monitoring univariate processes, with emphasis on methods that can handle autocorrelation or nonnormality; and

2. Monitoring profiles, with emphasis on methods that can handle high dimensionality, autocorrelation, nonnormality, or variance heterogeneity.

The literature review motivates the main contributions of this thesis, which are detailed in Chapters 3 and 4.

In Chapter 3, we formulate DFTC-VE, a distribution-free tabular CUSUM chart in which the marginal variance, the variance parameter, and the chart's control limits are estimated from a training data set automatically—that is, without the need for any intervention or trial-and-error experimentation by the user. We compare DFTC-VE's performance with the performance of competing distribution-free charts that also incorporate variance estimation. In addition, we study how the use of our automated variance estimators affects the performance of distribution-free SPC charts that are designed to use the exact values of the marginal variance and the variance parameter of the monitored process.

In Chapter 4, we turn to the development of a distribution-free monitoring scheme for high-dimensional profile-type data. Rapid advancements in data-acquisition technology, such as the development of laser range sensors, have motivated researchers and practitioners to adapt conventional statistical process control (SPC) techniques for use with large data sets called profiles that contain information about the relationship between the following: (a) a selected quality characteristic (response); and (b) an input (design, decision) variable, where the input variable can be assigned values throughout the experimental region of interest. For such data, a single realization of an in-control process may consist of n pairs $\{(x_i, y_i) : i = 1, \ldots, n\}$ of observations that can be described by the statistical

model $y_i = f_0(x_i) + \varepsilon_i$, where $f_0(\cdot)$ is a given function that defines the in-control relationship between the input variable x_i and the corresponding mean response $\mathrm{E}[y_i] = f_0(x_i)$; and ε_i is a random noise term, which is typically assumed to be independent and identically distributed (i.i.d.) normal. Chapter 4 details WDFTC, a wavelet-based distribution-free CUSUM chart that can detect shifts in the mean of a profile data set $\{(x_i, y_i) : i = 1, \ldots, n\}$, where the complexity of the functional relationship between the input variable x_i and the corresponding mean response $\mathrm{E}[y_i]$ may require a large number n of design points to yield a sufficiently accurate approximation of that relationship over the entire experimental region of interest. Moreover, WDFTC is designed to handle situations in which the noise components $\{\varepsilon_i : i = 1, \ldots, n\}$ associated with a complex profile may exhibit the following anomalous properties: (i) heterogeneity of variance across all the design points in the experimental region of interest; (ii) marked departures from normality (for example, nonzero skewness that is frequently encountered in certain types of manufacturing operations—see Stanfield et al. [48]); and (iii) substantial probabilistic dependencies (for example, nonzero correlations that arise because some of the corresponding points in the experimental region of interest are close to each other in space or time—see Lada et al. [28] and Stanfield et al. [48]).

In Chapter 5, we summarize the main findings of this research, and we make recommendations for future work.

CHAPTER 2

LITERATURE REVIEW

2.1. Statistical Process Control (SPC) for Univariate Processes

Montgomery [37] explains how to calculate control limits for the classical Shewhart and tabular CUSUM charts when those charts are used to monitor shifts in the mean of a process that consists of independent and identically distributed (i.i.d.) observations sampled from a known normal distribution. However, it is rarely the case that the exact distribution of the monitored process is known to the user of an SPC chart, and there is always the risk of simply assuming a wrong distribution. This can cause an excessive number of false alarms or an insufficient number of true alarms owing to miscalibrated control limits, ultimately resulting in excessive operating costs for the chart. Naturally, one can resort to distribution-free charts instead; but obtaining appropriate control limits becomes more difficult when those control limits must work for every possible distribution of the monitored process. For this reason, we study distribution-free charts whose control limits can be obtained by an automated technique that does not require either an excessively large training data set or cumbersome trial-and-error experimentation with such a training data set.

Beyond the problem of the monitored process having an unknown distribution (which is sometimes markedly nonnormal), in many SPC applications it is simply incorrect to assume that successive observations of the monitored process are independent—especially in applications involving relatively short time intervals between those observations or repeated measurements taken by the same operator on the same unit or workpiece. When classical SPC charts for i.i.d. observations are applied to autocorrelated processes, those charts may perform poorly in terms of the following performance measures: (a) ARL_0, the expected number of observations sampled from the in-control target process before a false

alarm is raised; and (b) ARL_1, the expected number of observations sampled from the target process with a specific out-of-control condition before a true alarm is raised [43]. Maragah and Woodall [36] show that retrospective Shewhart charts generate an increased number of false alarms when they are applied to processes with positive lag-one autocorrelation. For correlated data, Runger and Willemain [44] use nonoverlapping batch means as their basic data items and apply classical Shewhart charts designed for i.i.d. normal data, exploiting the well-known property that under broadly applicable conditions, the batch means are asymptotically i.i.d. normal as the batch size increases. For brevity, the chart of Runger and Willemain [44] is called R&W in the rest of this thesis.

Johnson and Bagshaw [21] and Kim et al. [26] develop CUSUM charts that use individual (raw, unbatched) observations as the basic data items; and in the rest of this thesis, these charts are called J&B and DFTC, respectively. Lu and Reynolds [33, 34] investigate the performance of the exponentially weighted moving average (EWMA) and CUSUM charts for a specific class of autocorrelated processes—namely, stationary and invertible first-order autoregressive–first-order moving average processes, which will simply be called ARMA(1,1) processes in the rest of this thesis. For this relatively specialized class of monitored processes, Lu and Reynolds conclude that the CUSUM and EWMA charts perform similarly when monitoring shifts in the process mean. However, the performance of such a model-based chart can be severely degraded when the hypothesized stochastic model on which the chart is based deviates significantly from the true probability law of the monitored process; and in general, definitive validation of a specific stochastic model for the monitored process can be difficult. Moreover, calibrating the control limits for a model-based chart can be extremely time-consuming unless the user is provided with an automated procedure for performing this calibration.

When developing distribution-free SPC charts, we must use one or more parameters of the monitored process, or suitable estimates of these parameters, to determine the control limits that yield the desired value of ARL_0. Such parameters include the marginal mean, the

marginal variance, and the variance parameter of the monitored process, which is the sum of covariances at all lags for the monitored process. The R&W chart uses the marginal variance of the batch means to calculate its control limits; and this quantity can be estimated by the usual sample variance of the batch means—provided the batch size is sufficiently large so that the batch means are approximately uncorrelated, and the batch count is sufficiently large to yield a stable estimator of the batch-means variance. To calculate control limits for the J&B and DFTC charts, we must know the exact values of the marginal variance and the variance parameter of the monitored process.

In Johnson and Bagshaw [21], Bagshaw and Johnson [4], and Kim et al. [26], the experimental studies of the J&B and DFTC charts were performed assuming exact knowledge of the relevant parameters of the monitored process. While such an assumption is convenient for performing simulation experiments, in most practical applications the user of an SPC chart does not know the exact values of these parameters. Instead, the process parameters must be estimated from a training data set (also called the Phase I data set) that is collected when the target process is known to be in control; then during the course of regular operation, the corresponding control limits can be used to monitor the working data set (also called the Phase II data set) for shifts that may occur in the future. When the monitored process is autocorrelated, accurately estimating the variance parameter can be substantially more difficult than accurately estimating the current mean of the process; and inaccurate variance estimators can severely degrade the performance of any SPC chart in which such estimators are used. Jensen et al. [17] provide a comprehensive literature review on the use of parameter estimation in SPC charts and recommend that the control limits should be updated as more data become available.

2.2. SPC for Linear and Nonlinear (Complex) Profiles

Kang and Albin [23] monitor a semiconductor manufacturing process that is characterized by a linear relationship between the following: (a) the pressure y in the chamber where

etching of the wafer occurs; and (b) the set point x for the mass flow controller that regulates the flow of gas into the etching chamber. Two quality characteristics (namely, the intercept a_0 and the slope a_1 in the linear statistical model $y = a_0 + a_1 x$ for $x_{\mathrm{LO}} \leq x \leq x_{\mathrm{HI}}$) are monitored using Hotelling's T^2 chart. Kim et al. [25] use two independent (univariate) exponentially weighted moving average (EWMA) charts to monitor the two regression parameters separately.

Although a linear form occurs frequently, many profile data sets (for example, radar signatures) exhibit nonlinearities and other complicated features such as discontinuities, cusps, and other types of nonsmooth, irregular behavior (Chicken et al. [9]). Woodall et al. [56] give an overview of using control charts to monitor both linear and nonlinear profile data as an application of SPC. Ding et al. [11] present a strategy for Phase I analysis of nonlinear profile data, where the Phase I data may be contaminated by out-of-control realizations of the profile; and the objective is to identify and eliminate all out-of-control realizations so that the remaining Phase I data can be used to calibrate the profile-monitoring scheme that will be applied in Phase II operation. Williams et al. [55] discuss an application of profile monitoring in the manufacturing of particle board, and they extend Hotelling's T^2 chart to monitor the coefficients of a parametric nonlinear regression model. Staudhammer et al. [51] develop profile charts for monitoring the thickness of a sawn board at selected points along the length of the board as it leaves a sawing machine in a lumber mill. They also monitor regression parameters to detect complex sawing defects. However, as Chicken et al. [9] point out, regression parameters may not adequately reflect the profile shifts, and fitting a sufficiently accurate parametric model may become difficult as well.

For most nonlinear profile-monitoring charts, the power to detect shifts in the mean of a profile can drop significantly if the monitored profile consists of a large number of components (that is, if the profile is "high-dimensional") (Fan [13]). Several dimension-reduction techniques have been proposed and incorporated into multivariate SPC charts for profile monitoring, including smoothing by regression (Kang and Albin [23]), functional

principal component analysis (Ramsay and Silverman [41]), and the use of the discrete wavelet transform (DWT) (Jin and Shi [19], Lada et al. [28], and Jeong et al. [18]).

Among such dimension-reduction techniques, wavelet-based approaches have gained popularity, especially for monitoring profiles that have highly complex or nonsmooth behavior; and such methods have been shown to be effective (Ganesan et al. [14]). These profiles are usually multiscale in nature, exhibiting substantially different critical features at different times and frequencies; see Ganesan et al. [14] and Kano et al. [24]. Jin and Shi [19, 20] use wavelets to monitor waveform signals (nonlinear profiles) from an automotive steel-stamping operation. To detect shifts in antenna data, Jeong et al. [18] apply a Hotelling's T^2–type chart to the wavelet coefficients of the observed nonlinear profiles. To monitor shifts in the mean of a nonlinear profile whose noise components are randomly sampled from a common normal distribution, Chicken et al. [9] track shifts in the mean of the corresponding discrete wavelet transform (DWT) using a likelihood ratio test to detect the change point. Chicken et al. [9] use trial-and-error simulations to estimate the upper control limit for the log-likelihood-ratio test statistic beyond which an associated series of sampled profiles is declared to be out-of-control.

Generally, a wavelet-based monitoring approach first uses wavelets to decompose a sampled profile into scaling and detail coefficients at various levels of resolution; then a data-denoising method such as principal component analysis (Jolliffe [22]) or a thresholding method (Donoho and Johnstone [12]) is used to reduce in magnitude or eliminate (that is, set to zero) all the estimated wavelet coefficients that are considered to be "unimportant" so that the surviving coefficients can be effectively monitored for possible shifts in the mean of the original sampled profiles. We exploit the capacity for parsimonious representation via wavelet coefficients in the formulation of WDFTC, a wavelet-based distribution-free tabular CUSUM chart to be presented in Chapter 4; and the wavelet-based dimension reduction is achieved by minimizing the weighted relative reconstruction error (Lada et al. [28]).

Beyond the challenge of coping effectively with the "curse of dimensionality," the assumption of i.i.d. normal errors is a severe constraint on the development of an effective wavelet-based control chart for monitoring profiles with deterministic and stochastic properties that may be irregular in some subregions of time or space. In our experience, SPC charts based on the assumption of i.i.d. normal noise components do not perform adequately when they are applied to processes whose responses (and hence the corresponding errors) exhibit substantial variance heterogeneity, pronounced nonnormality, or significant correlations (Kim et al. [26], Lee et al. [32]). Motivating examples are given in Section 4.1.1, where the profile-monitoring chart M^* of Chicken et al. [9], constructed for i.i.d. normal errors, is tested on profiles with normal or nonnormal errors and with zero or nonzero correlations between pairs of components in the same profile. Little has been done on the development and practical implementation of a monitoring scheme for high-dimensional profiles with nonnormal, correlated responses. Qiu [40] proposes a distribution-free multivariate CUSUM chart based on log-linear modeling, but the method is only applied to test processes with three quality characteristics; and as Qiu remarks, the performance of the proposed chart is unknown for high-dimensional profiles.

In Chapter 4, we develop and evaluate procedure WDFTC with a closed form of control limit based on a Brownian-motion approximation, which monitors high-dimensional profile data having any nonsingular probability distribution in order to detect various types of shifts in the mean profile. We evaluate WDFTC by applying it not only to artificial data with normal and nonnormal noise components, but also to synthetic data whose statistical model is derived from a lumber-manufacturing process.

CHAPTER 3

DISTRIBUTION-FREE TABULAR CUSUM CHART WITH AUTOMATED VARIANCE ESTIMATION

Statistical process control (SPC) charts are often used to monitor key performance characteristics of a production process, such as the process mean, and to detect any irregularities represented by gradual or sudden shifts in those quantities. While the correct detection of shifts is of great importance, timely detection of those shifts is equally critical. The application context determines the type of performance characteristic used to track the status of the process. For example, a specialist in computer-network security may want to detect network intrusions as soon as they occur by closely tracking shifts in the mean number of audit events in successive one-second time intervals [26, 27]. On the other hand, a manufacturing engineer seeking to use a coordinate measuring machine in an implementation of SPC may want to track shifts in the standard deviation of measurement error that might indicate the operator is having problems while using the machine.

Throughout this chapter, we take the performance characteristic to be the process mean; and we seek an SPC procedure for rapidly detecting shifts in the mean of an autocorrelated process, without any assumptions about the specific functional form of the probability law governing the monitored process. We let ARL_0 denote the in-control average run length—that is, the expected number of observations taken from the monitored process when it is in control (and thus yields the desired value of the process mean) before a false out-of-control alarm is raised. Similarly, let ARL_1 denote the average run length corresponding to a specific out-of-control condition—that is, the expected number of observations taken from the monitored process before a true out-of-control alarm is raised when the process mean deviates from its in-control value by a specific amount. Among several SPC charts

that yield a user-selected value of ARL_0, we prefer the chart that yields the smallest values of ARL_1 corresponding to a range of relevant out-of-control conditions.

We formulate and evaluate distribution-free statistical process control (SPC) charts for monitoring shifts in the mean of an autocorrelated process when a training data set is used to estimate the marginal variance of the process and the variance parameter (i.e., the sum of covariances at all lags). We adapt two alternative variance estimators for automated use in DFTC-VE, a distribution-free tabular CUSUM chart, based on the simulation-analysis methods of standardized time series and a simplified combination of autoregressive representation and nonoverlapping batch means. Extensive experimentation revealed that these variance estimators did not seriously degrade DFTC-VE's performance compared with its performance using the exact values of the marginal variance and the variance parameter. Moreover, DFTC-VE's performance compared favorably with that of other competing distribution-free SPC charts.

The rest of this chapter is organized as follows. Section 2 provides background information, including some motivating examples, notation, and key assumptions. Section 3 details the following alternative variance-estimation techniques that have been adapted from the simulation literature for automated use in DFTC-VE: (a) an overlapping area variance estimator based on the method of standardized time series; and (b) a less computationally intensive variance estimator based on a simplified combination of the methods of autoregressive representation and nonoverlapping batch means. Section 4 presents the proposed DFTC-VE chart for rapidly detecting shifts in the mean of an autocorrelated process. Section 5 summarizes our experimental performance evaluation of DFTC-VE versus the following: (i) distribution-free charts that use the exact values of the marginal variance and the variance parameter; and (ii) distribution-free charts that incorporate either of the variance-estimation procedures (a) or (b) above. We use the following types of test processes at various points in Sections 2 and 5: stationary first-order autoregressive (AR(1)) processes; stationary first-order exponential autoregressive (EAR(1)) processes; stationary

and invertible ARMA(1,1) processes; and an $M/M/1$ queue-waiting-time process. Section 6 contains conclusions and recommendations for future study. Appendix A contains the proof of a key result underlying DFTC-VE's variance estimator (b) above, together with tables of standard errors for all the estimated ARLs reported in Section 5.

3.1. Background

Throughout this chapter, we distinguish two sets of data: (i) a training (or Phase I) data set $\{X_i : i = 1,2,\ldots,n\}$ consisting of individual observations taken from the target process when it is known to be in control; and (ii) a working (or Phase II) data set $\{Y_i : i = 1,2,\ldots\}$ consisting of individual observations taken from the target process when it must be monitored for deviations from the in-control condition. We assume that the Phase I process $\{X_i : i = 1,2,\ldots,n\}$ is covariance stationary with $\mu \equiv \mathrm{E}[X_i]$ and $\sigma^2 \equiv \mathrm{E}\big[(X_i-\mu)^2\big]$ respectively denoting the marginal mean and variance of the process. The usual sample mean and variance of the training data set,

$$\widehat{\mu} = \overline{X}(n) = n^{-1}\sum_{i=1}^{n} X_i \tag{1}$$

and

$$\widehat{\sigma}^2 = S_n^2 = (n-1)^{-1}\sum_{i=1}^{n}(X_i-\widehat{\mu})^2, \tag{2}$$

are used to estimate μ and σ^2, respectively.

By a certain abuse of notation that should cause no confusion, we will always use μ and σ^2 to denote the marginal mean and variance of the data set that is currently under discussion, either in Phase I or Phase II. Thus in both Phase I and Phase II, we write the in-control condition as $\mu = \mu_0$, where μ_0 is the desired value of the process mean; and in Phase II, we write the out-of-control condition as $\mu = \mu_0 + \eta\sigma$, where η takes selected positive values. To obtain concrete results, we assume that $\widehat{\sigma}^2$ converges to σ^2 fast enough as the size n of the training data set grows so that $\widehat{\sigma}^2$ is a sufficiently accurate estimator of

σ^2. Moreover, we assume that a shift in the process mean in Phase II is not accompanied by any other change in the joint probability distribution governing the monitored process; in particular, we assume that the marginal variance σ^2 of the process is the same for both the in-control and out-of-control conditions.

For the DFTC-VE chart, we will also need an estimator of the variance parameter

$$\Omega^2 \equiv \lim_{n\to\infty} n\mathrm{Var}\left[\overline{X}(n)\right] = \sum_{\ell=-\infty}^{\infty} \mathrm{Cov}(X_i, X_{i+\ell}) \tag{3}$$

based on the Phase I (training) data set. As with the marginal variance, we assume that a shift in the process mean in Phase II is not accompanied by a change in the value of the variance parameter defined by Equation (3). In Section 3.2 we present two methods for computing an estimator $\widehat{\Omega}^2$ of Ω^2 from the Phase I data set. Under our assumptions about the values of σ^2 and Ω^2 being unchanged across the in-control condition and all relevant out-of-control conditions, the statistics $\widehat{\sigma}^2$ and $\widehat{\Omega}^2$ computed from the Phase I data set $\{X_i\}$ can be used as estimators of the corresponding parameters of the Phase II data set.

3.1.1 Motivating examples

To illustrate the need for new SPC charts that are specifically designed to monitor processes exhibiting pronounced autocorrelations or marked departures from normality (or both), we summarize the results of experiments in which classical SPC charts are applied to three test processes whose characteristics are typical of many practical applications. In the experimentation with the first test process, we use two classical charts that are designed for i.i.d. normal data—namely, the Shewhart chart and the tabular CUSUM chart. In the experimentation with the second and third test processes, we compare the performance of the following: (a) the model-based CUSUM chart for ARMA(1,1) processes that is due to Lu and Reynolds [34]; and (b) the distribution-free tabular CUSUM chart DFTC for all types of correlated processes that is due to Kim et al. [26].

The first motivating example is the AR(1) process,

$$Y_i = \mu + \varphi_Y(Y_{i-1} - \mu) + \varepsilon_i \text{ for } i = 1, 2, \ldots, \tag{4}$$

where: $\{\varepsilon_i : i = 1, 2, \ldots\} \overset{\text{i.i.d.}}{\sim} \mathrm{N}(0, \sigma_\varepsilon^2)$; we have $-1 < \varphi_Y < 1$ so that Equation (4) defines a process with a steady-state distribution; and we take $Y_0 \sim \mathrm{N}(\mu, \sigma^2)$, where $\sigma^2 = \sigma_\varepsilon^2/(1 - \varphi_Y^2)$, to ensure that the process starts in steady state. (For brevity and simplicity, we omit the corresponding details for the Phase I process, which is also defined by Equation (4) with $\mu = \mu_0$.)

The second motivating example is the EAR(1) process,

$$Y_i = \varphi_Y Y_{i-1} + U_i \varepsilon_i \quad \text{for} \quad i = 1, 2, \ldots, \tag{5}$$

where: $0 < \varphi_Y < 1$; the $\{U_i\}$ are the outcomes of independent Bernoulli trials having success probability $1 - \varphi_Y$ so that $\Pr\{U_i = u\} = \varphi_Y^{1-u}\big(1 - \varphi_Y\big)^u$ for $u = 0, 1$ and $i = 1, 2, \ldots$; the $\{\varepsilon_i\}$ are i.i.d. observations sampled from the exponential distribution having mean μ; and we sample the initial condition Y_0 from the exponential distribution having mean μ. With this setup, Equation (5) defines a stationary process with the exponential marginal distribution having mean μ. (We omit the corresponding details for the Phase I process, which is also defined by Equation (5) with $\mu = \mu_0$.) The EAR(1) process is widely used to model various phenomena that arise in the analysis of queueing and network processes (Billard and Mohamed [6]).

Remark 1. For a test process defined by Equation (4) or Equation (5), the autocorrelation function declines geometrically in the autoregressive parameter φ_Y,

$$\mathrm{Corr}[Y_i, Y_{i+\ell}] = \varphi_Y^{|\ell|} \text{ for } \ell = 0, \pm 1, \pm 2, \ldots \, .$$

The final motivating example is an ARMA(1,1) process,

$$Y_i = \mu + \varphi_1(Y_{i-1} - \mu) + \varepsilon_i - \theta_1 \varepsilon_{i-1} \text{ for } i = 1, 2, \ldots, \tag{6}$$

where $\{\varepsilon_i : i = 1, 2, \ldots\} \overset{\text{i.i.d.}}{\sim} \mathrm{N}(0, \sigma_\varepsilon^2)$; we have $-1 < \varphi_1, \theta_1 < 1$ so that Equation (6) defines a stationary and invertible ARMA(1,1) process; and we take the initial condition

$$\begin{bmatrix} Y_0 \\ \varepsilon_0 \end{bmatrix} \sim \mathrm{N}_2 \left(\begin{bmatrix} \mu \\ 0 \end{bmatrix}, \begin{bmatrix} \sigma^2 & \sigma_{Y\varepsilon} \\ \sigma_{Y\varepsilon} & \sigma_\varepsilon^2 \end{bmatrix} \right),$$

where

$$\sigma^2 = \mathrm{Var}[Y_i] = \left(\frac{1 + \theta_1^2 - 2\varphi_1\theta_1}{1 - \varphi_1^2} \right) \sigma_\varepsilon^2 \text{ and } \sigma_{Y\varepsilon} \equiv \mathrm{Cov}[Y_i, \varepsilon_i] = \sigma_\varepsilon^2$$

to ensure that the process defined by Equation (6) starts in steady state. For consistency with the experimental results for the ARMA(1,1) process as reported by Lu and Reynolds [34], we take the marginal variance $\sigma^2 = 1.0$, the residual variance $\sigma_\varepsilon^2 = 0.47451$, the autoregressive parameter $\varphi_1 = 0.8$, and the moving-average parameter $\theta_1 = 0.16859$; see Equations (4) and (5) of Lu and Reynolds [33]. Using the notation of Lu and Reynolds [34], we have

$$\psi = \frac{(1 - \varphi_1\theta_1)(\varphi_1 - \theta_1)}{\varphi_1(1 + \theta_1^2 - 2\varphi_1\theta_1)} = 0.9$$

so that the ARMA(1,1) process defined by Equation (6) has the autocorrelation function

$$\mathrm{Corr}(Y_i, Y_{i+\ell}) = \begin{cases} 1, & \text{if } \ell = 0, \\ \psi\varphi_1^{|\ell|}, & \text{if } \ell = \pm 1, \pm 2, \ldots. \end{cases}$$

In particular, note that the lag-one correlation in this process is $\mathrm{Corr}[Y_i, Y_{i+1}] = \psi\varphi_1 = 0.72$. (We omit the corresponding details for the Phase I process, which is also defined by Equation (6) with $\mu = \mu_0$.)

Given below are the descriptions of the classical SPC charts that were applied to our

motivating examples, with special emphasis on the parameters (or parameter estimates) of the in-control process that are required by each chart.

Classical Tabular CUSUM Charts: Set the reference-value parameter $k = 0.5$. Using the ARL approximation on p. 27 of Siegmund [47], calculate the control-limit parameter h that yields the desired value of ARL_0. Calculate $\widehat{\sigma}^2$ from the training data set $\{X_i : i = 1, 2, \ldots, 10{,}000\}$. Set the control limit $H = h\widehat{\sigma}$ and the reference value $K = k\widehat{\sigma}$. For $i = 1, 2, \ldots$, take the ith observation Y_i from the working data set; compute the corresponding two-sided tabular CUSUM

$$S^{\pm}(i) = \left\{ \begin{array}{ll} 0, & \text{if } i = 0, \\ \max\left\{0,\ S^{\pm}(i-1) \pm (Y_i - \mu_0) - K\right\}, & \text{if } i = 1, 2, \ldots; \end{array} \right\} \qquad (7)$$

and raise an out-of-control alarm if $S^{+}(i) \geq H$ or $S^{-}(i) \geq H$.

Classical Shewhart Charts: Based on the desired ARL_0, set the false-alarm probability $\alpha = 1/\mathrm{ARL}_0$. Take $z_{1-\alpha/2} = \Phi^{-1}(1 - \alpha/2)$, where $\Phi(\cdot)$ is the cumulative distribution function of the standard normal distribution. Calculate $\widehat{\sigma}^2$ from the training data set $\{X_i : i = 1, 2, \ldots, 10{,}000\}$; and set the control limit $H = z_{1-\alpha/2}\widehat{\sigma}$. For $i = 1, 2, \ldots$, take the ith observation Y_i from the working data set; and raise an out-of-control alarm if $|Y_i - \mu_0| \geq H$.

For both charts, we set $\mathrm{ARL}_0 = 10{,}000$; and we ran 4,000 independent replications of each experiment, including generation of Phase I and Phase II data, to obtain ARLs for both in-control and out-of-control situations. For the known-parameter case, we simply used the exact parameter values instead of the parameter estimates computed from the Phase I data set. Table 1 displays the results of applying the classical tabular CUSUM and Shewhart charts designed for i.i.d. normal processes to the AR(1) test process defined by Equation (4). The training data set has expected value $\mu = \mu_0$, the in-control mean, while the working data set has expected value $\mu = \mu_0 + \eta\sigma$, the out-of-control mean, where $\eta \equiv (\mu - \mu_0)/\sigma$

is the shift in mean for the out-of-control situation expressed in standard deviations of the monitored process rather than in the original measurement units for that process.

Table 1. Two-sided ARLs for the classical Shewhart and tabular CUSUM charts when they are applied to the AR(1) process defined by Equation (4) with $\varphi_Y = 0.7$ and target $ARL_0 = 10{,}000$

Shift	Tabular CUSUM		Shewhart Chart	
$\eta = (\mu - \mu_0)/\sigma$	σ Known	σ Unknown	σ Known	σ Unknown
0	74	73	11044	11024
0.25	60	61	7320	7552
0.5	40	40	3251	3375
0.75	26	26	1449	1451
1	18	18	656	692
1.5	10	10	170	174
2	7	7	58	59
2.5	5	5	22	23
3	4	4	11	10
4	3	3	3	3

As shown in Table 1, when the classical tabular CUSUM chart designed for i.i.d. normal processes was applied to the AR(1) process with $\varphi_Y = 0.7$, the chart did not yield the desired value of ARL_0. Specifically, the value of ARL_0 for the classical tabular CUSUM chart was 74 without estimation of σ^2 (so that the exact value of σ^2 was used); and the value of ARL_0 was 73 with estimation of σ^2 (so that the sample variance defined by Equation (2) was used as the estimator of σ^2). The results in Table 1 illustrate the problems that arise when the classical tabular CUSUM chart designed for i.i.d. normal processes is used to monitor shifts in the mean of an autocorrelated process. Rowlands and Wetherill [43] also discuss the sensitivity of the classical CUSUM chart to autocorrelation in the monitored process.

From Table 1, we also see that in the AR(1) test process the classical Shewhart chart delivered a value for ARL_0 that was approximately 10% above the target value of 10,000. We considered such performance to be acceptable, because it represented a rate of occurrence (and hence a cost) of false alarms that was 9% lower than expected. This approximate achievement of the desired in-control performance is mainly due to the AR(1) test process having a normal marginal distribution as required by the classical Shewhart chart. At the end of this section, we will also comment on the performance of the classical Shewhart

chart in the AR(1) test process for out-of-control conditions in the range $0 < \eta \leq 1.5$.

A model-based SPC chart can perform poorly when the monitored process does not follow the hypothesized joint probability distribution for which the chart's control limits are specifically designed. Table 2 illustrates the need for distribution-free charts by comparing the performance of the following: (a) the model-based CUSUM chart for ARMA(1,1) processes that is due to Lu and Reynolds [34] and is labeled CUSUM(LR) in the table; and (b) the distribution-free CUSUM chart DFTC for correlated processes of all types that is due to Kim et al. [26]. Table 2 summarizes the results of using the CUSUM(LR) and DFTC charts to monitor the ARMA(1,1) and EAR(1) test processes defined by Equations (6) and (5), respectively. It should be noted that the results for DFTC in Table 2 are based on using a batch size of $m = 3$ in the ARMA(1,1) process and a batch size of $m = 4$ in the EAR(1) process; see Section 3.2 of Kim et al. [26] and Section 3 of this chapter for a detailed explanation of the reasons for using batch means rather than raw (unbatched) observations when the DFTC chart is applied to a process with a lag-one correlation between successive raw observations that exceeds 0.5.

Table 2. Two-sided ARLs for the CUSUM(LR) and DFTC charts when they are applied to the following processes with target $ARL_0 = 10{,}000$: (a) an ARMA(1,1) process defined by Equation (6) with $\varphi_1 = 0.8$, $\theta_1 = 0.16859$, and $\psi = 0.9$; and (b) an EAR(1) process defined by Equation (5) with $\varphi_Y = 0.8$

Shift	ARMA(1,1)		EAR(1)	
$\eta = (\mu - \mu_0)/\sigma$	CUSUM(LR)	DFTC	CUSUM(LR)	DFTC
0	10001	11918	6450	11622
0.5	315	413	314	441
1.0	125	190	123	201
1.5	77	122	75	129
2	56	93	53	96
2.5	43	74	42	76
3	36	61	34	64
4	26	46	25	48
5	21	37	20	39

Lu and Reynolds [34] formulate the CUSUM(LR) chart based on the assumption that

the ARMA(1,1) model defined by Equation (6) is generally an adequate model of the monitored process. Of course this chart performs well when it is used to monitor an ARMA(1,1) process—provided that the user has performed an extensive set of trial-and-error simulation experiments to estimate the control limits that yield the target value of ARL_0 for the particular ARMA(1,1) process at hand. In practice, however, there is no guarantee that the monitored process can always be adequately modeled by an ARMA(1,1) process; and the CUSUM(LR) chart can perform poorly in terms of ARL_0 if the stochastic behavior of the monitored process deviates significantly from that of an ARMA(1,1) process.

Table 2 shows that when the CUSUM(LR) chart was used to monitor the EAR(1) process defined by Equation (5), the CUSUM(LR) chart delivered a value for ARL_0 that was 36% below the target value. We considered such performance to be unacceptable, because it represented a rate of occurrence (and hence a cost) of false alarms that was 55% higher than expected. On the other hand, the DFTC chart delivered a value for ARL_0 that was 12% above the target value, which represented a rate of occurrence of false alarms that was 14% lower than expected. Although admittedly DFTC delivered values of ARL_1 that were somewhat larger (in absolute terms) than the corresponding values of ARL_1 delivered by CUSUM(LR), we concluded that DFTC's overall performance was more robust than that of CUSUM(LR).

Returning to the AR(1) test process with $\varphi_Y = 0.7$, we also compare the results in Table 1 for the classical Shewhart chart with the corresponding results in the first part of Table 4 of Kim et al. [26] for the DFTC chart; and we see that the DFTC chart significantly outperformed the classical Shewhart chart for the selected out-of-control conditions in the range $0 < \eta \leq 1.5$. Thus the results in Table 1 also illustrate the inability of the classical Shewhart chart designed for i.i.d. normal processes to monitor properly shifts in the mean of an autocorrelated process when those shifts have small or medium magnitudes.

We believe that the experimental results discussed in this section provide compelling evidence of the need for new distribution-free SPC charts that can handle autocorrelated processes and that automatically provide the necessary variance estimators. Because the DFTC chart outperformed not only the classical Shewhart and CUSUM charts but also the CUSUM(LR) chart in almost all our test processes, we sought to extend the DFTC chart

by incorporating a suitable variance estimator from the literature on steady-state simulation analysis.

3.2. Variance estimators

3.2.1 Standardized time series (STS) overlapping area estimator

Alexopoulos et al. [2] show that the so-called standardized time series (STS) overlapping area estimator for Ω^2 outperforms a number of other variance estimators in terms of bias and variance; therefore we chose to incorporate this variance estimator into a version of DFTC-VE. First, we form $n-m+1$ overlapping batches, each of size m, from the training data so that the ith batch consists of the observations $\{X_{i+j} : j = 0,\ldots,m-1\}$ for $i = 1,2,\ldots,n-m+1$. Although the sample-size-to-batch-size ratio $b \equiv n/m$ is fixed, we let the batch size m increase so that the overall sample size $n = bm$ also increases; and all this data must be taken from the training data set to compute the STS estimator of Ω^2. The sample mean computed from the ith overlapping batch with (intermediate) batch size j is represented by

$$\overline{X}_i^*(j) \equiv \frac{1}{j}\sum_{\ell=0}^{j-1} X_{i+\ell} \text{ for } i = 1,2,\ldots,n-m+1 \text{ and } j = 1,2,\ldots,m.$$

Using the weight function

$$f(t) = \sqrt{840}\left(3t^2 - 3t + 1/2\right) \text{ for } t \in [0,1], \tag{8}$$

we define

$$Z_i^*(f;m) \equiv \frac{1}{m^{3/2}}\sum_{j=1}^{m} f\left(\tfrac{j}{m}\right) j\left[\overline{X}_i^*(m) - \overline{X}_i^*(j)\right] \text{ for } i = 1,2,\ldots,n-m+1 \tag{9}$$

as the weighted area under the standardized time series computed from the ith overlapping batch of size m in the training data set; see Equations (2) through (4) of Alexopoulos et al.

[2]. Finally, the overlapping STS area estimator for Ω^2 is defined by

$$\widehat{\Omega}^2_{\mathrm{A}} \equiv \frac{1}{n-m+1} \sum_{i=1}^{n-m+1} \left[Z_i^*(f;m) \right]^2. \tag{10}$$

See Sections 4.1 and 4.2 of Alexopoulos et al. [2] for a complete discussion of standardized time series and variance estimators that are computed from overlapping batches.

We use the weight function defined by Equation (8) to obtain the following key properties provided $b > 1$:

$$\left.\begin{array}{rcl} \mathrm{E}\left[\widehat{\Omega}^2_{\mathrm{A}}\right] & = & \Omega^2 + o(1/m) \\ \mathrm{Var}\left[\widehat{\Omega}^2_{\mathrm{A}}\right] & \doteq & \dfrac{0.819\Omega^4}{b} \end{array}\right\} \text{ as } m \to \infty; \tag{11}$$

in particular, see Examples 1 and 5 of Alexopoulos et al. [2]. Other weight functions and even other STS estimators for Ω^2 are available for use; our selection here has been based on the comparatively good analytical and empirical performance of the overlapping area estimator defined by Equation (10) with the weight function defined by Equation (8). The remaining unresolved problem is the selection of the batch size m, which affects not only the bias and variance of the STS estimator $\widehat{\Omega}^2_{\mathrm{A}}$ as shown in Equation (11) but also the convergence of $\widehat{\Omega}^2_{\mathrm{A}}$ to its limiting distribution. Asymptotically as $m \to \infty$, the STS estimator $\widehat{\Omega}^2_{\mathrm{A}}$ has approximately a scaled chi-squared distribution with mean Ω^2 as explained in Section 4.2 of Alexopoulos et al. [1].

For use with DFTC-VE, we propose an automated batch-size determination algorithm that uses the same sequential procedure as in Lada and Wilson [30] and Lada et al. [29]; but instead of using *nonoverlapping* batch means as the basic data items to be tested for independence and normality, we use the "STS-weighted-area" statistics similar to those defined in Equation (9) that are computed from nonoverlapping batches in the training data set. For $i = 1, 2, \ldots, b$, the ith nonoverlapping batch of size m in the training data set is $\{X_{(i-1)m+j} : j = 1, 2, \ldots, m\}$; and we seek a sufficiently large value of m so that the corresponding "STS-weighted-area" statistics $\{Z_i(f;m) : i = 1, 2, \ldots, b\}$ computed from Equations (12) and (13) below approximately constitute a random sample from a normal distribution. Such a batch size m is large enough to ensure approximate convergence of

the final overlapping area estimator $\widehat{\Omega}^2_{\mathrm{A}}$ defined by Equation (10) to its appropriate limiting distribution. See Theorem 3 of Alexopoulos et al. [2] for the exact asymptotic distribution of $\widehat{\Omega}^2_{\mathrm{A}}$ as $m \to \infty$. A formal statement of our batch-size determination algorithm for the STS overlapping area estimator is given on p. 23 below.

Remark 2. Using the final batch size m delivered by the algorithm given above, we see that for $i = 1, 2, \ldots, n-m+1$, the quantity $Z_i^*(f;m)$ given by Equation (9) is the signed, weighted area under the standardized time series defined on the ith overlapping batch $\{X_{i+j} : j = 0, 1, \ldots, m-1\}$ of size m. The weight function $f(\cdot)$ is carefully selected to ensure that $\mathrm{E}\big[Z_i^*(f;m)\big] = 0$ and $\lim_{m\to\infty} \mathrm{E}\Big\{ \big[Z_i^*(f;m)\big]^2 \Big\} = \Omega^2$ for $i = 1, 2, \ldots, n-m+1$. The final STS variance estimator $\widehat{\Omega}^2_{\mathrm{A}}$ defined by Equation (10) is the average of all the $\Big\{ \big[Z_i^*(f;m)\big]^2 : i = 1, 2, \ldots, n-m+1 \Big\}$ taken over the entire training data set.

Remark 3. The first part of Equation (11) reveals that the bias of the STS variance estimator is $o(1/m)$. To ensure sufficiently small bias as well as adequate convergence of the STS estimator to its limiting distribution, the final batch size delivered by the batch-size algorithm on p. 23 below is usually taken to be three times the batch size required to pass the independence test in Step **[3]** and the normality test in Step **[5]**. This inflation of the final batch size also improves the adequacy of the Brownian-motion approximation to the behavior of the CUSUM statistics used by the J&B and DFTC charts. If the size of the training data set, n, is not large enough to ensure that the $\big\{Z_i(f;m) : i = 1, 2, \ldots, b\big\}$ defined by Equation (13) pass both the independence test in Step **[3]** and the normality test in Step **[5]**, then we take the final batch size $m \leftarrow \lfloor n/20 \rfloor$, which ensures that $b = n/m \geq 20$. This design for the batch-size algorithm ensures that no matter how the algorithm terminates, the final STS variance estimator $\widehat{\Omega}^2_{\mathrm{A}}$ in Equation (10) has approximately a scaled chi-squared distribution with at least 48 degrees of freedom (Alexopoulos et al. [1]).

Remark 4. In the batch-size algorithm given on p. 23 below, we used $\alpha_{\mathrm{ran}} = 0.20$ for the size of the randomness test, and we used the normality-test parameters $\alpha_{\mathrm{nor}}(1) = 0.05$, and $\gamma = 0.184206$. These parameter values are based on the extensive computational experience of Lada and Wilson [30] in testing certain sets of nonoverlapping batch means for randomness and normality before the final delivered set of nonoverlapping batch means

Batch-Size Algorithm for the STS Overlapping Area Estimator of Ω^2

[1] Take an initial sample of size $n' \leftarrow 4{,}096$ from the training data set of size $n \leftarrow 10{,}000$; and divide the initial sample into $b \leftarrow 256$ adjacent nonoverlapping batches of size $m \leftarrow 16$. Set the randomness test size $\alpha_{\text{ran}} \leftarrow 0.20$. Set the initial normality test size $\alpha_{\text{nor}}(1) \leftarrow 0.05$, the normality test parameter $\gamma \leftarrow 0.184206$, and the normality test iteration counter $k \leftarrow 1$.

[2] Compute the following statistics from the b nonoverlapping batches of the current batch size m, where the weight function $f(\cdot)$ is defined by Equation (8):

$$\overline{X}_i(j) \equiv \frac{1}{j} \sum_{\ell=1}^{j} X_{m(i-1)+\ell} \quad \text{for} \quad i = 1,2,\ldots,b \quad \text{and} \quad j = 1,2,\ldots,m, \tag{12}$$

$$Z_i(f;m) \;\equiv\; \frac{1}{m^{3/2}} \sum_{j=1}^{m} f\left(\tfrac{j}{m}\right) j \left[\overline{X}_i(m) - \overline{X}_i(j)\right] \quad \text{for} \quad i = 1,2,\ldots,b. \tag{13}$$

[3] Apply the von Neumann test for randomness [54] to the $\{Z_i(f;m) : i = 1,2,\ldots,b\}$ defined by Equation (13) using the significance level α_{ran}. If the randomness test is passed, then go to Step **[5]**; otherwise, go to Step **[4]**.

[4] Increase the batch size m and update the overall sample size n' according to $m \leftarrow \lfloor \sqrt{2}m \rfloor$ and $n' \leftarrow bm$, respectively, where $\lfloor \cdot \rfloor$ is the floor function.

- **(a)** If $n' > n$, then return the final batch size $m \leftarrow \lfloor n/20 \rfloor$ and stop.
- **(b)** Otherwise, obtain the required additional observations from the training data set and go to Step **[2]**.

[5] Apply the Shapiro-Wilk normality test [46] to the current set of statistics $\{Z_i(f;m) : i = 1,2,\ldots,b\}$ defined by Equation (13) using the significance level $\alpha_{\text{nor}}(k) \leftarrow \alpha_{\text{nor}}(1) \exp\left[-\gamma(k-1)^2\right]$. If the normality test is passed with the current batch size m, then return the final batch size $m \leftarrow 3m$ and stop; otherwise go to Step **[6]**.

[6] Increase the normality-test iteration counter k, batch size m, and overall sample size n' according to $k \leftarrow k+1$, $m \leftarrow \lfloor \sqrt{2}m \rfloor$, and $n' \leftarrow bm$, respectively.

- **(a)** If $n' > n$, then return the final batch size $m = \lfloor n/20 \rfloor$ and stop.
- **(b)** Otherwise, obtain the required additional observations from the training data set; recompute the nonoverlapping batch statistics defined by Equations (12) and (13); and go to Step **[5]**.

is used to compute a spectral estimator of the variance parameter of a simulation output process.

3.2.2 "Quick-and-dirty" autoregressive (QDAR) variance estimator

As an alternative to the STS variance estimator, we develop the so-called "quick-and-dirty" autoregressive (QDAR) variance estimator, which is based on a simplified combination of the methods of autoregressive representation and nonoverlapping batch means that also underlies the simulation analysis procedures ASAP3 (Steiger et al. [52]) and SBatch (Lada et al. [29]). By design the QDAR variance estimator is simpler to compute and easier to automate than the STS variance estimators. If we can find a sufficiently large batch size m such that the nonoverlapping batch means $\left\{ \overline{X}_i(m) : i = 1,2,\ldots,b \right\}$ defined by Equation (12) and computed from the training data set are adequately modeled by an AR(1) process,

$$\overline{X}_i(m) = \mu + \varphi_{\overline{X}(m)} \left[\overline{X}_{i-1}(m) - \mu \right] + \varepsilon_i(m) \;\text{ for }\; i = 1,2,\ldots, \tag{14}$$

with $\left|\varphi_{\overline{X}(m)}\right| < 1$ and $\{\varepsilon_i(m) : i = 1,2,\ldots\} \overset{\text{i.i.d.}}{\sim} \mathrm{N}\left[0, \sigma^2_{\varepsilon(m)}\right]$, then the variance parameter of the AR(1) process defined by Equation (14) is given by

$$\begin{aligned} \Omega^2_{\overline{X}(m)} &= \sum_{\ell=-\infty}^{\infty} \mathrm{Cov}\left[\overline{X}_i(m), \overline{X}_{i+\ell}(m) \right] \\ &= \Omega^2 / m = \mathrm{Var}\left[\overline{X}(m) \right] \left[\frac{1 + \varphi_{\overline{X}(m)}}{1 - \varphi_{\overline{X}(m)}} \right]. \end{aligned} \tag{15}$$

As detailed below, the key idea underlying the QDAR variance estimator is to do the following: (i) determine a batch size m sufficiently large to ensure the approximate validity of the AR(1) model defined by Equation (14) as a characterization of the behavior of the nonoverlapping batch means $\left\{ \overline{X}_i(m) \right\}$ computed from the training data set; (ii) compute approximately unbiased estimators of $\mathrm{Var}\left[\overline{X}(m) \right]$ and $\varphi_{\overline{X}(m)}$ from the training data set; and (iii) insert these estimators into Equation (15) to obtain a simplified, intuitively appealing estimator of the variance parameter Ω^2 for the original (unbatched) process.

In developing and testing the DFTC chart, Kim et al. [26] find that in practice the adequacy of the Brownian-motion approximation to the behavior of the CUSUM on which DFTC is based requires a lag-one correlation of at most 0.5 in the time series of basic data items used to compute that CUSUM. To compensate for the additional uncertainty introduced by estimation of the variance parameter, in the formulation of the QDAR variance estimator we impose the more-stringent upper bound

$$\varphi_{\overline{X}(m)} = \mathrm{Corr}\left[\overline{X}_i(m), \overline{X}_{i+1}(m)\right] \leq 0.4 \tag{16}$$

on the basic data items used to compute the CUSUM on which DFTC-VE is based—that is, we require a sufficiently large batch size m so that the nonoverlapping batch means $\left\{\overline{X}_i(m) : i = 1,2,\ldots,b\right\}$ computed from the training data set satisfy Equation (16) with probability close to one. For additional considerations justifying Equation (16), see p. 77 of Bagshaw and Johnson [4]. Furthermore, Steiger et al. [52] and Lada et al. [29] find that if the batch size m is sufficiently large to ensure that $\varphi_{\overline{X}(m)} \leq 0.8$, then their variance estimators similar to the QDAR variance estimator are sufficiently stable in practice to yield highly reliable and accurate confidence-interval estimators of the steady-state mean. On the basis of all these considerations, we concluded that if we take the batch size large enough to satisfy Equation (16) with probability close to one, then we should be able to use the QDAR variance estimator effectively in distribution-free SPC charts requiring an estimator of the variance parameter.

When the lag-one correlation between individual (unbatched, raw) observations is greater than 0.5, Kim et al. [26] recommend the use of batching in Phase II to ensure that the lag-one correlation in the data to be monitored does not exceed 0.5; and thus in Phase II, we use the batch size m determined in Phase I so as to satisfy Equation (16) with probability close to one. A formal statement of the algorithm for computing the QDAR variance estimator is given on pp. 26–27 below.

Remark 5. In the QDAR algorithm given above, the parameter values $b_{\min} = 1{,}024$ and $\alpha_{\mathrm{cor}} = 0.01$ are based on the extensive computational experience of Steiger et al. [52] with the ASAP3 procedure and Lada et al. [29] with the SBatch procedure. In ASAP3

QDAR Algorithm for Estimating Ω^2

[1] Choose the initial batch size m, where $m = 1$ by default. Choose the initial batch count $b \geq b_{\min}$, where $b_{\min} = 1{,}024$ by default. Take the initial sample of size $n' = mb$ from the training data set of size n, where $n = 10{,}000$ by default. Also set the size α_{cor} of the test for acceptable lag-one correlation in the batch means and the upper bound ζ on the lag-one correlation in Equation (16) as follows: $\alpha_{\text{cor}} \leftarrow 0.01$ and $\zeta \leftarrow 0.4$.

[2] For the current sample of size $n' = mb$ taken from the training data set, calculate the sample statistics for $\left\{ \overline{X}_i(m) : i = 1, 2, \ldots, b \right\}$, the current set of b nonoverlapping batch means based on batch size m,

$$\left.\begin{aligned} \overline{X}(m,b) \equiv b^{-1} \sum_{i=1}^{b} \overline{X}_i(m), \quad S^2(m,b) \equiv (b-1)^{-1} \sum_{i=1}^{b} \left[\overline{X}_i(m) - \overline{X}(m,b) \right]^2, \\ \widehat{\rho}_{\overline{X}(m)}(m,b) \equiv (b-1)^{-1} \sum_{i=1}^{b-1} \left[\overline{X}_i(m) - \overline{X}(m,b) \right] \left[\overline{X}_{i+1}(m) - \overline{X}(m,b) \right] \Big/ S^2(m,b), \\ \widehat{\varphi}_{\overline{X}(m)} \equiv 2\widehat{\rho}_{\overline{X}(m)}(m,b) - \left[\widehat{\rho}^{(1)}_{\overline{X}(m)}(m,b/2) + \widehat{\rho}^{(2)}_{\overline{X}(m)}(m,b/2) \right] \Big/ 2, \end{aligned}\right\} \quad (17)$$

where $\widehat{\rho}_{\overline{X}(m)}(m,b)$ denotes the standard lag-one sample correlation between the batch means based on b nonoverlapping batches each of size m; and we let $\widehat{\rho}^{(1)}_{\overline{X}(m)}(m,b/2)$ and $\widehat{\rho}^{(2)}_{\overline{X}(m)}(m,b/2)$ respectively denote similar estimators of the lag-one correlation between the batch means based on the first and last $b/2$ nonoverlapping batches each of size m.

[3] If $\widehat{\varphi}_{\overline{X}(m)} \leq \sin\left[\sin^{-1}(\zeta) - z_{1-\alpha_{\text{cor}}} / \sqrt{b} \right]$ (where $z_{1-\alpha_{\text{cor}}} = \Phi^{-1}(1-\alpha_{\text{cor}}) = 2.33$), then go to Step **[4]**; else estimate the required batch size as follows:

$$\mathscr{Q} \leftarrow \left\lceil \ln\left\{ \sin\left[\sin^{-1}(\zeta) - z_{1-\alpha_{\text{cor}}} / \sqrt{b} \right] \right\} \Big/ \ln\left[\widehat{\varphi}_{\overline{X}(m)} \right] \right\rceil ; \; m \leftarrow \lceil \text{mid}\{1.1, \mathscr{Q}, 2\} m \rceil,$$

where $\lceil \cdot \rceil$ is the ceiling function and $\text{mid}\{u,v,w\}$ denotes the median of the real numbers u, v, w; retrieve additional data from the training data set and update n' if necessary to ensure that $b \leftarrow \lfloor n'/m \rfloor$ satisfies $b \geq b_{\min}$; and go to Step **[2]**.

[4] Compute the estimator of the variance parameter for the batch means,

$$\widehat{\Omega}^2_{\overline{X}(m)} = S^2(m,b) \left[\frac{b-1}{b - \widehat{C}(m,b)} \right] \left[\frac{1 + \widehat{\varphi}_{\overline{X}(m)}}{1 - \widehat{\varphi}_{\overline{X}(m)}} \right],$$

where

$$\widehat{C}(m,b) = \frac{1+\widehat{\varphi}_{\overline{X}(m)}}{1-\widehat{\varphi}_{\overline{X}(m)}} - \frac{2\widehat{\varphi}_{\overline{X}(m)}\left[1-\widehat{\varphi}^{b}_{\overline{X}(m)}\right]}{b\left[1-\widehat{\varphi}_{\overline{X}(m)}\right]^2};$$

and deliver the final estimator of the variance parameter for the original process,

$$\widehat{\Omega}^2_{\mathrm{Q}} \equiv m\widehat{\Omega}^2_{\overline{X}(m)}. \tag{18}$$

and SBatch, the final delivered set of nonoverlapping batch means is used to compute a correlation-adjusted batch-means estimator of the variance parameter of a simulation-generated output process.

Remark 6. In Step **[3]** of the QDAR algorithm, we take the parameter values $\alpha_{\mathrm{cor}} = 0.01$ and $\zeta = 0.4$ so that with approximate probability $1-\alpha_{\mathrm{cor}} = 0.99$, the lag-one correlation of the final set of batch means is at most 0.4 as required by Equation (16).

Remark 7. The formula for $\widehat{\varphi}_{\overline{X}(m)}$ in the third line of Equation (17) is the jackknifed estimator of $\varphi_{\overline{X}(m)}$ and thus has bias of the form $O(b^{-2})$, whereas the standard correlation estimator $\widehat{\rho}_{\overline{X}(m)}(m,b)$ has bias of the form $O(b^{-1})$. This property coupled with the requirement $b \geq b_{\min} = 1{,}024$ ensures that $\widehat{\varphi}_{\overline{X}(m)}$ is an approximately unbiased estimator of $\varphi_{\overline{X}(m)}$.

Remark 8. The final variance estimator (18) follows from (15), Remark 7, and the property

$$S^2(m,b)\left[\frac{b-1}{b-\widehat{C}(m,b)}\right] \text{ is an (approximately) unbiased estimator of } \mathrm{Var}\left[\overline{X}(m)\right]. \tag{19}$$

See Appendix A for a derivation of (19).

3.3. DFTC-VE: Distribution-free tabular CUSUM chart with automated variance estimation

Using the estimators given in Equations (1), (2), (10), and (18), we modify the DFTC chart of Kim et al. [26] to incorporate automated parameter estimation based on a training data set to yield procedure DFTC-VE. The formal statement of DFTC-VE is given on p. 28.

DFTC-VE: Distribution-Free Tabular CUSUM Chart with Variance Estimation

[1] Compute $\widehat{\mu}$ and $\widehat{\sigma}^2$ from the Phase I data set using Equations (1) and (2), respectively. Compute $\widehat{\Omega}^2$ from the Phase I data set using either Equation (10) or Equation (18).

[2] Set $K = k\widehat{\sigma}$; choose the desired value of the two-sided ARL_0; and calculate H, the solution to the equation

$$\frac{\widehat{\Omega}^2}{2K^2}\left\{\exp\left[\frac{2K\left(H+1.166\widehat{\Omega}\right)}{\widehat{\Omega}^2}\right]-1-\frac{2K\left(H+1.166\widehat{\Omega}\right)}{\widehat{\Omega}^2}\right\}=2\text{ARL}_0. \tag{20}$$

[3] For $i = 1,2,\ldots,$ take the ith observation Y_i from the Phase II data set; update the CUSUMs $S^{\pm}(i)$ in Equation (7); and raise an out-of-control alarm if $S^{+}(i) \geq H$ or $S^{-}(i) \geq H$.

Remark 9. For the choice of the reference parameter value, Kim et al. [26] recommend $k = 0.1$. A search method (such as the bisection algorithm) can be used to solve Equation (20) for the control limit H.

3.4. Experiments

We compare the performance of the following distribution-free SPC charts in terms of their average run lengths: (i) the DFTC-VE chart using the estimators $\widehat{\sigma}^2$ and $\widehat{\Omega}^2$ computed from Phase I together with the corresponding estimated control limits so as to operate the chart in Phase II; and (ii) the DFTC chart using the exact values of σ^2 and Ω^2 together with the corresponding exact values of the control limits so to operate the chart in Phase II. We also compare the performance of the DFTC-VE chart with that of other distribution-free SPC charts—namely, the R&W chart and the J&B chart—to which we added Phase I estimation of the required process parameters, as outlined below.

R&W Chart with Estimated Marginal Variance: From the Phase I data set, determine the smallest batch size m such that

$$\text{Corr}\left[\overline{X}_i(m),\overline{X}_{i+1}(m)\right] \leq 0.10, \tag{21}$$

where the Phase I nonoverlapping batch means $\{\overline{X}_i(m) : i = 1,2,\ldots,b\}$ are computed as in the first line of Equation (17); see pp. 288–289 of Runger and Willemain [44]. Choose the desired value of ARL_0 and find z_{on} such that

$$\frac{m}{1-\Phi(z_{\mathrm{on}})+\Phi(-z_{\mathrm{on}})} = \mathrm{ARL}_0. \tag{22}$$

Compute $\widehat{\mathrm{Var}}\big[\overline{X}(m)\big]$, the estimated variance of the batch means, from the entire Phase I data set in the same way that the statistic $S^2(m,b)$ is computed in the first line of Equation (17). For $j = 1,2,\ldots$, raise an out-of-control alarm after observation $i = jm$ in Phase II if

$$\big|\overline{Y}_j(m) - \mu_0\big| \geq z_{\mathrm{on}} \cdot \sqrt{\widehat{\mathrm{Var}}\big[\overline{X}(m)\big]}. \tag{23}$$

J&B Chart with Estimated Variance Parameter: Compute $\widehat{\Omega}^2$ from the Phase I data set using either Equation (10) or Equation (18). Set

$$H = \sqrt{2\mathrm{ARL}_0\widehat{\Omega}^2}.$$

For $i = 1,2,\ldots$, take the ith observation Y_i from the Phase II data set; compute the updated CUSUMs

$$S^{\pm}(i) = \left\{ \begin{array}{ll} 0, & \text{for } i = 0, \\ \max\big\{S^{\pm}(i-1) \pm (Y_i - \mu_0), 0\big\} & \text{for } i = 1,2,\ldots; \end{array} \right\}$$

and raise an out-of-control alarm if $S^+(i) \geq H$ or $S^-(i) \geq H$.

In our experimental performance evaluation, we considered stationary test processes with normal and nonnormal marginals. We used a single test process with normal marginals, the AR(1) process defined by Equation (4), and the following test processes with nonnormal marginals: (i) the EAR(1) process defined by Equation (5); and (ii) the queue-waiting-time process in a steady-state $M/M/1$ queueing system as detailed in Section 3.4.3 below. For each distribution-free chart, we took the target value of 10,000 for ARL_0; and we performed

4,000 independent replications of each chart applied to each test process. On each replication of each SPC chart, the training data set had marginal mean $\mu = \mu_0$, the in-control mean; and the working data set had marginal mean $\mu = \mu_0 + \eta\sigma$, where the shift parameter η took the values 0, 0.25, 0.5, 0.75, 1, 1.5, 2, 2.5, 3, and 4. Thus for each SPC chart, the Phase II experimentation included the in-control condition as well as out-of-control conditions corresponding to "small," "medium," and "large" shifts in the mean.

Throughout this section, we let m_1 and m_2 denote the batch sizes used in Phase I and Phase II, respectively. Note that a new set of Phase I data was obtained on each replication; and thus our batch-size determination algorithms delivered a different batch size on each replication. Over the 4,000 independent replications of each SPC chart applied to each test process, we computed $\overline{m}_1$ and $\overline{m}_2$, the average batch sizes used in Phase I and Phase II, respectively. On each replication of the R&W chart applied to a test process, in Phase I we use the batch size m_1 that achieves an approximate lag-one correlation of 0.1 for the batch means; and then in Phase II we took $m_2 = m_1$. In particular, for the AR(1) and EAR(1) testing processes, we use m_1 from Runger and Willemain [44]; and for the queue-waiting-time process, we use m_1 from Kim et al. [26], who study the required batch size for highly nonnormal, correlated processes.

For the CUSUM-type charts with variance estimation (J&B and DFTC-VE), the batch-size determination algorithm varied depending on which variance estimator was used. On each replication of an SPC chart using the STS variance estimator $\widehat{\Omega}^2_{\mathrm{A}}$, in Phase I we used the final batch size m_1 delivered by the algorithm in Section 3.2.1; and in Phase II we used the batch size $m_2 = 1$ so as to compensate for the computational overhead associated with the STS batch-size determination algorithm used in Phase I. Thus the SPC charts with the STS variance estimator always used raw (unbatched) observations in Phase II. On each replication of an SPC chart with the QDAR variance estimator $\widehat{\Omega}^2_{\mathrm{Q}}$, in Phase I we used the final batch size m_1 delivered by the QDAR batch-size determination algorithm in Section 3.2.2 to ensure that with approximate probability $1 - \alpha_{\mathrm{cor}} = 0.99$, Equation (16) was satisfied; and then in Phase II we took $m_2 = m_1$.

To provide a baseline for evaluating the performance of the selected SPC charts that incorporate a variance-estimation procedure, for each test process we also present the performance in Phase II of the DFTC chart using the exact values of σ^2 and Ω^2. On each replication of DFTC applied to each test process, we set the Phase II batch size m_2 according to the procedure specified in Section 3.2 of Kim et al. [26] so as to ensure that with approximate probability 0.99, the lag-one correlation between the batch means used in DFTC was at most 0.5. Corresponding to each table of estimated ARLs that is presented in this section is a matching table of standard errors for those ARLs that is presented in Appendix B.

3.4.1 AR(1) processes

For an AR(1) process defined by Equation (4), the marginal variance and the variance parameter are given by

$$\sigma^2 = \frac{\sigma_\varepsilon^2}{1-\varphi_Y^2} \quad \text{and} \quad \Omega^2 = \sigma^2\left(\frac{1+\varphi_Y}{1-\varphi_Y}\right) = \frac{\sigma_\varepsilon^2}{\left(1-\varphi_Y\right)^2}, \tag{24}$$

respectively. Table 3 summarizes the performance of the selected SPC charts in the AR(1) processes defined by varying the autoregressive parameter φ_Y over the values $0.25, 0.5, 0.7$, and 0.9.

First we consider the impact of variance estimation on the performance of the R&W, J&B, and DFTC-VE charts. As Table 3 shows, the R&W chart with variance estimation delivered values of ARL_0 that were reasonably close to the target value of 10,000 for all values of the autoregressive coefficient φ_Y. Moreover, for large shifts ($\eta > 1.5$) and for all values of φ_Y under study, the R&W chart outperformed the other charts (including DFTC) with respect to the delivered values of ARL_1. Thus we concluded that in the AR(1) process, variance estimation did not significantly degrade the effectiveness of the R&W chart in detecting large shifts in the process mean.

Table 3 also shows that in some situations, the performance of the J&B chart was degraded by variance estimation. In particular, when the J&B chart used the STS variance estimator $\widehat{\Omega}_{\mathrm{A}}^2$, the delivered values of ARL_0 were generally 10% to 12% below the target

Table 3. Two-sided ARLs in terms of number of raw observations for an AR(1) process

φ_Y	Shift	R&W	J&B		DFTC-VE		DFTC
	η		Area	QDAR	Area	QDAR	Phase II
		$\overline{m}_1 = 4$	$\overline{m}_1 = 94$	$\overline{m}_1 = 1$	$\overline{m}_1 = 93$	$\overline{m}_1 = 1$	
		$\overline{m}_2 = 4$	$\overline{m}_2 = 1$	$\overline{m}_2 = 1$	$\overline{m}_2 = 1$	$\overline{m}_2 = 1$	$\overline{m}_2 = 1$
0.25	0	10226	8796	9335	10758	10821	10846
	0.25	4294	724	728	275	274	270
	0.5	1181	365	366	110	111	111
	0.75	360	244	244	69	69	69
	1	133	183	184	50	50	50
	1.5	28	122	122	32	33	32
	2	10	92	92	24	24	24
	2.5	5	74	74	19	19	19
	3	4	61	61	16	16	16
	4	4	46	46	12	12	12
		$\overline{m}_1 = 8$	$\overline{m}_1 = 101$	$\overline{m}_1 = 2$	$\overline{m}_1 = 104$	$\overline{m}_1 = 2$	
		$\overline{m}_2 = 8$	$\overline{m}_2 = 1$	$\overline{m}_2 = 2$	$\overline{m}_2 = 1$	$\overline{m}_2 = 2$	$\overline{m}_2 = 1$
0.5	0	10200	8789	10258	10597	13129	11356
	0.25	4378	970	1017	443	517	434
	0.5	1166	488	513	178	229	180
	0.75	374	326	344	111	147	112
	1	141	245	258	80	108	82
	1.5	33	163	172	52	71	53
	2	14	122	129	38	53	39
	2.5	9	98	104	30	42	31
	3	8	82	87	25	35	26
	4	8	61	65	19	27	19
		$\overline{m}_1 = 19$	$\overline{m}_1 = 188$	$\overline{m}_1 = 8$	$\overline{m}_1 = 184$	$\overline{m}_1 = 7$	
		$\overline{m}_2 = 19$	$\overline{m}_2 = 1$	$\overline{m}_2 = 8$	$\overline{m}_2 = 1$	$\overline{m}_2 = 7$	$\overline{m}_2 = 3$
0.7	0	10454	8741	10157	10267	10826	11376
	0.25	4191	1307	1393	714	925	729
	0.5	1159	662	705	289	445	310
	0.75	381	446	473	182	291	198
	1	151	332	354	132	217	144
	1.5	43	222	238	85	143	94
	2	24	166	178	63	108	69
	2.5	20	133	143	49	87	55
	3	19	111	120	41	72	46
	4	19	83	91	30	55	34
		$\overline{m}_1 = 58$	$\overline{m}_1 = 487$	$\overline{m}_1 = 30$	$\overline{m}_1 = 485$	$\overline{m}_1 = 30$	
		$\overline{m}_2 = 58$	$\overline{m}_2 = 1$	$\overline{m}_2 = 30$	$\overline{m}_2 = 1$	$\overline{m}_2 = 30$	$\overline{m}_2 = 7$
0.9	0	11051	9005	10846	11567	9910	11668
	0.25	5359	2440	2546	1838	2182	1728
	0.5	1816	1228	1296	747	1076	755
	0.75	654	817	866	466	714	481
	1	307	612	658	339	537	352
	1.5	110	408	439	218	359	227
	2	69	306	334	158	271	167
	2.5	59	244	270	125	219	133
	3	58	203	227	104	185	111
	4	58	152	174	77	142	83

value, corresponding to rates of occurrence of false alarms that were 11% to 14% higher than expected. When the J&B chart used the variance estimator $\widehat{\Omega}^2_{\mathrm{Q}}$, the delivered values of ARL_0 deviated from the target value by percentages ranging from -6.7% to 8.5%. By contrast, from Table 3 of Kim et al. [26], we see that when the J&B chart uses the exact value of Ω^2 in the AR(1) process, the delivered values of ARL_0 are only 1% to 4% above the target value. From a similar analysis of the results for the out-of-control conditions in Table 3 of the current chapter and in Table 3 of Kim et al. [26], we concluded that use of either variance estimator $\widehat{\Omega}^2_{\mathrm{A}}$ or $\widehat{\Omega}^2_{\mathrm{Q}}$ did not significantly degrade the performance of the J&B chart with respect to the delivered values of ARL_0.

From the results in Table 3 we concluded that in the AR(1) process, the performance of the DFTC-VE chart was not significantly degraded in comparison with that of the DFTC chart using the exact values of σ^2 and Ω^2. In particular, the DFTC-VE chart using the STS variance estimator $\widehat{\Omega}^2_{\mathrm{A}}$ delivered nearly the same performance in the AR(1) process as that of the DFTC chart for both the in-control and out-of-control conditions and for all values of φ_Y. Although the performance of the DFTC-VE chart using the QDAR variance estimator $\widehat{\Omega}^2_{\mathrm{Q}}$ was somewhat worse than that of the DFTC chart when φ_Y took the values 0.5, 0.7, and 0.9, we judged the overall performance of DFTC-VE using $\widehat{\Omega}^2_{\mathrm{Q}}$ to be acceptable for the following reasons: (i) its rate of generation of false alarms was never significantly higher than expected; and (ii) its rate of generation of true alarms in each out-of-control condition was not significantly lower than that achieved by DFTC or by DFTC-VE using the alternative variance estimator $\widehat{\Omega}^2_{\mathrm{A}}$.

Beyond evaluating the impact of variance estimation on the performance of the R&W, J&B, and DFTC-VE charts in the AR(1) process, we evaluate the performance of these charts in comparison with each other. For small shifts ($\eta < 1$) and small-to-moderate correlations ($0.25 \leq \varphi_Y \leq 0.7$), the DFTC-VE chart with either variance estimator $\widehat{\Omega}^2_{\mathrm{A}}$ or $\widehat{\Omega}^2_{\mathrm{Q}}$ outperformed R&W with respect to the delivered values of ARL_1. For medium-to-large shifts ($\eta \geq 1.5$), the R&W chart outperformed the DFTC-VE chart. These conclusions agree with the findings of Kim et al. [26] about the relative performance of the R&W chart and the DFTC chart when the exact values of σ^2 and Ω^2 are known. Finally we observe that in the AR(1) process, DFTC-VE outperformed J&B with variance estimation in all

cases.

3.4.2 EAR(1) processes

For an EAR(1) process defined by Equation (5), the marginal variance and the variance parameter are given by

$$\sigma^2 = \sigma_\varepsilon^2 = \mu^2 \text{ and } \Omega^2 = \sigma^2 \left(\frac{1+\varphi_Y}{1-\varphi_Y} \right), \tag{25}$$

respectively. Because of the apparent similarities between the formulas for Ω^2 in Equations (24) and (25), we might expect that the distribution-free SPC charts using variance estimation would exhibit similar performance characteristics in the AR(1) and EAR(1) processes. Table 4 summarizes the experimental results obtained for all four distribution-free SPC charts when they were applied to the EAR(1) process.

Although the performance of the CUSUM-based charts (that is, J&B, DFTC-VE, and DFTC) was nearly the same in the EAR(1) and AR(1) processes, the R&W chart completely failed in the EAR(1) process, delivering values of ARL_0 whose percentage deviations from the target value of 10,000 ranged from -92% to -53%. This failure of the R&W chart is primarily due to the nonnormality of the EAR(1) process. The R&W chart is based on the assumption that the batch means are normal, as clearly seen in Equations (22) and (23). Unfortunately in the EAR(1) process, a batch size just large enough to satisfy Equation (21) is not large enough to ensure a sufficient degree of convergence to normality in the batch means so that Equations (22) and (23) can be applied.

From the results in Table 4 we concluded that in the EAR(1) process, variance estimation did not significantly degrade the performance of the DFTC-VE chart in comparison with that of the DFTC chart using the exact values of σ^2 and Ω^2. We found that in the EAR(1) process, the DFTC-VE chart using the STS variance estimator $\widehat{\Omega}_{\text{A}}^2$ delivered nearly the same performance as that of the DFTC chart for both the in-control and out-of-control conditions and for all values of φ_Y. We also found that in the EAR(1) process, the performance of the DFTC-VE chart using the QDAR variance estimator $\widehat{\Omega}_{\text{Q}}^2$ was acceptable but

Table 4. Two-sided ARLs in terms of number of raw observations for an EAR(1) process

φ_Y	Shift	R&W	J&B		DFTC-VE		DFTC
	η		Area	QDAR	Area	QDAR	Phase II
		$\overline{m}_1 = 4$	$\overline{m}_1 = 92$	$\overline{m}_1 = 1$	$\overline{m}_1 = 90$	$\overline{m}_1 = 1$	
		$\overline{m}_2 = 4$	$\overline{m}_2 = 1$	$\overline{m}_2 = 1$	$\overline{m}_2 = 1$	$\overline{m}_2 = 1$	$\overline{m}_2 = 1$
0.25	0	823	8820	8864	10486	10201	10557
	0.25	430	726	729	283	281	279
	0.5	247	367	366	112	112	112
	0.75	139	244	245	70	70	70
	1	80	183	184	51	50	51
	1.5	29	122	122	32	32	32
	2	11	92	92	24	24	24
	2.5	5	73	74	19	19	19
	3	4	61	61	16	16	16
	4	4	46	46	12	12	12
		$\overline{m}_1 = 8$	$\overline{m}_1 = 98$	$\overline{m}_1 = 2$	$\overline{m}_1 = 98$	$\overline{m}_1 = 2$	
		$\overline{m}_2 = 8$	$\overline{m}_2 = 1$	$\overline{m}_2 = 2$	$\overline{m}_2 = 1$	$\overline{m}_2 = 2$	$\overline{m}_2 = 1$
0.5	0	1449	8918	9816	10480	13135	11497
	0.25	737	968	1032	458	530	451
	0.5	385	489	520	179	233	183
	0.75	204	326	345	112	148	113
	1	104	244	258	80	108	81
	1.5	36	163	172	51	71	52
	2	14	122	129	38	53	38
	2.5	9	98	103	30	42	30
	3	8	81	86	25	35	25
	4	8	61	65	19	27	19
		$\overline{m}_1 = 19$	$\overline{m}_1 = 201$	$\overline{m}_1 = 8$	$\overline{m}_1 = 201$	$\overline{m}_1 = 7$	
		$\overline{m}_2 = 19$	$\overline{m}_2 = 1$	$\overline{m}_2 = 8$	$\overline{m}_2 = 1$	$\overline{m}_2 = 7$	$\overline{m}_2 = 3$
0.7	0	2826	8941	9535	10973	10621	11392
	0.25	1214	1339	1406	748	933	734
	0.5	564	668	705	301	443	316
	0.75	284	444	470	184	289	196
	1	141	331	353	132	215	142
	1.5	45	221	235	84	142	91
	2	22	166	177	61	107	68
	2.5	19	132	143	49	86	54
	3	19	110	120	40	72	45
	4	19	83	91	30	55	34
		$\overline{m}_1 = 58$	$\overline{m}_1 = 496$	$\overline{m}_1 = 29$	$\overline{m}_1 = 496$	$\overline{m}_1 = 30$	
		$\overline{m}_2 = 58$	$\overline{m}_2 = 1$	$\overline{m}_2 = 29$	$\overline{m}_2 = 1$	$\overline{m}_2 = 30$	$\overline{m}_2 = 7$
0.9	0	4685	9121	10202	11897	10195	12153
	0.25	2267	2475	2650	1941	2207	1770
	0.5	1094	1246	1311	761	1074	767
	0.75	560	822	865	468	713	479
	1	312	610	653	331	529	345
	1.5	109	404	437	212	352	221
	2	62	304	331	155	268	163
	2.5	58	241	268	122	217	130
	3	58	201	225	101	182	107
	4	58	151	172	75	140	81

somewhat worse than that of the DFTC chart when φ_Y took the values 0.5, 0.7, and 0.9. Finally comparing the performance of the J&B and DFTC-VE charts in the EAR(1) process, we concluded that DFTC-VE was more effective in detecting all shifts of the process mean at all levels of the autoregressive parameter φ_Y.

3.4.3 $M/M/1$ queue waiting times

The final test process consists of successive waiting times in the queue for an $M/M/1$ queueing system with arrival rate λ and service rate ν, where $\lambda < \nu$. Let $\{A_i : i = 1,2,\ldots\}$ denote a sequence of i.i.d. exponential random variables having mean $1/\lambda$. We take A_1 to be the arrival time of the first customer; and for $i = 2,3,\ldots$, we take A_i to be the time between the arrivals of the $(i-1)$st customer and the ith customer. Let $\{B_i : i = 1,2,\ldots\}$ denote a sequence of i.i.d. exponential random variables having mean $1/\nu$. For $i = 1,2,\ldots$, we take B_i to be the service time of the ith customer. Let $\tau = \lambda/\nu$ denote the steady-state server utilization. For $i = 1,2,\ldots$, let Y_i denote the waiting time in the queue for the ith customer. To ensure that $\{Y_i : i = 1,2,\ldots\}$ starts in steady state, we take $Y_1 = 0$ with probability $1-\tau$, and we sample Y_1 from an exponential distribution having mean $1/(\nu-\lambda)$ with probability τ. Subsequent waiting times in the queue are computed from the relation

$$Y_i = \max\{0, Y_{i-1} + B_{i-1} - A_i\} \quad \text{for} \quad i = 2,3,\ldots. \tag{26}$$

The steady-state marginal mean and variance of the queue-waiting time process defined by Equation (26) are given by

$$\mu = \frac{\tau^2}{\lambda(1-\tau)} \quad \text{and} \quad \sigma^2 = \frac{\tau^3(2-\tau)}{\lambda^2(1-\tau)^2},$$

respectively, and the variance parameter is given by

$$\Omega^2 = \frac{\tau^3\left(\tau^3 - 4\tau^2 + 5\tau + 2\right)}{\lambda^2(1-\tau)^4};$$

see, for example, p. 287 of Steiger and Wilson [53].

In our simulation experiments with this test process, we set the service rate to ν to 1.0. We set the arrival rate λ to 0.3 and 0.6, resulting in the server utilizations of 0.3 and 0.6, respectively. Note that for each out-of-control condition corresponding to a shift in the mean of size $\eta\sigma$, we simply added $\eta\sigma$ to each individual waiting time after it was generated according to Equation (26). Table 5 summarizes the experimental results obtained for all four distribution-free SPC charts when they were applied to the $M/M/1$ queue-waiting-time process.

Table 5. Two-sided ARLs in terms of number of raw observations for an $M/M/1$ process

τ	Shift	R&W	J&B		DFTC-VE		DFTC
	η		Area	QDAR	Area	QDAR	Phase II
		$\overline{m}_1 = 11$	$\overline{m}_1 = 500$	$\overline{m}_1 = 4$	$\overline{m}_1 = 499$	$\overline{m}_1 = 4$	
		$\overline{m}_2 = 11$	$\overline{m}_2 = 1$	$\overline{m}_2 = 4$	$\overline{m}_2 = 1$	$\overline{m}_2 = 4$	$\overline{m}_2 = 2$
0.3	0	408	9197	9972	13653	13547	9236
	0.25	279	1113	1169	603	689	596
	0.5	192	554	581	228	303	238
	0.75	122	366	387	137	193	146
	1	76	273	290	98	143	105
	1.5	25	182	193	63	94	68
	2	11	136	145	46	70	50
	2.5	11	109	117	37	56	40
	3	11	91	97	30	47	33
	4	11	68	74	23	36	25
		$\overline{m}_1 = 55$	$\overline{m}_1 = 499$	$\overline{m}_1 = 22$	$\overline{m}_1 = 499$	$\overline{m}_1 = 24$	
		$\overline{m}_2 = 55$	$\overline{m}_2 = 1$	$\overline{m}_2 = 22$	$\overline{m}_2 = 1$	$\overline{m}_2 = 24$	$\overline{m}_2 = 10$
0.6	0	880	9371	10824	16100	11149	13504
	0.25	745	2354	2543	1948	2119	1830
	0.5	581	1157	1243	709	994	746
	0.75	464	765	828	429	646	463
	1	372	571	621	304	482	337
	1.5	224	378	415	193	322	217
	2	136	282	315	142	243	161
	2.5	86	226	253	112	196	128
	3	66	188	213	92	165	107
	4	56	141	162	69	126	81

As in the case of the EAR(1) test process, the R&W chart failed in the $M/M/1$ queue-waiting-time process, delivering values of ARL_0 whose percentage deviations from the target value of 10,000 were -96% and -91% for server utilizations of 0.3 and 0.6, respectively. We concluded that the reasons for this failure were the same as for the EAR(1) process—nonnormality of the batch means when the R&W chart used batch sizes that were just large enough to satisfy Equation (21) but not nearly large enough to justify the use of

Equations (22) and (23) in calibrating the chart. We concluded that the R&W procedure cannot be used reliably to monitor autocorrelated processes exhibiting significant departures from normality.

From Table 5 we concluded that in the $M/M/1$ queue-waiting-time process, the performance of the DFTC-VE chart was similar to that of the DFTC chart using the exact values of σ^2 and Ω^2. In particular, we found that the DFTC-VE chart using the STS variance estimator $\widehat{\Omega}^2_{\mathrm{A}}$ delivered nearly the same performance as that of the DFTC chart for all out-of-control conditions and for both values of τ. For the in-control condition, however, the DFTC-VE chart using the STS variance estimator $\widehat{\Omega}^2_{\mathrm{A}}$ significantly overestimated ARL_0. We also found that the performance of the DFTC-VE chart using the QDAR variance estimator $\widehat{\Omega}^2_{\mathrm{Q}}$ was acceptable but somewhat worse than that of the DFTC-VE chart when $\tau = 0.6$. Finally comparing the performance of the J&B and DFTC-VE charts in the $M/M/1$ queue-waiting-time process, we concluded that DFTC-VE was more effective in detecting all shifts of the process mean at both levels of the server utilization.

3.5. Conclusion

In this chapter, we developed and evaluated DFTC-VE, a distribution-free tabular CUSUM chart for monitoring shifts in the mean of an autocorrelated process, which incorporates automated variance estimation from a training data set. We examined how the variance estimation affected the performance of DFTC-VE in comparison with that of its predecessor DFTC, which requires the exact values of the relevant process parameters. Moreover, we compared the performance of the DFTC-VE chart with that of other distribution-free SPC charts. When used in conjunction with all the distribution-free charts considered in this chapter, the proposed variance-estimation procedures did not cause dramatic performance degradation in comparison with the corresponding chart performance using the exact values of the relevant process parameters. Also, among the distribution-free charts considered, we observed that the DFTC-VE chart incorporating the STS variance estimator in Equation (10) performed very well; however, the improved performance came at the expense of additional algorithmic complexity. The QDAR variance estimator in Equation (18)—which is much easier to understand, implement, and apply—also performed reasonably well when

it was used in the DFTC-VE chart.

CHAPTER 4

MONITORING HIGH-DIMENSIONAL PROFILES USING A WAVELET-BASED DISTRIBUTION-FREE TABULAR CUSUM CHART

In this chapter, we develop and evaluate procedure WDFTC, which monitors high-dimensional profile having any nonsingular probability distribution in order to detect various types of shifts. It is almost always the case that the existing SPC charts used for profile monitoring assume the observations are multivariate normal [40]. Several researchers have shown that the wavelet-based SPC charts work well for specific processes, but they are not designed for nonnormal profiles (see, for example, Aradhye et al. [3] and Jeong et al. [18]). To overcome such a limitation of existing multivariate SPC charts, Qiu [40] proposed a distribution-free multivariate CUSUM-type SPC chart based on log-linear modeling; but as the author noted in the paper, the performance of the proposed method is questionable, as the log-linear modeling becomes challenging when the dimension of the observations increases, a common characteristic of profile data. Combined with the challenge of monitoring high-dimensional data, the normal assumption poses a natural challenge when developing a profile chart. We evaluate our proposed method by applying it to both normal and nonnormal data.

4.1. Background

To facilitate our discussion of the development of a distribution-free chart for monitoring high-dimensional profiles, we consider a vector-valued stochastic process of the form

$$\boldsymbol{Y}_j = \boldsymbol{f}(\boldsymbol{x}) + \boldsymbol{\varepsilon}_j, \;\; j = 1, 2, \ldots, \tag{27}$$

where: $\boldsymbol{x} = (x_1, \ldots, x_n)^{\mathrm{T}}$ is the $n \times 1$ vector consisting of n selected values of the input variable to be used in generating the jth observed profile (note that $\boldsymbol{x}$ is the same for all

profiles; and throughout this chapter, we let $\boldsymbol{A}^{\mathrm{T}}$ denote the transpose of a vector or matrix $\boldsymbol{A}$); $\boldsymbol{Y}_j = (y_{1,j}, \ldots, y_{n,j})^{\mathrm{T}}$ is the $n \times 1$ vector consisting of the n respective values of the response variable; $\boldsymbol{f}(\boldsymbol{x}) = [\, f(x_1), \ldots, f(x_n)\,]^{\mathrm{T}}$ is the $n \times 1$ vector consisting of the n respective expected values of the response variable; and $\boldsymbol{\varepsilon}_j = (\varepsilon_{1,j}, \ldots, \varepsilon_{n,j})^{\mathrm{T}}$ is the associated $n \times 1$ noise (error) vector with mean $\mathrm{E}[\boldsymbol{\varepsilon}_j] = \boldsymbol{0}_n$ (the $n \times 1$ vector of zeros) and covariance matrix $\mathrm{Cov}[\boldsymbol{\varepsilon}_j] = \mathrm{E}[\boldsymbol{\varepsilon}_j \boldsymbol{\varepsilon}_j^{\mathrm{T}}] = \boldsymbol{\Sigma}_0$. The relevant univariate functional relationship holds for each point of the jth profile; thus we have $y_{i,j} = f(x_i) + \varepsilon_{i,j}$ for $i = 1, \ldots, n$, where $\varepsilon_{i_1,j}$ and $\varepsilon_{i_2,j}$ may be nonnormal and correlated for $i_1 \neq i_2$. We distinguish two process states: (a) $\boldsymbol{Y}_j$ is in-control when $\mathrm{E}[\boldsymbol{Y}_j] = \boldsymbol{f}_0 = [\, f_0(x_1), \ldots, f_0(x_n)\,]^{\mathrm{T}}$ for a given in-control function $f_0(\cdot)$ relating the input variable to the corresponding mean response; and (b) $\boldsymbol{Y}_j$ is out-of-control when $\mathrm{E}[\boldsymbol{Y}_j] = \boldsymbol{f}_1 = [\, f_1(x_1), \ldots, f_1(x_n)\,]^{\mathrm{T}} \neq \boldsymbol{f}_0$ for any other function $f_1(\cdot)$. Without loss of generality, throughout the rest of this chapter we assume the mean in-control profile $\boldsymbol{f}_0$ is centered so that $\boldsymbol{1}_n^{\mathrm{T}} \boldsymbol{f}_0 = \sum_{i=1}^{n} f_0(x_i) = 0$, where $\boldsymbol{1}_n$ is the $n \times 1$ vector of ones.

Whether it is in-control or out-of-control, the jth observed profile $\boldsymbol{Y}_j$ (for $j = 1, 2, \ldots$) is assumed to have the same covariance matrix $\mathrm{Cov}[\boldsymbol{Y}_j] = \boldsymbol{\Sigma}_0$. For the ith component $Y_{i,j}$ of the jth profile ($i = 1, \ldots, n$), we let $\sigma_i^2 = \left[\boldsymbol{\Sigma}_0\right]_{i,i}$ denote the component's marginal variance. Suppose that the profile length n has the form $n = 2^J$ for some positive integer J and that the DWT defined by a given wavelet system with the coarsest level of resolution $L \in \{0, \ldots, J-1\}$ is applied to both $\boldsymbol{f}_0$ and $\boldsymbol{Y}_j$ for $j = 1, 2, \ldots$ [35, 39]. Then we have 2^L scaling coefficients, representing the coarser features of the profile—i.e., the profile features that are prominent at the lower levels of resolution. Let $\boldsymbol{W}$ denote the corresponding DWT matrix and let $\boldsymbol{d}_j = \boldsymbol{W}\boldsymbol{Y}_j = (d_{1,j}, \ldots, d_{n,j})^{\mathrm{T}}$ denote the DWT of the jth profile, while $\boldsymbol{\theta}_0 = \boldsymbol{W}\boldsymbol{f}_0 = (\theta_{1,0}, \ldots, \theta_{n,0})^{\mathrm{T}}$ denotes the DWT of the mean in-control profile $\boldsymbol{f}_0$. The covariance matrix of $\boldsymbol{d}_j$ is given by $\mathrm{Cov}[\boldsymbol{d}_j] = \boldsymbol{\Lambda}_0 = \boldsymbol{W}\boldsymbol{\Sigma}_0\boldsymbol{W}^{\mathrm{T}}$.

Throughout this chapter, we compare and analyze different profile-monitoring charts based on the in-control average run length (ARL_0) and the out-of-control average run length (ARL_1) expressed in terms of the number of individual profiles $\{\boldsymbol{Y}_j : j = 1, 2, \ldots\}$ that are observed before raising a false alarm (under the in-control condition) or a true alarm (under a specific out-of-control condition).

4.1.1 Motivating Example

In this subsection, we demonstrate the need for a distribution-free SPC chart that effectively monitors high-dimensional profiles exhibiting variance heterogeneity, nonnormality, or stochastic dependencies among profile components. In particular, we examine the performance of the wavelet-based profile-monitoring chart M^* of Chicken et al. [9], which is designed to monitor nonlinear profiles described by Equation (27), where the jth error vector $\boldsymbol{\varepsilon}_j = (\varepsilon_{1,j}, \ldots, \varepsilon_{n,j})^{\mathrm{T}}$ is assumed to consist of i.i.d. $N(0, \sigma_*^2)$ components; that is, the noise terms $\{\varepsilon_{i,j} : i = 1, \ldots, n\}$ are assumed to be independent normal random variables with mean 0 and standard deviation σ_*. Therefore each observed profile has the covariance matrix $\boldsymbol{\Sigma}_0 = \sigma_*^2 \boldsymbol{I}_n$, where $\boldsymbol{I}_n$ denotes the $n \times n$ identity matrix.

We consider two motivating examples in which the above assumptions on the error vectors $\{\boldsymbol{\varepsilon}_j\}$ are violated. In the first motivating example (ME_1), we explore the effect on the performance of the chart M^* arising from correlated normal noise components with heterogeneous variances as detailed below.

- The autocorrelation function for the noise components of each profile in ME_1 is taken from Sachs and MacGibbon ([45], p. 484)—namely, the damped sinusoidal form

$$\begin{aligned} \rho(\ell) = \mathrm{Corr}[y_{i,j}, y_{i+\ell,j}] &= \mathrm{Corr}[\varepsilon_{i,j}, \varepsilon_{i+\ell,j}] \\ &= (-\alpha_2)^{|\ell/2|} \left[\frac{\sin(|\ell|\omega + \xi)}{\sin(\xi)} \right] \quad \text{for } \ell = 0, \pm 1, \ldots, \pm(n-1), \end{aligned} \tag{28}$$

  where we take $\alpha_1 = 4/3$, $\alpha_2 = -8/9$, the angular frequency $\omega = \cos^{-1}\left[\alpha_1 / 2\sqrt{-\alpha_2}\right]$ $\cong 0.785$, and the phase constant $\xi = \tan^{-1}[\tan(\omega)(1-\alpha_2)/(1+\alpha_2)] \cong 1.51$. This gives, for example, $\rho(1) = 0.71$ and $\rho(2) = 0.052$.

- The marginal variances for the components of each profile in ME_1 are similar to

those used in Example 2 of Gao ([15]),

$$\sigma_i^2 = \mathrm{Var}[y_{i,j}] = \mathrm{Var}[\varepsilon_{i,j}] \\ = \sigma_0^2 \left(1 + \left\{0.5 - 2.5\left[(i-1)/n - 0.515\right]^2\right\}^2\right)^2 \text{ for } i = 1,\ldots,n, \tag{29}$$

where $\sigma_0^2 = 9.50$. The resulting marginal variances σ_i^2 (for $i = 1,\ldots,n$) take values between 9.5 and 14.8; and the componentwise correlations take values between -0.71 and 0.71.

The covariances between pairs of profile components in ME_1 are then given by $\mathrm{Cov}[y_{i_1,j}, y_{i_2,j}]$ $= \left[\boldsymbol{\Sigma}_0\right]_{i_1,i_2} = \sigma_{i_1}\sigma_{i_2}\rho(i_1 - i_2)$ for $i_1, i_2 = 1,\ldots,n$.

In the second motivating example (ME_2), we explore the effect on the performance of the chart M^* arising from nonnormal marginal distributions for the components of the profiles that are randomly sampled during Phase II operation. By contrast with ME_1, test process ME_2 has noise components that are mutually independent shifted exponential random variables with mean 0 and variance 1 for $j = 1, 2, \ldots$.

In both test processes ME_1 and ME_2, we add out-of-control shifts and noise terms to the mean in-control profile $\boldsymbol{f}_0$ defined by $n = 512$ equally spaced points on the piecewise smooth function of Mallat ([35], p. 458) as depicted in Figure 1; and we monitor the observed profiles $\{\boldsymbol{Y}_j : j = 1, 2, \ldots\}$ in Phase II operation using procedure M^*.

When monitoring nonnormal profiles, we consider two different simulation-based methods to calibrate (estimate) the control limits for an SPC chart that was originally developed under the assumption of normally distributed profile components, possibly with nonzero componentwise correlations.

Calibration Method CM_A: Generate a preliminary (Phase I) data set consisting of normally distributed profiles that have the same in-control mean vector and the same covariance matrix as the nonnormal profiles to be monitored. Obtain the required control limit(s) for the normally distributed profiles via trial-and-error simulations designed to yield the prespecified target value of ARL_0. Use the resulting control

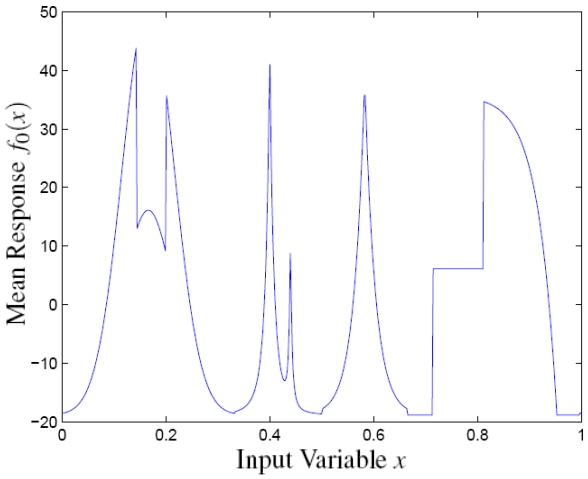

Figure 1. Mallat's piecewise smooth function

limit(s) to monitor the nonnormal profiles in regular Phase II operation.

Calibration Method $\mathrm{CM_B}$: Obtain the required control limit(s) via trial-and-error simulations using the same type of in-control, nonnormal profiles that are to be monitored in Phase II; then use the resulting control limit(s) to detect out-of-control conditions in Phase II operation.

Exploiting the $\mathrm{CM_A}$-based control limit(s), we can illustrate the risk of monitoring nonnormal profiles with existing SPC charts that were originally designed for normal profiles. A similar approach is taken by Qiu [40], wherein he demonstrates how excessively large rates of occurrence for false alarms (or equivalently, values for ARL_0 that are substantially below the user-specified nominal level) can occur when SPC charts based on the normal assumption are applied to nonnormal profiles. On the other hand, $\mathrm{CM_B}$ enables us to compare the performance of different SPC charts in terms of the resulting values of ARL_1 (or equivalently, the rates of occurrence of true alarms) for a specific out-of-control condition, because each chart's control limits have been calibrated to yield the target value of ARL_0 when the monitored nonnormal profiles are in-control.

The profile-monitoring chart M^* was applied as follows:

Chart M^* : Calculate $\boldsymbol{\theta}_0 = \boldsymbol{W}\boldsymbol{f}_0$. Let j_0 denote the unknown change point (profile index)

after which out-of-control profiles occur in Phase II operation, where $j_0 \geq 0$. Calculate $\boldsymbol{d}_j = \boldsymbol{W}\boldsymbol{Y}_j$ for $j = 1, \ldots, G$, where G is assumed to be large enough so that $G \geq j_0$. For $j = 1, \ldots, G$, calculate the basic statistic $w_j = (n/\sigma_*^2) \sum_{i=1}^{n} (d_{i,j} - \theta_{i,0})^2$ measuring the standardized discrepancy between $\boldsymbol{d}_j$ and $\boldsymbol{\theta}_0$ as well as its "thresholded" version $\widetilde{w}_j = (n/\sigma_*^2) \sum_{i=1}^{n} [\mathrm{thr}(d_{i,j} - \theta_{i,0})]^2$, where the VisuShrink thresholding operator $\mathrm{thr}(\cdot)$ of Donoho and Johnstone [12] is applied to each component of the difference $\boldsymbol{d}_j - \boldsymbol{\theta}_0$. Given a candidate value $u \in \{0, 1, \ldots, G-1\}$ for the unknown change point j_0, calculate the associated likelihood-ratio parameter estimator $\widehat{\gamma} = (G-u)^{-1} \left(\sum_{j=u+1}^{G} \widetilde{w}_j \right) - u^{-1} \left(\sum_{j=1}^{u} \widetilde{w}_j \right)$ for $u \neq 0$ and $\widehat{\gamma} = G^{-1} \sum_{j=1}^{G} \widetilde{w}_j$ for $u = 0$; and finally evaluate the log-likelihood-ratio statistic $h(u) = (\widehat{\gamma}/2) \sum_{j=u+1}^{G} [(w_j/n) - 1]$ so as to find $\widehat{j_0} = \arg\max \left\{ h(u) : u = 0, 1, \ldots, G-1 \right\}$, the estimated change point. For an upper control limit (UCL) obtained via trial-and-error simulations designed to yield the target value $\mathrm{ARL}_0 = 200$, raise an out-of-control alarm at time (profile) index G if $h(\widehat{j_0}) > \mathrm{UCL}$.

Following the approach of Chicken et al. [9], we express the overall size of a shift $\boldsymbol{f}_1 - \boldsymbol{f}_0$ in the mean profile in terms of the squared Euclidean distance between $\boldsymbol{f}_0$ and $\boldsymbol{f}_1$,

$$a = \|\boldsymbol{f}_1 - \boldsymbol{f}_0\|_2^2 = \sum_{i=1}^{n} [\, f_1(x_i) - f_0(x_i) \,]^2 .$$

Recall that we set $n = 512$; and in Phase II operation of M^*, we take uniform local shifts added to $\boldsymbol{f}_0$ for the component indices $i \in \{89, 90, \ldots, 96\}$ (i.e., 8 shifted components) and $i \in \{241, 242, \ldots, 256\}$ (i.e., 16 shifted components) so as to yield a selected value of the overall shift size $\sqrt{a}$. This local shift was also used by Jeong et al. [18] and Chicken et al. [9].

Table 6 contains the estimated ARLs and the associated standard errors delivered by M^* based on 1,000 independent replications of the test process ME_1 when a uniform local shift of overall size $\sqrt{a}$ was added to the in-control mean profile $\boldsymbol{f}_0$, with the same values of $\sqrt{a}$ used by Chicken et al. [9]. To apply M^*, we estimated σ_* using the average of the median absolute deviations of the $n/2$ highest-level detail coefficients from each observed profile as proposed by Chicken et al. [9]. Note that in test problem ME_1, the calibration methods

CM_A and CM_B coincide. Comparing the values of ARL_1 in Table 6 with the corresponding values of ARL_1 in Table 1 of Chicken et al. [9], we concluded that the performance of M^* was unacceptable for all the specified out-of-control conditions.

Table 6. ARLs delivered by M^* for test process ME_1

Shift Type	$\sqrt{a}$	Est. ARL	Std. Err.
In-Control	0	199.	0.83
Local Shift	0.1	198.	0.71
	0.2	198.	0.71
	0.3	196.	0.71
	0.4	194.	0.69
	0.5	189.	0.68

Table 7 contains the estimated ARLs and the associated standard errors delivered by M^* based on 1,000 independent replications of the nonnormal test process ME_2 when a uniform local shift of overall size $\sqrt{a}$ was added to the in-control mean profile $\boldsymbol{f}_0$. The performance of M^* was evaluated using both calibration methods CM_A and CM_B. The control limit obtained from CM_A resulted in an extremely small value of ARL_0 for M^*, which translated into an unacceptably large rate of occurrence of false alarms; and in this anomalous situation, we omitted applying M^* to ME_2 with the specified out-of-control conditions. Comparing the values of ARL_1 in Table 7 with the corresponding values of ARL_1 in Table 1 of Chicken et al. [9], we concluded that when procedure M^* was calibrated using method CM_B, the performance of M^* was unacceptable for all the specified out-of-control conditions.

It was clear from the results for both test processes ME_1 and ME_2 that the performance of M^* became problematic in the presence of stochastic dependence, heterogeneous variances, and nonnormality of the sampled profiles. Such characteristics are common in high-dimensional profile data, but most existing profile-monitoring charts, including M^*, require the monitored profile to have i.i.d. normal noise components for successful application of the chart. This conclusion will be placed into a more complete perspective in the fourth section of this chapter, where we summarize the results of a comprehensive experimental performance evaluation of WDFTC versus M^* and some other commonly used

Table 7. ARLs delivered by M^* for test process ME_2

Shift Type	$\sqrt{a}$	Calibration Method			
		CM_A		CM_B	
		Est. ARL	Std. Err.	Est. ARL	Std. Err.
In-Control	0	3.32	0.09	200.	0.81
Local Shift	0.1	–	–	191.	0.98
	0.2	–	–	165.	0.86
	0.3	–	–	110.	0.62
	0.4	–	–	59.	0.37
	0.5	–	–	30.	0.20

profile-monitoring schemes.

4.1.2 Wavelet Transform Overview

In this subsection, we briefly review the wavelet transform. Let $\mathscr{L}^2[0,1]$ denote the space of real-valued square-integrable functions defined on the unit interval $[0,1]$. The wavelet transform of a function $g \in \mathscr{L}^2[0,1]$ is used to obtain a representation of g as an infinite series involving orthonormal basis functions. A scaling function $\phi \in \mathscr{L}^2[0,1]$ has several key properties that give rise to the associated wavelet function $\psi \in \mathscr{L}^2[0,1]$; and from ψ, we can derive an orthonormal set of basis functions for $\mathscr{L}^2[0,1]$ analogous to the trigonometric functions used in the Fourier series representation. For simplicity in the following discussion, we assume that ϕ and ψ are the Haar scaling and wavelet functions, respectively; see Ogden ([39], pp. 7–23) or Mallat ([35], p. 291).

For a function $g \in \mathscr{L}^2[0,1]$, the wavelet representation of g is given by

$$\begin{aligned} g(z) &= \lim_{B\to\infty} \sum_{\ell=-\infty}^{B-1} \sum_{m=0}^{\lceil 2^\ell \rceil - 1} \langle g, \psi_{\ell,m}\rangle \psi_{\ell,m}(z) \\ &= \lim_{B\to\infty} \sum_{m=0}^{2^B-1} \langle g, \phi_{B,m}\rangle \phi_{B,m}(z) \ \text{ for almost all } \ z \in [0,1], \end{aligned} \tag{30}$$

where: $h_{\ell,m}(z) = 2^{\ell/2} h(2^\ell z - m)$ for $h = \psi, \phi$; and for $g_1, g_2 \in \mathscr{L}^2[0,1]$ we let $\langle g_1, g_2\rangle = \int_0^1 g_1(z) g_2(z)\,\mathrm{d}z$ denote the inner product operator ([39]). The Bth partial sum $P_B(g)$ on

the far right-hand side of Equation (30) can be viewed as an approximation to g that becomes progressively more accurate as B increases. In Equation (30), the quantities $\{\mathsf{C}_{\ell,m} = \langle g, \phi_{\ell,m}\rangle\}$ are called the scaling coefficients of g, and the quantities $\{\mathsf{D}_{\ell,m} = \langle g, \psi_{\ell,m}\rangle\}$ are called the detail coefficients of g. In practice, a physical measuring device can only measure a signal (function) g to a finite level of resolution; thus we take $g \approx P_J(g)$ for some finest (highest) level of resolution J; furthermore, the successive function-approximation operations must stop at some coarsest (lowest) level of resolution L, where $L < J$. As a result, one obtains an approximate representation of g based on its DWT,

$$\begin{aligned} g(z) &\approx \sum_{m=0}^{2^J-1} \mathsf{C}_{J,m}\phi_{J,m}(z) \\ &= \sum_{m=0}^{2^L-1} \mathsf{C}_{L,m}\phi_{L,m}(z) + \sum_{\ell=L}^{J-1}\sum_{m=0}^{2^\ell-1} \mathsf{D}_{\ell,m}\psi_{\ell,m}(z) \ \text{ for almost all } \ z \in [0,1], \end{aligned} \tag{31}$$

where: (i) the scaling functions $\{\phi_{L,m}(z)\}$ represent the low-frequency components of $g(z)$—that is, the smooth parts of $g(z)$; and (ii) the wavelet functions $\{\psi_{\ell,m}(z)\}$ represent the high-frequency components of $g(z)$—that is, the local behavior of $g(z)$.

In the context of monitoring deviations from an in-control profile defined by the function $f_0(x)$ for $x_{\text{LO}} \le x \le x_{\text{HI}}$, we exploit the wavelet transform by taking $g(z) = f_0[x_{\text{LO}} + z(x_{\text{HI}} - x_{\text{LO}})]$ in Equations (30) and (31). Because the in-control function $f_0(x) = g[(x - x_{\text{LO}})/(x_{\text{HI}} - x_{\text{LO}})]$ as approximated in Equation (31) is originally represented using $n = 2^J$ scaling coefficients $\{\mathsf{C}_{J,m} : m = 0, 1, \ldots, 2^J - 1\}$ of $f_0(\cdot)$ at the finest level of resolution, we see that $f_0(x)$ can also be represented using the 2^L coarsest-level scaling coefficients $\{\mathsf{C}_{L,m} : m = 0, 1, \ldots, 2^L - 1\}$ of $f_0(\cdot)$ together with the $n - 2^L$ detail coefficients $\{\mathsf{D}_{\ell,m} : \ell = L, L+1, \ldots, J-1; m = 0, 1, \ldots, 2^\ell - 1\}$ of $f_0(\cdot)$. Therefore monitoring deviations from an $n \times 1$ in-control mean profile vector $\boldsymbol{f}_0$ defined by the function $f_0(\cdot)$ is equivalent to monitoring deviations from the $n \times 1$ vector of scaling and detail coefficients in the DWT of $f_0(\cdot)$. Let $\boldsymbol{W}$ denote the $n \times n$ orthogonal matrix associated with the DWT of $f_0(\cdot)$. Given a randomly sampled $n \times 1$ in-control profile $\boldsymbol{Y}_j$, the linear transformation $\boldsymbol{d}_j = \boldsymbol{W}\boldsymbol{Y}_j$ yields estimates of the scaling and detail coefficients of $f_0(\cdot)$, where if necessary the rows of $\boldsymbol{W}$

have been suitably interchanged to ensure that the first 2^L components of $\boldsymbol{d}_j$ are the estimated scaling coefficients of $f_0(\cdot)$, and the last $n-2^L$ components of $\boldsymbol{d}_j$ are the estimated detail coefficients of $f_0(\cdot)$.

Because of its simplicity, the Haar wavelet is frequently used in existing wavelet-based SPC schemes (Ganesan et al. [14], Jeong et al. [18]), especially when the in-control function $f_0(\cdot)$ is piecewise constant. For smoother functions, other wavelet systems such as the Daubechies or Symmlet wavelets are often used (Lada et al. [28]; Ganesan et al. [14]). In this chapter we use the Symmlet wavelet with the number of vanishing moments equal to eight because the Symmlet 8 wavelet yields a smoother approximation to $f_0(\cdot)$ than the Haar wavelet does.

4.2. Wavelet-Based Distribution-Free Tabular CUSUM Profile Monitoring Chart

In this section, we present steps of WDFTC for monitoring high-dimensional profiles with general variance-covariance structures. The idea is to reduce the dimensionality of the original sampled profile $\boldsymbol{Y}_i$ for better detection power and then monitor the dimension-reduced data by aggregating the selected components of the corresponding DWT in the form of the (univariate) Hotelling's T^2-type statistic. We first present overall steps of WDFTC and then discuss each step in detail. Specifically, we discuss how to reduce dimensionality, the impacts of componentwise covariances, the batching method to reduce the impacts of the covariances, and estimation of the parameters required to perform WDFTC in practical applications. Table 8 gives a list of all key notation needed in the formulation of WDFTC.

4.2.1 New High-Dimensional Profile Monitoring Chart

Procedure WDFTC combines the DWT with the distribution-free tabular CUSUM chart of Kim et al. [26] and Lee et al. [32] and focuses on monitoring key components of the DWT determined by a wavelet-based dimension-reduction technique that will be explained in the following subsection. Table 8 provides a list of all key notation needed in the formulation of WDFTC.

WDFTC begins by computing the wavelet coefficient vector $\boldsymbol{\theta}_0 = \boldsymbol{W}\boldsymbol{f}_0$ for the in-control

Table 8. Notation summary

$\boldsymbol{f}_0$ = the $n \times 1$ in-control mean profile, which is assumed to satisfy the centering condition $\mathbf{1}_n^{\mathrm{T}} \boldsymbol{f}_0 = 0$;

$\boldsymbol{\theta}_0 = \boldsymbol{W}\boldsymbol{f}_0$, the $n \times 1$ DWT of the in-control mean profile $\boldsymbol{f}_0$, where the first 2^L components of $\boldsymbol{\theta}_0$ are the scaling coefficients and the last $n - 2^L$ components of $\boldsymbol{\theta}_0$ are the detail coefficients;

$\boldsymbol{\theta}_0^{\#}$ = the $n \times 1$ version of $\boldsymbol{\theta}_0$ in which p elements are selected for retention and $n - p$ elements are set to zero and so as to minimize the weighted relative reconstruction error (33) incurred by using $\boldsymbol{W}^{-1}\boldsymbol{\theta}_0^{\#}$ as an approximation to $\boldsymbol{f}_0$;

$\boldsymbol{\vartheta}_0^{\#}$ = the $p \times 1$ version of $\boldsymbol{\theta}_0^{\#}$ in which the $n - p$ nonselected elements of $\boldsymbol{\theta}_0^{\#}$ have been deleted;

$\boldsymbol{f}_0^{\#} = \boldsymbol{W}^{-1}\boldsymbol{\theta}_0^{\#}$, the approximate in-control mean profile reconstructed from $\boldsymbol{\theta}_0^{\#}$;

$\boldsymbol{Y}_j$ = the jth $n \times 1$ observed profile for $j = 1, \ldots, N$ in Phase I and for $j = 1, 2, \ldots$ in Phase II;

$\boldsymbol{Y}_k(r) = r^{-1} \sum_{u=1}^{r} \boldsymbol{Y}_{(k-1)r+u}$, the kth $n \times 1$ batch-means vector based on nonoverlapping batches of size r for $k = 1, \ldots, \lfloor N/r \rfloor$ in Phase I and for $k = 1, 2, \ldots$ in Phase II;

$\boldsymbol{d}_j = \boldsymbol{W}\boldsymbol{Y}_j$, the $n \times 1$ DWT of the jth observed profile $\boldsymbol{Y}_j$;

$\boldsymbol{d}_k(r) = r^{-1} \sum_{u=1}^{r} \boldsymbol{d}_{(k-1)r+u}$, the kth $n \times 1$ batch-means DWT vector computed from nonoverlapping batches of size r;

$\boldsymbol{d}_j^{\#}$ = the $p \times 1$ reduced-dimension version of $\boldsymbol{d}_j$ in which the $n - p$ elements of $\boldsymbol{d}_j$ corresponding to the nonselected (zero-valued) elements of $\boldsymbol{\theta}_0^{\#}$ have been deleted to yield $\boldsymbol{d}_j^{\#}$;

$\boldsymbol{d}_k^{\#}(r) = r^{-1} \sum_{u=1}^{r} \boldsymbol{d}_{(k-1)r+u}^{\#}$, the kth $p \times 1$ batch-means vector of reduced-dimension DWTs based on nonoverlapping batches of size r;

$\boldsymbol{\Lambda}_0 = \mathrm{E}\big[\big(\boldsymbol{d}_j - \mathrm{E}[\boldsymbol{d}_j]\big) \big(\boldsymbol{d}_j - \mathrm{E}[\boldsymbol{d}_j]\big)^{\mathrm{T}} \big]$, the $n \times n$ covariance matrix of $\boldsymbol{d}_j$, assumed to be the same for both in-control and out-of-control conditions;

$\boldsymbol{\Lambda}_0(r) = \boldsymbol{\Lambda}_0 / r$, the $n \times n$ covariance matrix of $\boldsymbol{d}_k(r)$;

$\boldsymbol{\Lambda}_0^{\#} = \mathrm{E}\big[\big(\boldsymbol{d}_j^{\#} - \mathrm{E}[\boldsymbol{d}_j^{\#}]\big) \big(\boldsymbol{d}_j^{\#} - \mathrm{E}[\boldsymbol{d}_j^{\#}]\big)^{\mathrm{T}} \big]$, the $p \times p$ covariance matrix of the reduced-dimension DWT $\boldsymbol{d}_j^{\#}$;

$\boldsymbol{\Lambda}_0^{\#}(r) = \boldsymbol{\Lambda}_0^{\#} / r$, the $p \times p$ covariance matrix of the reduced-dimension batch-means DWT $\boldsymbol{d}_k^{\#}(r)$;

$\overline{\boldsymbol{d}}_N^{\#} = N^{-1} \sum_{j=1}^{N} \boldsymbol{d}_j^{\#}$, the $p \times 1$ sample mean of the the reduced-dimension DWTs $\{\boldsymbol{d}_j^{\#} : j = 1, \ldots, N\}$ computed from the profiles observed in Phase I;

$\widehat{\boldsymbol{\Lambda}}_0^{\#} = (N-1)^{-1} \sum_{j=1}^{N} \big(\boldsymbol{d}_j^{\#} - \overline{\boldsymbol{d}}_N^{\#}\big) \big(\boldsymbol{d}_j^{\#} - \overline{\boldsymbol{d}}_N^{\#}\big)^{\mathrm{T}}$, the $p \times p$ sample covariance matrix of the reduced-dimension DWTs $\{\boldsymbol{d}_j^{\#} : j = 1, \ldots N\}$ computed from the profiles observed in Phase I;

$\widetilde{\boldsymbol{\Lambda}}_0^{\#}$ = version of $\widehat{\boldsymbol{\Lambda}}_0^{\#}$ that has been regularized (thresholded) according to Algorithm CMR below;

$\widetilde{\boldsymbol{\Lambda}}_0^{\#}(r) = \widetilde{\boldsymbol{\Lambda}}_0^{\#} / r$, the $p \times p$ estimated covariance matrix of the reduced-dimension DWTs $\{\boldsymbol{d}_k^{\#}(r) : k = 1, \ldots, \lfloor N/r \rfloor\}$ based on the regularized sample covariance matrix $\widetilde{\boldsymbol{\Lambda}}_0^{\#}$.

mean profile $\boldsymbol{f}_0$, where the first 2^L components of $\boldsymbol{\theta}_0$ are the scaling coefficients, and the last $n - 2^L$ components of $\boldsymbol{\theta}_0$ are the detail coefficients. As described in the next subsection,

we seek an "optimal" set of p wavelet coefficients selected from the components of $\boldsymbol{\theta}_0$ to constitute the respective nonzero components of the $n \times 1$ vector $\boldsymbol{\theta}_0^{\#}$ so that the following conditions hold: (a) we take $2^L \leq p \leq n$, selecting all 2^L scaling coefficients and the $p - 2^L$ largest-magnitude detail coefficients of $\boldsymbol{\theta}_0$ to form the nonzero components of $\boldsymbol{\theta}_0^{\#}$; and (b) as an approximation to $\boldsymbol{f}_0$, the inverse transform $\boldsymbol{W}^{-1}\boldsymbol{\theta}_0^{\#}$ minimizes the weighted relative reconstruction error (WRRE) over all $p \in \{2^L, \ldots, n\}$. Let $\boldsymbol{\vartheta}_0^{\#}$ denote the $p \times 1$ reduced-dimension version of $\boldsymbol{\theta}^{\#}$ in which all the nonselected (zero-valued) components have been deleted; and let $\boldsymbol{d}_j^{\#}$ denote the corresponding $p \times 1$ reduced-dimension version of the DWT of the jth profile $\boldsymbol{Y}_j$ for $j = 1, 2, \ldots$. Let $\boldsymbol{\Lambda}_0^{\#}$ denote the $p \times p$ covariance matrix of $\boldsymbol{d}_j^{\#}$, and let $\widetilde{\boldsymbol{\Lambda}}_0^{\#}$ denote the regularized (thresholded) estimator of $\boldsymbol{\Lambda}_0^{\#}$ computed from the Phase I data. WDFTC computes the batch-means vectors $\boldsymbol{d}_k^{\#}(r) = r^{-1}\sum_{u=1}^{r} \boldsymbol{d}_{(k-1)r+u}^{\#}$ based on nonoverlapping batches of size r observed in Phase I for $k = 1, \ldots, \lfloor N/r \rfloor$. All the sample information about in-control deviations from $\boldsymbol{\theta}_0$ within the kth batch of r profiles observed in Phase I is combined in Hotelling's statistic $T_k^2(r) = \left[\boldsymbol{d}_k^{\#}(r) - \boldsymbol{\vartheta}_0^{\#}\right]^{\mathrm{T}} \left[\widetilde{\boldsymbol{\Lambda}}_0^{\#}/r\right]^{-1} \left[\boldsymbol{d}_k^{\#}(r) - \boldsymbol{\vartheta}_0^{\#}\right]$ for $k = 1, \ldots, \lfloor N/r \rfloor$. Procedure WDFTC determines its control limit analytically for a given target value of ARL_0 using an approach adapted from Kim et al. [26] based on the sample mean and variance of the statistics $\{T_k^2(r) : k = 1, \ldots, \lfloor N/r \rfloor\}$ observed in Phase I. Then in Phase II (regular) operation, the CUSUM procedure of Lee et al. [32] is applied to the associated statistics $\{T_k^2(r) : k = 1, 2, \ldots\}$ to detect out-of-control conditions. A formal algorithmic statement of WDFTC is given below.

4.2.2 Dimension Reduction

In this subsection, we discuss WDFTC's dimension-reduction technique. Jin and Shi [19] use a universal thresholding scheme for wavelet shrinkage, but such a scheme assumes uncorrelated normal components and thus does not always work for nonnormal components. Instead, we propose an extension of the method of Lada et al. [28] that exploits the concept of weighted relative reconstruction error. We seek to select a (relatively) small number p of the components of $\boldsymbol{\theta}_0 = (\theta_{1,0}, \ldots, \theta_{n,0})^{\mathrm{T}} = \boldsymbol{W}\boldsymbol{f}_0$, including all 2^L scaling coefficients and the $p - 2^L$ largest-magnitude detail coefficients (provided $p > 2^L$); and the modified vector $\boldsymbol{\theta}_0^{\#} = (\theta_{1,0}^{\#}, \ldots, \theta_{n,0}^{\#})^{\mathrm{T}}$ is obtained from $\boldsymbol{\theta}_0$ by setting to zero the $n - p$ nonselected

Procedure WDFTC

Phase I—Using the in-control mean profile $\boldsymbol{f}_0$ and the randomly sampled in-control profiles $\{\boldsymbol{Y}_j : j = 1,\ldots,N\}$, perform the following steps:

[1] Choose the optimal set of p nonzero components for the $n \times 1$ vector $\boldsymbol{\theta}_0^{\#}$ by selecting from the associated components of $\boldsymbol{\theta}_0$ so as to minimize the weighted relative reconstruction error (33) incurred by using $\boldsymbol{W}^{-1}\boldsymbol{\theta}_0^{\#}$ to approximate $\boldsymbol{f}_0$, where $2^L \le p \le n$. Assign $\boldsymbol{\vartheta}_0^{\#}$ as the $p \times 1$ version of $\boldsymbol{\theta}_0^{\#}$ from which all nonselected components have been deleted.

[2] Apply the covariance-matrix regularization scheme of Algorithm CMR below to the sample covariance matrix $\widehat{\boldsymbol{\Lambda}}_0^{\#}$ of the reduced-dimension DWTs $\{\boldsymbol{d}_j^{\#} : j = 1,\ldots,N\}$, thereby obtaining the regularized sample covariance matrix $\widetilde{\boldsymbol{\Lambda}}_0^{\#}$.

[3] Compute the batch size r using Algorithm BSD below.

[3a] For $k = 1,2,\ldots,\lfloor N/r \rfloor$, compute the kth batch-means vector $\boldsymbol{Y}_k(r) = r^{-1}\sum_{u=1}^{r}\boldsymbol{Y}_{(k-1)r+u}$, the associated batch-means DWT $\boldsymbol{d}_k(r) = \boldsymbol{W}\boldsymbol{Y}_k(r)$, and the reduced-dimension batch-means DWT $\boldsymbol{d}_k^{\#}(r)$ to obtain the Hotelling's statistic

$$T_k^2(r) = \left[\boldsymbol{d}_k^{\#}(r) - \boldsymbol{\vartheta}_0^{\#}\right]^{\mathrm{T}}\left[\widetilde{\boldsymbol{\Lambda}}_0^{\#}(r)\right]^{-1}\left[\boldsymbol{d}_k^{\#}(r) - \boldsymbol{\vartheta}_0^{\#}\right]. \tag{32}$$

[3b] From the Phase I statistics $\{T_k^2(r) : k = 1,\ldots,\lfloor N/r \rfloor\}$, compute the usual sample mean $\widehat{\mu}_{T^2(r)}$ and the sample variance $\widehat{\sigma}^2_{T^2(r)}$.

[4] Calculate the root H of the equation

$$\frac{\widehat{\sigma}^2_{T^2(r)}}{2K^2}\left(\exp\left\{\frac{2K\left[H + 1.166\widehat{\sigma}_{T^2(r)}\right]}{\widehat{\sigma}^2_{T^2(r)}}\right\} - 1 - \left\{\frac{2K\left[H + 1.166\widehat{\sigma}_{T^2(r)}\right]}{\widehat{\sigma}^2_{T^2(r)}}\right\}\right) = 2\mathrm{ARL}_0,$$

where $K = 0.1\widehat{\sigma}_{T^2(r)}$.

Phase II—For $k = 1,2,\ldots$, compute the kth batch-means vector $\boldsymbol{Y}_k(r)$ from the latest nonoverlapping batch of r profiles $\{\boldsymbol{Y}_j : j = (k-1)r+1,\ldots,kr\}$ observed in Phase II, and perform the following steps:

[5] Compute $\boldsymbol{d}_k(r)$ and its reduced-dimension counterpart $\boldsymbol{d}_k^{\#}(r)$ to obtain the associated Hotelling's statistic $T_k^2(r)$ as in Equation (32).

[6] Raise an alarm after observing the kth batch-means vector $\boldsymbol{Y}_k(r)$ if $S^+(k) \ge H$ or $S^-(k) \ge H$, where

$$S^{\pm}(k) = \begin{cases} 0, & \text{if } k = 0, \\ \max\left\{0,\ S^{\pm}(k-1) \pm \left[T_k^2(r) - \widehat{\mu}_{T^2(r)}\right] - K\right\}, & \text{if } k = 1,2,\ldots. \end{cases}$$

components of $\boldsymbol{\theta}_0$ so that the reconstructed vector $\boldsymbol{f}_0^{\#} = \boldsymbol{W}^{-1}\boldsymbol{\theta}_0^{\#}$ is a sufficiently accurate approximation to $\boldsymbol{f}_0$. In the following discussion, we will write $\boldsymbol{\theta}_0^{\#}$ and $\boldsymbol{f}_0^{\#}$ as $\boldsymbol{\theta}_0^{\#}(p)$ and $\boldsymbol{f}_0^{\#}(p)$, respectively, to emphasize the dependence of these vectors on p. When we use $\boldsymbol{f}_0^{\#}(p)$ as an approximation to $\boldsymbol{f}_0$, the relative reconstruction error is $\|\boldsymbol{f}_0^{\#}(p) - \boldsymbol{f}_0\|_2 / \|\boldsymbol{f}_0\|_2$ (Lada et al. [28]); and the corresponding data-compression ratio is p/n. For a given value of $p \in \{2^L, \ldots, n\}$ and weight $q \in [0,1]$ assigned to the data-compression ratio, we define the weighted relative reconstruction error (WRRE) as follows:

$$\begin{aligned} \mathrm{WRRE}(p; \boldsymbol{f}_0, q) &= (1-q)\left[\frac{\|\boldsymbol{W}^{-1}\boldsymbol{\theta}_0^{\#}(p) - \boldsymbol{f}_0\|_2}{\|\boldsymbol{f}_0\|_2}\right] + q\left(\frac{p}{n}\right) \\ &= (1-q)\left[\frac{\|\boldsymbol{f}_0^{\#}(p) - \boldsymbol{f}_0\|_2}{\|\boldsymbol{f}_0\|_2}\right] + q\left(\frac{p}{n}\right); \end{aligned} \tag{33}$$

and we choose p (and implicitly, $\boldsymbol{\theta}_0^{\#}(p)$) to minimize $\mathrm{WRRE}(p; \boldsymbol{f}_0, q)$,

$$p = \underset{u = 2^L, \ldots, n}{\arg\min}\ \mathrm{WRRE}(u; \boldsymbol{f}_0, q)\,. \tag{34}$$

Remark 10. There is a potential problem in using the dimension-reduction scheme of Equations (33) and (34) if $\mathbf{1}_n^{\mathrm{T}} \boldsymbol{f}_0 \neq 0$ and all the components of $\boldsymbol{f}_0$ have large magnitudes. In this situation, the relative reconstruction error can be negligibly small in comparison with the data-compression ratio for all feasible values of p so that Equation (34) yields $p = 2^L$; and then the only nonzero components of $\boldsymbol{\theta}_0^{\#}(p)$ are the scaling coefficients in $\boldsymbol{\theta}_0$, which can yield a low-resolution approximation to $\boldsymbol{f}_0$. The centering condition $\mathbf{1}_n^{\mathrm{T}} \boldsymbol{f}_0 = 0$ avoids this problem.

In the formulation (33), the weight q can be adjusted to achieve an effective trade-off between the relative reconstruction error and the data-compression ratio. In many applications of profile monitoring, the reduced dimension p must be sufficiently small to ensure that the Hotelling's statistics $\{T_k^2(r) : k = 1, 2, \ldots\}$ computed in Phase II have adequate power to detect shifts in the mean profile. On the other hand, p must be sufficiently large so that the selected scaling and detail coefficients in the DWT of an out-of-control profile can accurately represent deviations from the in-control mean profile. Setting the weight

$q = 0.5$ yields the same value of p as for the method of Lada et al. [28]. For profiles of moderate dimension (that is, $n \leq 1{,}000$), we found that $q = 0.5$ generally yielded satisfactory results. On the other hand, for profiles of dimension $n > 1{,}000$, we found that $q > 0.5$ was required to obtain acceptable results. In this chapter, we use $q = 0.7$ to handle profiles of dimension $n = 2{,}048$.

The effectiveness of the dimension-reduction scheme in WDFTC also depends on the coarsest level of resolution, L, based on the application at hand. For the choice of L to be used with WDFTC, we adapt the approach of Lada and Wilson [30] and use the default value $L = \lceil J/2 \rceil$, where $J = \log_2(n)$. In some cases we also use slightly smaller values of L than the default value (for example, $\lceil J/2 \rceil - 1$ or $\lceil J/2 \rceil - 2$), but only if such values of L yield a meaningful dimension reduction compared with that of the default value.

In some applications, the mean in-control profile $\boldsymbol{f}_0$ and its DWT $\boldsymbol{\theta}_0$ may not be known exactly. To estimate $\boldsymbol{f}_0$ in such cases, we use the centered sample mean

$$\widehat{\boldsymbol{f}}_0 = \left(\boldsymbol{I}_n - n^{-1}\mathbf{1}_n\mathbf{1}_n^{\mathrm{T}}\right)\left(N^{-1}\sum_{j=1}^{N}\boldsymbol{Y}_j\right)$$

of the profiles observed in Phase I, where $\boldsymbol{I}_n$ denotes the $n \times n$ identity matrix. Moreover from the DWT $\widehat{\boldsymbol{\theta}}_0 = \boldsymbol{W}\widehat{\boldsymbol{f}}_0$, we obtain the associated estimators $\widehat{\boldsymbol{\theta}}_0^{\#}(p)$ and $\widehat{\boldsymbol{f}}_0^{\#}(p) = \boldsymbol{W}^{-1}\widehat{\boldsymbol{\theta}}_0^{\#}(p)$ to be used in Equations (33) and (34) of the dimension-reduction scheme as well as the estimator $\widehat{\boldsymbol{\vartheta}}_0^{\#}$ to be used in computing Hotelling's statistic $T_k^2(r) = \left[\boldsymbol{d}_k^{\#}(r) - \widehat{\boldsymbol{\vartheta}}_0^{\#}\right]^{\mathrm{T}}\left[\widetilde{\boldsymbol{\Lambda}}_0^{\#}(r)\right]^{-1}\left[\boldsymbol{d}_k^{\#}(r) - \widehat{\boldsymbol{\vartheta}}_0^{\#}\right]$ for $k = 1, 2, \ldots$ in both Phases I and II of WDFTC.

4.2.3 Covariance-Matrix Regularization

In this subsection, we explain the covariance-matrix regularization method in step **[2]** of WDFTC that is applied to the $p \times p$ sample covariance matrix $\widehat{\boldsymbol{\Lambda}}_0^{\#} = (N-1)^{-1}\sum_{j=1}^{N}\left(\boldsymbol{d}_j^{\#} - \overline{\boldsymbol{d}}_N^{\#}\right)\left(\boldsymbol{d}_j^{\#} - \overline{\boldsymbol{d}}_N^{\#}\right)^{\mathrm{T}}$ of the reduced-dimension DWTs $\{\boldsymbol{d}_j^{\#} : j = 1, \ldots, N\}$ computed from the profiles observed in Phase I, where $\overline{\boldsymbol{d}}_N^{\#} = N^{-1}\sum_{j=1}^{N}\boldsymbol{d}_j^{\#}$. Commenting on the use of wavelets by Jin and Shi [20] for diagnosis of process faults, Woodall et al. [56] state that the use of Hotelling's T^2 statistic may not be efficient because high correlations between the components of each profile $\boldsymbol{Y}_j$ may lead to overparametrization—that is, an excessive value for

the dimension of p of the $\{\boldsymbol{d}_j^{\#}\}$. Moreover if $p > 200$, then estimating the $p \times p$ covariance matrix $\boldsymbol{\Lambda}_0^{\#}$ can also be difficult, especially if there is a limited amount of training data (see, for example, Hoffbeck and Landgrebe [16], Daniels and Kass [10], and Ledoit and Wolf [31]). In particular if the size N of the Phase I data set is insufficient or the joint distribution of each in-control random vector $\boldsymbol{d}_j^{\#}$ is singular, then $\widehat{\boldsymbol{\Lambda}}_0^{\#}$ is not guaranteed to be positive definite so that the associated Hotelling's T^2 statistic is not guaranteed to exist. In this chapter we make the following (mild) assumptions: (a) the $n \times 1$ profile vector $\boldsymbol{Y}_j$ has a nonsingular joint probability density function that depends on the current in-control or out-of-control condition; and (b) the covariance matrix $\mathrm{Cov}[\boldsymbol{Y}_j]$ is the same for both in-control and out-of-control conditions. Under assumptions (a) and (b), different profile components may have different continuous marginal distributions that may be nonnormal. In this broadly applicable setting if $N \geq p+1$, then $\widehat{\boldsymbol{\Lambda}}_0^{\#}$ is positive definite with probability one (see, for example, Proposition 2 of Porta Nova and Wilson [38]). To avoid problems with Hotelling's T^2 statistic when $p > 200$, we adapt the covariance-regularization method of Bickel and Levina [5] and use $\widetilde{\boldsymbol{\Lambda}}_0^{\#}$, the resulting thresholded version of $\widehat{\boldsymbol{\Lambda}}_0^{\#}$ in WDFTC. Although the main asymptotic results of Bickel and Levina [5] are based on the assumption that the profiles $\{\boldsymbol{Y}_j\}$ are randomly sampled from a Gaussian (normal) or sub-Gaussian distribution, we have found the authors' approach to be useful in formulating a covariance-matrix regularization procedure for WDFTC that is reasonably robust against violations of the normality assumption. As we shall see in the next subsection, the batch-size determination Algorithm CMR is also designed to avoid large departures from normality in the basic random vectors from which the relevant Hotelling's T^2 statistic is computed.

In the context of profile monitoring with WDFTC, the basic idea of the covariance-matrix regularization method of Bickel and Levina [5] is that if p and N are sufficiently large and $\log(p)/N$ is sufficiently small, then the $p \times p$ sample covariance matrix $\widehat{\boldsymbol{\Lambda}}_0^{\#}$ can be (hard) thresholded at a positive level τ depending on N and p such that with high probability, the thresholded covariance matrix $\widetilde{\boldsymbol{\Lambda}}_0^{\#}$ is positive definite and close to the theoretical covariance matrix $\boldsymbol{\Lambda}_0^{\#} = \mathrm{E}\big[\big(\boldsymbol{d}_j^{\#} - \mathrm{E}[\boldsymbol{d}_j^{\#}]\big)\big(\boldsymbol{d}_j^{\#} - \mathrm{E}[\boldsymbol{d}_j^{\#}]\big)^{\mathrm{T}}\big]$ in a certain sense. We adapt the thresholding scheme of Bickel and Levina [5] to WDFTC so that when it is applied to $\widehat{\boldsymbol{\Lambda}}_0^{\#}$, the following elements remain intact (i.e., are not subject to the thresholding operation):

(a) the $2^L \times 2^L$ submatrix of sample covariances of the estimated scaling coefficients (i.e, $\big[\widehat{\boldsymbol{\Lambda}}_0^{\#}\big]_{u,v}$ for $u,v = 1,\ldots,2^L$); and (b) the diagonal elements (i.e, $\big[\widehat{\boldsymbol{\Lambda}}_0^{\#}\big]_{u,u}$ for $u = 1,\ldots,p$). With the threshold τ, WDFTC's covariance-regularization scheme maps $\widehat{\boldsymbol{\Lambda}}_0^{\#}$ into the matrix $\boldsymbol{R}\big(\widehat{\boldsymbol{\Lambda}}_0^{\#}; L, \tau\big)$ whose (u,v) element is

$$\big[\boldsymbol{R}\big(\widehat{\boldsymbol{\Lambda}}_0^{\#}; L, \tau\big)\big]_{u,v} = \begin{cases} \big[\widehat{\boldsymbol{\Lambda}}_0^{\#}\big]_{u,v}, & \text{if } (u \le 2^L \text{ and } v \le 2^L) \text{ or } (u = v) \\ \big[\widehat{\boldsymbol{\Lambda}}_0^{\#}\big]_{u,v} 1\big(\big|\big[\widehat{\boldsymbol{\Lambda}}_0^{\#}\big]_{u,v}\big| \ge \tau\big), & \text{otherwise,} \end{cases} \tag{35}$$

where $1(\cdot)$ is the indicator function. Algorithm CMR below determines the estimated threshold $\widehat{\tau}$ and the "regularized" version $\widetilde{\boldsymbol{\Lambda}}_0^{\#}$ of the sample covariance matrix $\widehat{\boldsymbol{\Lambda}}_0^{\#}$ based on that threshold.

Algorithm CMR

[1] Divide the Phase I data set into two disjoint subsets of size $N_1 = \lfloor 0.4N \rfloor$ and $N_2 = N - N_1$, respectively.

[2] For $\ell = 1, 2$, calculate the sample covariance matrix $\widehat{\boldsymbol{\Lambda}}_0^{\#}(\ell, N_\ell)$ from the Phase I data subset of size N_ℓ.

[3] Compute the estimated threshold,

$$\widehat{\tau} = \underset{\tau \ge 0}{\arg\min} \sum_{\substack{(u \ne v) \text{ and} \\ (2^L < u \text{ or } 2^L < v)}} \Big\{ \big[\widehat{\boldsymbol{\Lambda}}_0^{\#}(1,N_1)\big]_{u,v} 1\big[\big|\big[\widehat{\boldsymbol{\Lambda}}_0^{\#}(1,N_1)\big]_{u,v}\big| \ge \tau\big] - \big[\widehat{\boldsymbol{\Lambda}}_0^{\#}(2,N_2)\big]_{u,v} \Big\}^2.$$

[4] Calculate the sample covariance matrix $\widehat{\boldsymbol{\Lambda}}_0^{\#}$ using the entire Phase I data set of size N and apply the covariance-regularization map of Equation (35) to $\widehat{\boldsymbol{\Lambda}}_0^{\#}$ using the threshold $\widehat{\tau}$:

$$\widetilde{\boldsymbol{\Lambda}}_0^{\#} = \boldsymbol{R}\big(\widehat{\boldsymbol{\Lambda}}_0^{\#}; L, \widehat{\tau}\big).$$

Remark 11. The estimated threshold $\widehat{\tau}$ can also be interpreted as the minimal magnitude for the sample covariances in $\widehat{\boldsymbol{\Lambda}}_0^{\#}$ to be considered "significant"; and this interpretation will play an important role in Algorithm BSD for determining the batch size r as detailed in the next subsection.

Remark 12. After a suitable batch size r is obtained from Algorithm BSD, we use $\widetilde{\boldsymbol{\Lambda}}_0^{\#}(r) = \widetilde{\boldsymbol{\Lambda}}_0^{\#}/r$ as our estimator of the covariance matrix of the reduced-dimension batch-means

DWTs $\{\boldsymbol{d}_k^{\#}(r) : k = 1, \ldots, \lfloor N/r \rfloor\}$ computed in Phase I; and then in both Phases I and II of WDFTC, we use $\widetilde{\boldsymbol{\Lambda}}_0^{\#}(r)$ to calculate the Hotelling's statistic $T_k^2(r) = \left[\boldsymbol{d}_k^{\#}(r) - \widehat{\boldsymbol{\vartheta}}_0^{\#}\right]^{\mathrm{T}} \left[\widetilde{\boldsymbol{\Lambda}}_0^{\#}(r)\right]^{-1} \left[\boldsymbol{d}_k^{\#}(r) - \widehat{\boldsymbol{\vartheta}}_0^{\#}\right]$ for $k = 1, 2, \ldots$.

4.2.4 Batch Size Determination

In this subsection, we explain the method used in WDFTC to determine the batch size r. In our experience excessive covariances between the components of the dimension-reduced DWTs $\{\boldsymbol{d}_j^{\#}\}$ can seriously distort the performance of a profile-monitoring chart based on a Hotelling's T^2–type statistic computed from the $\{\boldsymbol{d}_j^{\#}\}$ obtained in Phase I of the chart's operation. In this situation we have obtained substantial improvements in the performance of WDFTC by reducing the magnitudes of the covariances between pairs of components of the dimension-reduced DWTs to manageable levels. The desired covariance reductions are achieved indirectly by aggregating the observed profiles $\{\boldsymbol{Y}_j : j = 1, 2, \ldots\}$ into nonoverlapping batches of size r so that the associated nonoverlapping batch means $\left\{\boldsymbol{Y}_k(r) = r^{-1} \sum_{u=1}^{r} \boldsymbol{Y}_{(k-1)r+u} : k = 1, 2, \ldots\right\}$ yield batch-means DWT vectors $\left\{\boldsymbol{d}_k(r) = \boldsymbol{W}\boldsymbol{Y}_k(r) : k = 1, 2, \ldots\right\}$ for which $\mathrm{Cov}[\boldsymbol{d}_k(r)] = \mathrm{Cov}[\boldsymbol{d}_j]/r = \boldsymbol{\Lambda}_0/r$ and $\mathrm{Cov}[\boldsymbol{d}_k^{\#}(r)] = \mathrm{Cov}[\boldsymbol{d}_j^{\#}]/r = \boldsymbol{\Lambda}_0^{\#}/r$, where r is taken to be just large enough to achieve effective covariance reductions. The formal statement of Algorithm BSD is given below.

The basic idea of Algorithm BSD is first to compute the average magnitude of the elements of the regularized sample covariance matrix $\widetilde{\boldsymbol{\Lambda}}_0^{\#}$ as delivered by Algorithm CMR, where the average magnitude is taken only over the elements that were subjected to the thresholding operation and survived in Algorithm CMR; then the ratio of this average to the estimated threshold $\widehat{\tau}$ is an estimate of the batch size r necessary to reduce the magnitudes of all relevant covariances between pairs of components of the reduced-dimension batch-means vector $\boldsymbol{d}_j^{\#}(r)$ to "nonsignificant" levels.

Remark 13. Algorithm BSD is designed to yield a batch size r sufficiently large so that all the off-diagonal elements of the regularized sample covariance matrix $\widetilde{\boldsymbol{\Lambda}}_0^{\#}(r) = \widetilde{\boldsymbol{\Lambda}}_0^{\#}/r$ have sufficiently small magnitudes to avoid aberrant behavior of the profile-monitoring statistic $T_k^2(r)$. In particular, the inflation factor $\sqrt{2}$ forces Algorithm BSD to yield a batch size

$r > 1$ for most processes, provided that Algorithm CMR delivers at least one nonzero off-diagonal element in the regularized sample covariance matrix $\widetilde{\mathbf{\Lambda}}_0^{\#}$, excluding the estimated covariances between pairs of scaling coefficients. Note that Algorithm BSD is based on the (regularized) sample covariance matrix and thus the resulting batch size may be too small without an inflation factor in some cases. The inflation factor $\sqrt{2}$ ensures that Algorithm BSD yields sufficiently large batch size for such cases.

Remark 14. When the true covariance matrix $\mathbf{\Lambda}_0^{\#}(r)$ is used to calculate the profile-monitoring statistic $T_k^2(r)$, then we have the in-control mean $\mathrm{E}[T_k^2(r)] = p$ regardless of the distribution of the profiles $\{\boldsymbol{Y}_j\}$, provided that the latter distribution is nonsingular. Thus one can check if the regularized matrix $\widetilde{\mathbf{\Lambda}}_0^{\#}(r)$ is a good estimate of $\mathbf{\Lambda}_0^{\#}(r)$ by comparing the sample average of the in-control statistics $\{T_k^2(r) : k = 1, \ldots, \lfloor N/r \rfloor\}$ to its theoretical expected value p.

Algorithm BSD

[1] Apply Algorithm CMR to obtain $\widehat{\tau}$, the estimated threshold, and $\widetilde{\mathbf{\Lambda}}_0^{\#}$, the regularized (thresholded) version of $\widehat{\mathbf{\Lambda}}_0^{\#}$.

[2] Let Q denote the number of nonzero off-diagonal elements of $\widetilde{\mathbf{\Lambda}}_0^{\#}$, excluding the estimated covariances between pairs of scaling coefficients,

$$Q = \sum_{\substack{(u \neq v) \text{ and} \\ (2^L < u \text{ or } 2^L < v)}} 1\big(\big|\big[\widetilde{\mathbf{\Lambda}}_0^{\#}\big]_{u,v}\big| > 0\big)\,.$$

[2a] If $Q = 0$, then return $r \leftarrow 1$ and stop; otherwise, go to step **[2b]**.

[2b] Calculate the average magnitude ζ of the nonzero off-diagonal elements of $\widetilde{\mathbf{\Lambda}}_0^{\#}$, excluding the estimated covariances between pairs of scaling coefficients,

$$\zeta = \frac{1}{Q} \sum_{\substack{(u \neq v) \text{ and} \\ (2^L < u \text{ or } 2^L < v)}} \big|\big[\widetilde{\mathbf{\Lambda}}_0^{\#}\big]_{u,v}\big|\,.$$

[3] Set the batch size $r \leftarrow \lceil \sqrt{2}\zeta/\widehat{\tau} \rceil$ and stop.

4.3. Experiments

In this section, we present experimental results for WDFTC in comparison with other existing profile-monitoring charts. The following three charts are considered: (a) HTW$_n$, the classical Hotelling's T^2 chart based on the full $n \times 1$ vector of wavelet coefficients for each observed profile; (b) HTW$_p$, a reduced-dimension variant of HTW$_n$ that is based on p preselected wavelet coefficients for each observed profile as detailed below; and (c) the M^* chart of Chicken et al. [9] as described in the second section of this chapter. Concise summaries of the steps of procedures HTW$_n$ and HTW$_p$ are given below.

Chart HTW$_n$: Compute the exact covariance matrix $\boldsymbol{\Lambda}_0 = \boldsymbol{W}\boldsymbol{\Sigma}_0\boldsymbol{W}^{\mathrm{T}}$ for the DWTs $\{\boldsymbol{d}_j : j = 1,\ldots,N\}$ of the profiles observed in Phase I, where $\boldsymbol{\Sigma}_0$ is assumed to be known. In terms of the prespecified false-alarm rate FAR $= 1/\mathrm{ARL}_0$, calculate the upper control limit UCL$_1$ for the "ideal" profile-monitoring statistic $T_j^2 = (\boldsymbol{d}_j - \boldsymbol{\theta}_0)^{\mathrm{T}}\boldsymbol{\Lambda}_0^{-1}(\boldsymbol{d}_j - \boldsymbol{\theta}_0)$ as the $1 - \mathrm{FAR}$ quantile of the chi-squared distribution with n degrees of freedom—that is, UCL$_1$ is the solution of the equation $\Pr\{\chi_n^2 \leq \mathrm{UCL}_1\} = 1 - \mathrm{FAR}$, where χ_n^2 denotes a chi-squared random variable with n degrees of freedom. After the jth profile is observed in Phase II, an out-of-control alarm is raised if $T_j^2 > \mathrm{UCL}_1$.

Chart HTW$_p$: Compute the exact covariance matrix $\boldsymbol{\Lambda}_0 = \boldsymbol{W}\boldsymbol{\Sigma}_0\boldsymbol{W}^{\mathrm{T}}$ for the DWTs $\{\boldsymbol{d}_j : j = 1,\ldots,N\}$ of the profiles observed in Phase I, where $\boldsymbol{\Sigma}_0$ is assumed to be known. Select the p largest-magnitude components of the DWT $\boldsymbol{\theta}_0 = \boldsymbol{W}\boldsymbol{f}_0$ of the mean in-control profile; and for the corresponding $p \times 1$ subvectors $\{\boldsymbol{d}_j^{\#} : j = 1,\ldots,N\}$ extracted from the DWTs of the profiles observed in Phase I, let $\boldsymbol{\Lambda}_0^{\#}$ denote the associated covariance matrix (a submatrix of $\boldsymbol{\Lambda}_0$). In terms of the prespecified false-alarm rate FAR $= 1/\mathrm{ARL}_0$, calculate the upper control limit UCL$_2$ for the "ideal" reduced-dimension profile-monitoring statistic $T_j^2 = \left(\boldsymbol{d}_j^{\#} - \boldsymbol{\theta}_0^{\#}\right)^{\mathrm{T}}\left(\boldsymbol{\Lambda}_0^{\#}\right)^{-1}\left(\boldsymbol{d}_j^{\#} - \boldsymbol{\theta}_0^{\#}\right)$ as the $1 - \mathrm{FAR}$ quantile of the chi-squared distribution with p degrees of freedom—that is, UCL$_2$ is the solution of the equation $\Pr\{\chi_p^2 \leq \mathrm{UCL}_2\} = 1 - \mathrm{FAR}$. An out-of-control alarm is raised after the jth profile if $T_j^2 > \mathrm{UCL}_2$.

Remark 15. The p components of $\boldsymbol{\theta}_0$ that are selected for use in HTW$_p$ may be different from the p components of $\boldsymbol{\theta}_0$ that minimize the weighted relative reconstruction error (33).

In all the experiments reported below, we used the exact values of the covariance matrices $\mathbf{\Sigma}_0$, $\mathbf{\Lambda}_0$, and $\mathbf{\Lambda}_0^{\#}$ as required for procedures HTW$_n$ and HTW$_p$. Recall that procedure M^* estimates σ_* from the average of the median absolute deviations of the $n/2$ coefficients at the highest levels of resolution for each of the profiles observed so far in Phase II operation. Moreover, WDFTC uses the regularized sample covariance matrix, $\widetilde{\mathbf{\Lambda}}_0^{\#}$, computed from the Phase I data set of size N. In this respect the procedures HTW$_n$ and HTW$_p$ have some advantage over procedures M^* and WDFTC in the experimental performance evaluation that may not carry over to practical applications in which $\mathbf{\Sigma}_0$, $\mathbf{\Lambda}_0$, and $\mathbf{\Lambda}_0^{\#}$ are unknown and must be estimated from a Phase I data set. For profiles of dimension $n = 512$, we applied WDFTC with a Phase I data set of size $N = 3{,}000$; and for profiles of dimension $n = 2{,}048$, we applied WDFTC with a Phase I data set of size $N = 5{,}000$.

In the first part of the experimental performance evaluation of WDFTC and its competitors HTW$_n$, HTW$_p$, and M^*, we applied those procedures to both normal and nonnormal profiles having both independent and correlated components such that the mean in-control profile $\boldsymbol{f}_0$ is defined by $n = 512$ equally spaced points on Mallat's piecewise smooth function as depicted in Figure 1. In the second part of the experimental performance evaluation, we applied WDFTC to a lumber manufacturing process (Staudhammer [49]) in which the mean in-control profile had $n = 2{,}048$ points. In both applications, we estimated the relevant in-control and out-of-control ARLs based on 1,000 independent replications of each test process. Corresponding to each table of estimated ARLs given in this section, there is a matching table of standard errors for those estimated ARLs that is given in the Appendix C to this chapter.

4.3.1 Profiles Based on Mallat's Piecewise Smooth Function

In the experiments with the mean in-control profile $\boldsymbol{f}_0$ based on Mallat's piecewise smooth function, we set the target value of $\text{ARL}_0 = 200$. The mean out-of-control profile has the form $\boldsymbol{f}_1 = \boldsymbol{f}_0 + \eta\mathbf{\Delta}\boldsymbol{\sigma}$, where: (a) the shift-size parameter $\eta \in \{0.25, 0.5, 0.75, 1, 2\}$; (b) the $n \times n$ shift-sign matrix $\mathbf{\Delta} = \text{diag}(\delta_1, \ldots, \delta_n)$ is a diagonal matrix with $\delta_i \in \{-1, 0, 1\}$ for $i = 1, \ldots, n$; and (c) $\boldsymbol{\sigma} = (\sigma_1, \ldots, \sigma_n)^{\text{T}}$ is the vector of marginal standard deviations of the components of $\boldsymbol{\varepsilon}_j$. Whereas procedure M^* is based on the Haar wavelet system, we

used the Symmlet 8 wavelet system in procedures WDFTC, HTW$_n$, and HTW$_p$. Because $n = 512$, the highest level of resolution $J = \log_2(n) = 9$; and selecting the coarsest level of resolution $L = \lceil J/2 \rceil = 5$ and the weight $q = 0.5$ in Equation (33) for the weighted relative reconstruction error, we obtain the "optimal" reduced dimension $p = 62$ from Equation (34). To make a fair comparison of WDFTC with HTW$_p$, we also set $p = 62$ in the latter chart.

In the following tables, Global Shift 1 refers to the situation in which $\delta_i = 1$ for $i = 1, \ldots, n$ so that there is a positive shift of size $\eta\sigma_i$ in the ith component of the mean profile for $i = 1, \ldots, n$. By contrast, Global Shift 2 refers to the situation in which $\delta_i = 1$ for $i = 1, \ldots, n/2$ and $\delta_i = -1$ for $i = (n/2)+1, \ldots, n$; therefore in the first half of the components of the mean profile, there are positive shifts of the respective amounts $\eta\sigma_1, \ldots, \eta\sigma_{n/2}$, and in the last half of the components of the mean profile there are negative shifts of the respective amounts $-\eta\sigma_{(n/2)+1}, \ldots, -\eta\sigma_n$. Local Shift 1 is specified as follows: $\delta_i = 1$ for $i \in \mathscr{A}_1 = \{73, 74, \ldots, 76\} \cup \{288, 289, \ldots, 296\}$, and $\delta_i = 0$ for $i \notin \mathscr{A}_1$. Therefore with Local Shift 1, the 13 selected components of the mean profile are increased by the respective amounts $\eta\sigma_i$ for $i \in \mathscr{A}_1$, while all other components of the mean profile remain unchanged. Local Shift 2 is specified as follows: $\delta_i = 1$ for $i \in \mathscr{A}_2 = \{3, 4, \ldots, 15\} \cup \{344, 345, \ldots, 347\}$, and $\delta_i = 0$ for $i \notin \mathscr{A}_2$. Therefore with Local Shift 2, the 17 selected components of the mean profile are increased by the respective amounts $\eta\sigma_i$ for $i \in \mathscr{A}_2$, while other components of the mean profile remain unchanged.

4.3.1.1 Multivariate Normal Marginals

Most existing profile-monitoring charts assume that the observed profiles $\{\boldsymbol{Y}_j : j = 1, 2, \ldots\}$ are i.i.d. multivariate normal vectors with a common marginal variance and zero correlations between each pair of components. With $\boldsymbol{f}_0$ based on Mallat's piecewise smooth function, we first consider the following three cases in which the error vector $\boldsymbol{\varepsilon}_j$ is multivariate normal with mean $\mathbf{0}_n$ and covariance matrix $\boldsymbol{\Sigma}_0$: (a) the components of $\boldsymbol{\varepsilon}_j$ are independent standard normal random variables so that $\boldsymbol{\Sigma}_0 = \boldsymbol{I}_n$; (b) the components of $\boldsymbol{\varepsilon}_j$ are correlated standard normal random variables with common correlation 0.5 so that $\boldsymbol{\Sigma}_0$ has all its diagonal elements equal to 1.0 and all its off-diagonal elements equal to 0.5; and (c)

the components of $\boldsymbol{\varepsilon}_j$ are correlated normal random variables with mean zero, marginal variances given by Equation (29), and pairwise correlations given by Equation (28) so that $\left[\boldsymbol{\Sigma}_0\right]_{u,v} = \sigma_u \sigma_v \rho(u-v)$ for $u, v = 1, \ldots, n$ as for the test processes ME_1 and ME_2.

Case (a): Error Vector Has Independent Standard Normal Components. Table 9 shows the values of ARL_0 and ARL_1 delivered by WDFTC and its competitors for Case (a). WDFTC required the average batch size $\bar{r} = 3$. All four charts yielded values for ARL_0 close to the target value of 200. To detect Global Shifts 1 and 2 with $\eta > 0.25$, WDFTC required one batch-mean vector (that is, 3 observed profiles), whereas each of the other charts required 1 observation. To detect Global Shifts 1 and 2 with $\eta = 0.25$, WDFTC sometimes required two batch means so that on average WDFTC required about 4 observations; by contrast, HTW_n required about 16 observations, while HTW_p and M^* each required about 2 observations. For the Local Shifts 1 and 2 with $0.25 \leq \eta \leq 1$, WDFTC significantly outperformed all the other charts, and HTW_n usually delivered the worst performance. The latter conclusion is not surprising, because high dimensionality degrades the performance of Hotelling's T^2–type charts (Fan [13]). For Local Shifts 1 and 2 with $0.25 \leq \eta \leq 1$, the performance of M^* was often similar to that of HTW_n and was always much worse than that of WDFTC. For example, to detect Local Shift 1 with $\eta = 0.5$, charts M^* and HTW_n each required about 145 observations, while WDFTC required about 35 observations.

Case (b): Error Vector Has Correlated Standard Normal Components. Table 10 shows the values of ARL_0 and ARL_1 delivered by WDFTC and its competitors for Case (b). As we saw in Case (a), WDFTC required the average batch size $\bar{r} = 3$, and all four charts yielded values for ARL_0 close to the target value of 200. However for Global Shift 1 and all levels of η, the introduction of a common correlation of 0.5 significantly increased the value of ARL_1 for all four charts compared with the results for Case (a). For example, in Case (a) to detect Global Shift 1 with $\eta = 0.5$, WDFTC required about 3 observations while HTW_n, HTW_p, and M^* each required about 1 observation; by contrast in Case (b) the corresponding values of ARL_1 for WDFTC, HTW_n, HTW_p, and M^* were about 134, 190, 180, and 86 observations, respectively. Overall in Case (b) for Global Shift 1, M^* significantly outperformed WDFTC, which in turn significantly outperformed HTW_n and HTW_p. To detect Global Shift 2 at all levels of η, WDFTC required about 3 observations

Table 9. ARLs for error vector with independent standard normal components

Shift Type	Shift Size	WDFTC $\bar{r}=3$	HTW_n	HTW_p	M^*
In-Control	0	189.97	210.56	197.81	196.07
Global Shift 1	0.25	3.80	16.13	2.22	2.12
	0.5	3.00	1.18	1.00	1.04
	0.75	3.00	1.00	1.00	1.00
	1	3.00	1.00	1.00	1.00
	2	3.00	1.00	1.00	1.00
Global Shift 2	0.25	3.86	16.32	2.34	1.42
	0.5	3.00	1.15	1.00	1.05
	0.75	3.00	1.00	1.00	1.00
	1	3.00	1.00	1.00	1.00
	2	3.00	1.00	1.00	1.00
Local Shift 1	0.25	114.41	191.78	164.74	183.95
	0.5	35.04	145.24	95.74	145.35
	0.75	16.04	101.19	44.07	88.77
	1	9.47	65.16	18.58	43.98
	2	3.06	6.35	1.39	2.65
Local Shift 2	0.25	112.09	179.01	166.38	197.08
	0.5	33.06	135.86	89.34	131.01
	0.75	15.14	84.56	40.69	73.69
	1	8.82	49.35	15.57	28.19
	2	3.04	3.16	1.31	2.05

while HTW_p required about 1 observation; on the other hand, the values of ARL_1 for HTW_n ranged from approximately 4 observations (for $\eta = 0.25$) to approximately 1 observation (for $\eta > 0.25$), and the values of ARL_1 for M^* ranged from approximately 9 observations (for $\eta = 0.25$) to approximately 2 observations (for $\eta = 2$). For Local Shifts 1 and 2 and all levels of η, WDFTC substantially outperformed M^*; and for Local Shifts 1 and 2 with $0.25 \leq \eta \leq 0.75$, WDFTC substantially outperformed HTW_p, which in turn outperformed HTW_n. For example, to detect Local Shift 2 with $\eta = 0.5$, WDFTC, HTW_n, HTW_p, and M^* required approximately 18, 93, 48, and 199 observations, respectively.

Case (c): Error Vector Has a General Normal Distribution. Table 11 shows the values of ARL_0 and ARL_1 delivered by WDFTC and its competitors for Case (c). For this test process, WDFTC required the average batch size $\bar{r} = 8$. As we saw in Cases (a) and (b), all four charts yielded values for ARL_0 close to the target value of 200. Because of the batching operation, WDFTC usually required at least 8 observations to detect shifts of any type. For Global Shift 1 with all levels of η, HTW_n and HTW_p outperformed WDFTC,

Table 10. ARLs for error vector with correlated standard normal components

Shift Type	Shift Size	WDFTC $\bar{r} = 3$	HTW$_n$	HTW$_p$	M^*
In-Control	0	188.73	210.65	198.30	200.48
Global Shift 1	0.25	174.96	210.21	197.38	153.84
	0.5	134.04	189.50	180.23	85.59
	0.75	78.60	171.14	149.24	39.77
	1	47.08	163.73	110.20	20.75
	2	12.46	94.86	29.09	3.02
Global Shift 2	0.25	3.01	3.79	1.07	8.84
	0.5	3.01	1.15	1.00	5.11
	0.75	3.01	1.00	1.00	3.89
	1	3.01	1.00	1.00	2.97
	2	3.01	1.00	1.00	1.97
Local Shift 1	0.25	71.47	177.53	140.86	200.55
	0.5	18.54	112.10	50.69	203.18
	0.75	8.75	58.10	16.30	198.04
	1	5.56	25.88	4.95	196.10
	2	3.01	1.54	1.00	168.72
Local Shift 2	0.25	65.51	163.70	135.48	201.31
	0.5	17.57	92.67	47.87	198.74
	0.75	8.38	43.98	13.69	197.23
	1	5.41	15.86	4.44	196.54
	2	3.01	1.13	1.00	164.76

and WDFTC substantially outperformed M^*. For example, to detect Global Shift 1 with $\eta = 0.5$, WDFTC, HTW$_n$, HTW$_p$, and M^* delivered ARL$_1$ values of approximately 8, 1, 1, and 175 observations, respectively. To detect Global Shift 2 at all levels of η, WDFTC required about 8 observations, while HTW$_p$ required about 1 observation; on the other hand the values of ARL$_1$ for HTW$_n$ ranged from approximately 3 observations (for $\eta = 0.25$) to approximately 1 observation (for $\eta > 0.25$), and the values of ARL$_1$ for M^* ranged from 5 observations (for $\eta = 0.25$) to 1 observation (for $\eta = 2$). For Local Shifts 1 and 2 and all levels of η, WDFTC substantially outperformed M^*; and for Local Shifts 1 and 2 with $0.25 \leq \eta \leq 0.75$, WDFTC substantially outperformed HTW$_p$, which in turn outperformed HTW$_n$. For example, to detect Local Shift 2 with $\eta = 0.5$, WDFTC, HTW$_n$, HTW$_p$, and M^* required approximately 12, 73, 39, and 201 observations, respectively.

Table 11. ARLs for error vector with a general normal distribution

Shift Type	Shift Size	WDFTC $\bar{r}=8$	HTW_n	HTW_p	M^*
In-Control	0	198.99	210.65	201.02	201.44
Global Shift 1	0.25	8.01	3.00	1.04	197.46
	0.5	8.01	1.00	1.00	175.29
	0.75	8.01	1.00	1.00	123.74
	1	8.01	1.00	1.00	77.63
	2	8.01	1.00	1.00	14.97
Global Shift 2	0.25	8.01	2.98	1.03	5.29
	0.5	8.01	1.00	1.00	3.03
	0.75	8.01	1.00	1.00	2.11
	1	8.01	1.00	1.00	1.98
	2	8.01	1.00	1.00	1.00
Local Shift 1	0.25	57.99	179.06	144.43	201.04
	0.5	17.22	113.26	63.98	200.99
	0.75	9.19	56.98	21.46	199.10
	1	8.08	25.91	7.74	198.38
	2	8.01	1.50	1.03	194.82
Local Shift 2	0.25	39.30	145.23	119.47	201.03
	0.5	11.96	73.24	38.99	200.64
	0.75	8.17	26.50	10.18	199.89
	1	8.01	9.13	2.84	199.82
	2	8.01	1.02	1.00	198.89

4.3.1.2 Multivariate Shifted Exponential Errors

To demonstrate the distribution-free aspect of WDFTC, in this subsection we consider two cases in which the error vector $\boldsymbol{\varepsilon}_j$ has a multivariate exponential distribution, but $\boldsymbol{f}_0$ is still based on Mallat's piecewise smooth function: (a) the components of $\boldsymbol{\varepsilon}_j$ are independent shifted standard exponential random variables with mean zero and standard deviation one; and (b) the components of $\boldsymbol{\varepsilon}_j$ are shifted standard exponential random variables generated via the NORTA method (Cario and Nelson [8]) so that a standard normal vector with common correlation 0.5 between each pair of components is transformed into $\boldsymbol{\varepsilon}_j$, yielding a pairwise correlations between components of $\boldsymbol{\varepsilon}_j$ that are slightly less than 0.5 on the average.

In the following tables, we only report the ARLs delivered by HTW_n and HTW_p using the control limits based on calibration method CM_A. In the previous subsection titled Motivating Examples, we concluded that the performance of M^* was not acceptable when the noise components have exponential marginals. Therefore in the following tables, we

only report the ARLs delivered by M^* using the control limits based on calibration method $\mathrm{CM_B}$.

Case (a): Error Vector Has Independent Shifted Standard Exponential Components. Table 12 shows the values of ARL_0 and ARL_1 delivered by WDFTC and its competitors for Case (a). WDFTC required the average batch size $\bar{r} = 3$. The small values of ARL_0 for HTW_n and HTW_p (approximately 11 and 36, respectively) led us to conclude that those charts were not robust against departures from normality. On the other hand by exploiting its readily computed, distribution-free control limits, WDFTC delivered $\mathrm{ARL}_0 \approx 194$, which did not deviate significantly from the target value of 200; moreover, WDFTC substantially outperformed M^* for Global Shift 1 and for Local Shifts 1 and 2 at all levels of η. For Global Shift 2, WDFTC delivered values of ARL_1 ranging from approximately 4 (for $\eta = 0.25$) to 3 (for $0.5 \leq \eta \leq 2$), while M^* delivered values of ARL_1 ranging from approximately 5 (for $\eta = 0.25$) to 1 (for $\eta = 2$). All in all, the performance of WDFTC in the case of shifted standard exponential errors provided good evidence of the chart's effectiveness and robustness.

Case (b): Vector Errors Has Correlated Shifted Standard Exponential Components. Table 13 shows the values of ARL_0 and ARL_1 delivered by WDFTC and its competitors for Case (b). WDFTC required the average batch size $\bar{r} = 3$. The extremely small values of ARL_0 for HTW_n and HTW_p (approximately 3 and 5, respectively) reinforced our conclusion that those charts were not robust against departures from normality. Both WDFTC and M^* delivered values of ARL_0 close to the target value of 200; but whereas the control limits for WDFTC are easily evaluated, the control limits for M^* must be estimated by cumbersome, compute-intensive simulation experiments. For Global Shift 1 with $\eta = 0.25$, WDFTC and M^* performed about the same, delivering ARL_1 values of approximately 170 and 176, respectively; but M^* significantly outperformed WDFTC for $0.5 \leq \eta \leq 2$. For Global Shift 2, WDFTC delivered values of ARL_1 ranging from about 6 (for $\eta = 0.25$) to about 3 (for $0.5 \leq \eta \leq 2$), while M^* delivered values of ARL_1 ranging from about 10 (for $\eta = 0.25$) to about 2 (for $\eta = 2$). For Local Shifts 1 and 2 with all values of η, WDFTC substantially outperformed M^*. For example, to detect Local Shift 2

Table 12. ARLs for error vector with independent shifted standard exponential components

Shift Type	Shift Size	WDFTC $\bar{r}=3$	HTW_n CM_A	HTW_p CM_A	M^* CM_B
In-Control	0	193.85	11.42	35.90	200.24
Global Shift 1	0.25	4.24	–	–	82.77
	0.5	3.00	–	–	14.43
	0.75	3.00	–	–	4.40
	1	3.00	–	–	2.01
	2	3.00	–	–	1.00
Global Shift 2	0.25	4.22	–	–	4.74
	0.5	3.00	–	–	2.90
	0.75	3.00	–	–	2.00
	1	3.00	–	–	1.78
	2	3.00	–	–	1.00
Local Shift 1	0.25	123.55	–	–	200.26
	0.5	39.20	–	–	197.22
	0.75	18.07	–	–	194.17
	1	10.55	–	–	186.59
	2	3.17	–	–	110.55
Local Shift 2	0.25	122.26	–	–	200.87
	0.5	35.62	–	–	198.08
	0.75	16.54	–	–	197.37
	1	9.87	–	–	189.51
	2	3.07	–	–	96.01

with $\eta = 0.5$, WDFTC required approximately 60 observations, while M^* required approximately 224 observations. The results for Case (b) provided further evidence of WDFTC's robustness and effectiveness.

4.3.2 Laser Range Sensor Data

In this subsection, we summarize the experimental results for an application of WDFTC to laser range sensor (LRS) data observed in a lumber-manufacturing process. LRS equipment can measure the thickness of a sawed board with a high degree of accuracy, and the development of such equipment has provided ample opportunities for quality engineers in the industry to improve and maintain the quality of the manufactured boards (Staudhammer [49]).

Figure 2 is a plot of a sample stream of board-thickness measurements taken along the length of a certain type of board from a particular sensor location as detailed in Staudhammer et al. [50]. Four laser sensors are set up to measure the thickness of sawed boards of

Table 13. ARLs for error vector with correlated shifted standard exponential components

Shift Type	Shift Size	WDFTC $\bar{r}=3$	HTW$_n$ CM$_A$	HTW$_p$ CM$_A$	M^* CM$_B$
In-Control	0	197.06	2.83	4.68	200.85
Global Shift 1	0.25	170.16	–	–	176.49
	0.5	163.11	–	–	115.93
	0.75	141.20	–	–	65.55
	1	110.05	–	–	39.91
	2	34.84	–	–	5.90
Global Shift 2	0.25	5.90	–	–	9.69
	0.5	3.01	–	–	5.61
	0.75	3.01	–	–	4.01
	1	3.01	–	–	3.17
	2	3.01	–	–	1.93
Local Shift 1	0.25	171.70	–	–	230.75
	0.5	61.63	–	–	233.84
	0.75	28.96	–	–	223.00
	1	15.94	–	–	217.93
	2	5.13	–	–	197.92
Local Shift 2	0.25	165.25	–	–	226.96
	0.5	59.68	–	–	224.39
	0.75	26.22	–	–	227.72
	1	14.37	–	–	228.05
	2	4.89	–	–	198.83

various types at two different locations on both sides of the board. In this subsection, we use the thickness measurements from one laser location only, but the measurements from all four laser locations can easily be incorporated to monitor various kinds of board defects as detailed below.

For each sawed board, over 2,000 thickness measurements are taken from each laser location; and the physical proximity of the locations on the board for successive thickness measurements naturally induces correlation between those measurements. On the other hand, Staudhammer [49] finds that there is no significant correlation between measurements taken on different boards, and she formulates statistical models to describe the variation in board thickness along the length of each individual board. Staudhammer proposes new profile-monitoring charts to detect various types of board defects, and she evaluates the performance of those charts using a comprehensive simulation study based on the proposed statistical models.

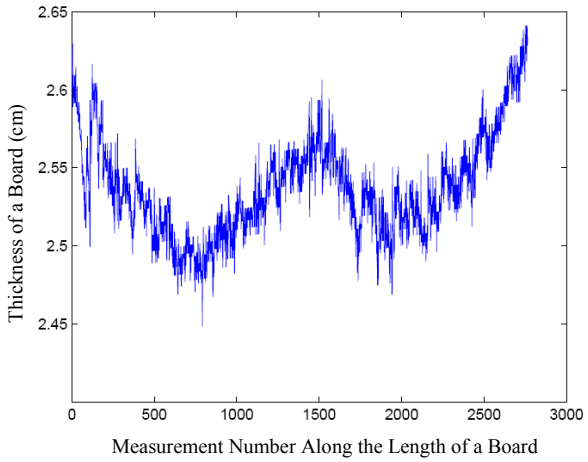

Figure 2. Sample stream of board-thickness measurements

From Equation (1) of Staudhammer et al. [50], we see that for a particular saw configuration, type of board, and side of the board, the statistical model for the thickness of the uth board (expressed in cm) as measured from the vth laser location at the ith horizontal distance x_i cm along the length of the board has the form

$$y_{uvi} = \mu_0 + \mathcal{B}_u + \mathcal{L}_v + \mathcal{B}\mathcal{L}_{uv} + \varepsilon_{uvi} \text{ for } i = 1, \ldots, n, \tag{36}$$

where: (a) μ_0 is the true mean in-control board thickness taken over the population of sawed boards defined by the given saw configuration, type of board, and side of the board; (b) $\mathcal{B}_u$ is the random board effect for the uth sample board, which is assumed to be i.i.d. $N(0, \sigma^2_{\mathcal{B}})$; (c) $\mathcal{L}_v$ is the random effect of the vth laser location, which is assumed to be i.i.d. $N(0, \sigma^2_{\mathcal{L}})$; (d) $\mathcal{B}\mathcal{L}_{uv}$ is the random effect arising from the interaction of the board and laser-location effects, which is assumed to be i.i.d. $N(0, \sigma^2_{\mathcal{B}\mathcal{L}})$; and (e) ε_{uvi} is the residual error associated with the thickness measurement taken on the uth board from the vth laser location at the ith distance x_i along the board so that the error process $\{\varepsilon_{uvi} : i = 1, \ldots n\}$ is assumed to be stationary and correlated with marginal distribution $N(0, \sigma^2_{\varepsilon})$. Staudhammer et al. [50] obtain the following parameter estimates for the model (36): $\widehat{\sigma}_{\mathcal{B}} = 0.0204$ cm, $\widehat{\sigma}_{\mathcal{L}} = 0.0052$ cm, and $\widehat{\sigma}_{\mathcal{B}\mathcal{L}} = 0.0238$ cm.

The authors find that for the board type BB considered in this chapter and for each value of u and v, the error process $\{\varepsilon_{uvi} : i = 1, \ldots n\}$ can be adequately represented by an ARIMA(1,1,1) time series model,

$$(1 - \alpha B)(\varepsilon_i - \varepsilon_{i-1}) = (1 - \beta B)\varepsilon_i \text{ for } i = 1, 2, \ldots, \tag{37}$$

where: (a) B is the backshift operator so that $(1 - \alpha B)\varepsilon_i = \varepsilon_i - \alpha\varepsilon_{i-1}$; and (b) the white-noise process $\{\varepsilon_i : i = 1, 2, \ldots\}$ consists of i.i.d. $N(0, \sigma_\varepsilon^2)$ random variables. The authors obtain the following parameter estimates for the error model (37): the autoregressive parameter $\widehat{\alpha} = 0.00053$ cm, the moving-average parameter $\widehat{\beta} = 0.00178$ cm, and the white-noise standard deviation $\widehat{\sigma}_\varepsilon = 0.00967$ cm. We apply WDFTC to board-thickness data generated according to the statistical model specified by Equations (36) and (37); and we compare the performance of WDFTC with that of the profile-monitoring charts proposed by Staudhammer et al. [50, 51] for detecting various types of defects in the lumber-manufacturing process.

Rasmussen et al. [42] identify common defects that can arise in lumber manufacturing. In the experimental performance evaluation, we consider four such defects: the machine positioning problem, taper, flare, and snake. Taken from Staudhammer [49], Figure 3 illustrates all four types of defects.

The machine positioning problem (MPP) is one of the simplest defects, resulting in a uniform change in thickness along the length of the board. The taper defect results in a gradual thickening or thinning along the length of the board. The flare is one of the more complex defects, which results in progressive board thickening only at the end of the board. The snake is another complex defect that causes high within-board variation of the board's thickness along the length of the board. For more-detailed descriptions of the these defects, see Staudhammer [49]. We use the following synthetic out-of-control conditions with various levels of severity to simulate all four types of defects as follows:

- For the MPP defect, we used the out-of-control mean $\mu_1 = \mu_0 + \delta$, where the shift $\delta \in \{0.0254, 0.0508, 0.0762, 0.1016\}$ (expressed in cm).

- For the taper defect, we took $\mathrm{E}[y_{uvi}] = \mu_0 + x_i\delta/x_n$ for $i = 1, \ldots, n$ and $\delta \in \{0.0508,$

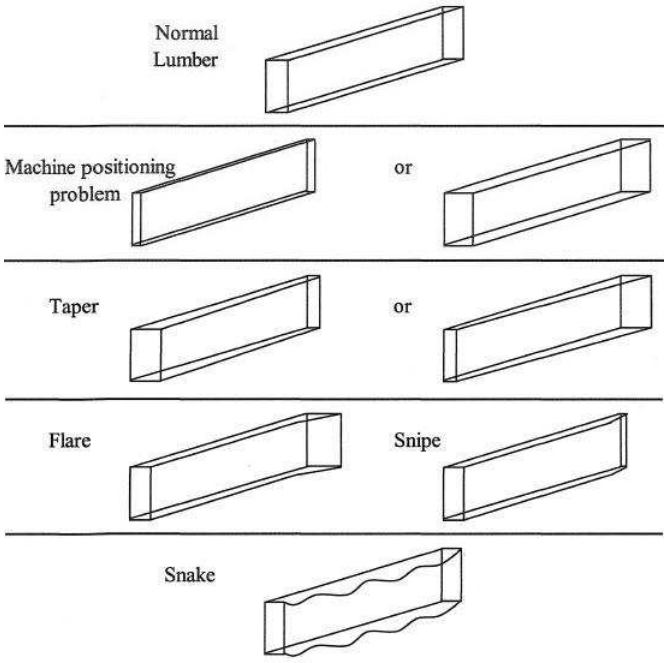

Figure 3. Common defects in lumber manufacturing

0.1016, 0.1524, 0.2032} (expressed in cm) so that the mean deviation from the in-control board thickness μ_0 increased in proportion to the horizontal distance x_i along the length of the board (where the board length $x_n = 244$cm).

- For the flare defect, we took

$$\mathrm{E}[y_{uvi}] = \begin{cases} \mu_0\,, & \text{if } x_i < x_n - 15 \text{ cm}\,, \\ \mu_0 + (x_i - x_{i_0})\delta/(x_n - x_{i_0})\,, & \text{if } x_i \geq x_n - 15 \text{ cm}\,, \end{cases}$$

for $i_0 = \max\{i : x_i < x_n - 15\}$ so that tapering occurs only along the last 15 cm of the board's length.

- For the snake defect, we took $\mathrm{E}[y_{uvi}] = \mu_0 + \mathsf{A}\sin(2\pi x_i/\mathsf{P})$, adding a waveform with the period $\mathsf{P} = 182.88$ cm and with the amplitude $\mathsf{A} \in \{0.0508, 0.1016, 0.1524, 0.2032\}$ (all in cm) for $i = 1, \ldots, n = 2{,}048$.

Table 14 summarizes the ARLs delivered by WDFTC when it is applied to the LRS data for the target false alarm rate FAR $= 0.0027$ alarms/profile (sampled board), which is equivalent to setting the target value $\mathrm{ARL}_0 = 370$. For ease of comparison, the last column

of Table 14 shows the results reported by Staudhammer et al. [50, 51] for their four profile-monitoring charts that are specifically designed to detect sawing defects of type MPP, taper, flare, and snake, respectively. Staudhammer et al. [50, 51] report their results using graphs of the corresponding rates of occurrence for true and false alarms; in Table 14 we convert those rates into the associated values of ARL_0 and ARL_1.

Table 14. ARLs for WDFTC and the profile charts (PCs) of Staudhammer [49]

Shift Type	Shift Size δ or A (cm)	WDFTC $\bar{r} = 6$	PC
MPP	0	358.15	333.33
	0.0254	208.35	20.00
	0.0508	60.60	3.33
	0.0762	30.05	1.25
	0.1016	17.85	1.00
Taper	0	358.15	200.00
	0.0508	190.50	
	0.1016	57.16	
	0.1524	29.07	
	0.2032	17.23	
Flare	0	358.15	50.00
	0.0508	282.76	
	0.1016	109.78	
	0.1524	50.34	
	0.2032	28.57	
Snake	0	358.15	76.92
	0.0508	112.80	
	0.1016	31.84	
	0.1524	16.62	
	0.2032	9.38	

Because $n = 2{,}048$, we considered this problem to exemplify high-dimensional profile monitoring; and therefore we set $q = 0.7$ to obtain more effective dimension reduction when minimizing the WRRE. With this choice of q, we solved (34) for various values of L. Ultimately we decided to set $L = 5$ because that choice resulted in a good data-compression ratio, and further meaningful dimension-reduction was not achieved by using smaller values of L. With $q = 0.7$ and $L = 5$, Equation (34) yielded $p = 92$, achieving a data-compression ratio of approximately 4.5%. WDFTC delivered the average batch size

$\bar{r} = 6$. In Staudhammer et al. [50, 51], various Shewhart-type profile-monitoring charts are tailored respectively to the detection of a specific types of defect; and the development of such highly specialized charts can require an extensive modeling-and-analysis effort. See, for example, the authors' approach to detecting the snake defect. Such modeling efforts are not required to apply WDFTC. It is also noteworthy that WDFTC can detect all the different types of defects without the need for frequent recalibration, although some defects are harder to detect than others (for example, the flare defect). From Table 14 we see the profile chart of Staudhammer et al. [50, 51] that is specifically designed for the MPP defect delivered substantially smaller values of ARL_1 than WDFTC delivered for this particular defect. For other kinds of defects, however, the profile charts of Staudhammer et al. [50, 51] delivered values of ARL_0 that were far below the target value of 370; for example, the values of ARL_0 for the charts designed to detect taper, flare, and snake defects were 200, 50, and 77, respectively. The authors acknowledge the difficulty of adjusting their profile charts to obtain the target ARL_0, because it will require estimating the tails of the run length distribution, which is a challenging problem. Overall, we concluded that WDFTC outperformed the profile-monitoring charts of Staudhammer et al. [50, 51] in this application.

4.4. Conclusion

In this chapter, we described WDFTC, a wavelet-based distribution-free chart for monitoring high-dimensional profiles whose components can have nonnormal distributions, variance heterogeneity, or substantial intercompany correlations. We also formulated the following: (a) an effective dimension-reduction technique based on the discrete wavelet transform and the concept of weighted relative reconstruction error; and (b) a covariance-matrix regularization scheme and a batch-size determination procedure that significantly improve the effectiveness of the associated Hotelling's T^2–type statistics. When tested on normal or nonnormal profiles with dimension $n = 512$ and with independent or correlated components, WDFTC was competitive with other commonly used charts, including the chart M^* of Chicken et al. [9]; moreover, WDFTC substantially outperformed all those charts for small- to medium-size local shifts. We found another advantage of WDFTC is

that its control limits are rapidly evaluated numerically instead of requiring calibration via cumbersome, time-consuming trial-and-error simulations. When WDFTC was applied to lumber-manufacturing profiles of dimension $n = 2{,}048$, we found that WDFTC was sufficiently versatile to detect a wide variety of defect types with reasonable sensitivity while maintaining the user-specified overall rate of generating false alarms. By contrast each of the profile-monitoring charts of Staudhammer et al. [50, 51] was specifically designed to detect a single defect type; and although we found that each such chart often outperformed WDFTC in detecting its relevant defect, it was extremely difficult to calibrate those specialized charts to deliver the target false-alarm rate when those charts are operated separately or jointly. Overall we concluded that WDFTC also outperformed the profile-monitoring charts of Staudhammer et al. [50, 51].

CHAPTER 5

CONTRIBUTIONS AND FUTURE RESEARCH

This thesis proposes new SPC methods for complex observations with characteristics common in modern data, such as nonnormality, cross- and serial-correlations, and heteroscedastic variances. In this chapter, we summarize the main contributions of the thesis and present future research directions.

5.1. Contributions

In Chapter 3, we develop and evaluate DFTC-VE, a distribution-free tabular CUSUM chart for monitoring shifts in the mean of an autocorrelated process, which incorporates automated variance estimation from a training data set. The proposed variance-estimation procedures did not cause dramatic performance degradation in comparison with the corresponding chart performance using the exact values of the relevant process parameters. Also, among the distribution-free charts considered, the DFTC-VE chart incorporating both proposed variance estimators perform reasonably well when it is used in the DFTC-VE chart.

Chapter 4 presents WDFTC, a wavelet-based profile monitoring chart that can monitor shifts in the mean of a high-dimensional profile with noisy components. We develop WDFTC by combining: (a) a new dimension-reduction method that uses the DWT and the WRRE; and (b) a new batch-size determination algorithm that uses a function of component-wise covariances and a sample variance-covariance matrix adjustment by regularization. WDFTC outperforms the existing profile monitoring charts, especially for small-to-medium local shifts in terms of the values of ARL_1 when tested on profiles with normal noise components. Using extreme nonnormal noise components, we also show that WDFTC is robust to the marked departures from the normality of the components of each profile and that WDFTC outperforms the existing profile monitoring charts in terms of the values of ARL_0 and ARL_1.

5.2. Future Research

Several future research topics can stem from the subject areas of this thesis.

1. In Chapter 3, we restrict attention to monitoring shifts in the process mean. While rapid detection of shifts in the mean of a process is of major interest, other characteristics of the monitored process (such as its standard deviation and selected quantiles of its marginal distribution) are also of interest in certain application contexts; and we believe that variants of the DFTC chart could also be adapted for rapidly monitoring shifts in these characteristics.

2. In Chapter 4, we focus on independent profiles that may be correlated component-wise only, but we believe that WDFTC could be extended to handle high-dimensional profiles that are correlated over time. Other possible extensions of WDFTC include:

    - Formulating a more effective and efficient covariance regularization scheme;
    - Formulating a more effective and efficient batching scheme;
    - Designing new wavelet systems that are specially tailored for high-dimensional profile monitoring in different application domains; and
    - Formulating a new dimension-reduction scheme that can be applied rapidly and that more effectively balances the need for an acceptably small reconstruction error against the need for an acceptably small data-compression ratio.

APPENDIX A

DERIVATION OF EQUATION (19)

To provide a justification for Equation (19), first we state and prove the following proposition.

Proposition 1. *If the process* $\{Y_i : i = 1, 2, \ldots\}$ *is covariance stationary, then*

$$\mathrm{E}\big[S^2(m,b)\big] = \frac{b}{b-1}\Big\{\mathrm{Var}\big[\overline{Y}(m)\big] - \mathrm{Var}\big[\overline{Y}(m,b)\big]\Big\}, \tag{38}$$

and

$$\mathrm{Var}\big[\overline{Y}(m,b)\big] = \frac{\mathrm{Var}\big[\overline{Y}(m)\big]}{b}\left[1 + 2\sum_{\ell=1}^{b-1}\big(1-\ell/b\big)\rho_{\overline{Y}(m)}(\ell)\right], \tag{39}$$

where

$$\rho_{\overline{Y}(m)}(\ell) \equiv \mathrm{Corr}\big[\overline{Y}_i(m), \overline{Y}_{i+\ell}(m)\big] \quad \textit{for} \quad \ell = 0, \pm 1, \pm 2, \ldots \tag{40}$$

is the autocorrelation function of the batch means process $\big\{\overline{Y}_i(m) : i = 1, \ldots, b\big\}$.

Proof. From the definition (17) of the nonoverlapping batch means sample variance $S^2(m,b)$,

we have

$$\begin{aligned}
\mathrm{E}\big[S^2(m,b)\big] &= \mathrm{E}\left(\frac{1}{b-1}\sum_{i=1}^{b}\Big\{\big[\overline{Y}_i(m)-\mu\big]-\big[\overline{Y}(m,b)-\mu\big]\Big\}^2\right) && (41)\\
&= \frac{1}{b-1}\sum_{i=1}^{b}\mathrm{E}\Big\{\big[\overline{Y}_i(m)-\mu\big]^2\Big\} \\
&\quad -\frac{2}{b-1}\mathrm{E}\left(\left\{\sum_{i=1}^{b}\big[\overline{Y}_i(m)-\mu\big]\right\}\big[\overline{Y}(m,b)-\mu\big]\right) \\
&\quad +\frac{1}{b-1}\sum_{i=1}^{b}\mathrm{E}\Big\{\big[\overline{Y}(m,b)-\mu\big]^2\Big\} && (42)\\
&= \frac{b}{b-1}\mathrm{Var}\big[\overline{Y}(m)\big] \\
&\quad -\frac{2}{b-1}\mathrm{E}\big\{b\big[\overline{Y}(m,b)-\mu\big]\big[\overline{Y}(m,b)-\mu\big]\big\} \\
&\quad +\frac{b}{b-1}\mathrm{Var}\big[\overline{Y}(m,b)\big] && (43)\\
&= \frac{b}{b-1}\Big\{\mathrm{Var}\big[\overline{Y}(m)\big]-\mathrm{Var}\big[\overline{Y}(m,b)\big]\Big\}. && (44)
\end{aligned}$$

Equation (39) is a well-known result for unbatched observations from a covariance stationary process—see, for example, Equation (9.8) of Law and Kelton (2000); and the proof for batch means proceeds along the same lines as for unbatched observations. ∎

To compute an approximately unbiased estimator of $\mathrm{Var}\big[\overline{Y}(m)\big]$, we exploit the special structure of the autocorrelation function (40) of the batch means based on the AR(1) model (14) of the batch means,

$$\rho_{\overline{Y}(m)}(\ell) = \varphi_{\overline{Y}(m)}^{|\ell|} \quad \text{for } \ell = 0, \pm 1, \pm 2, \ldots . \tag{45}$$

Now it is easily proved using, for example, the argument justifying equation (A.3) of Sargent, Kang, and Goldsman (1992) that

$$C(m,b) \equiv 1 + 2\sum_{\ell=1}^{b-1}\big(1-\ell/b\big)\varphi_{\overline{Y}(m)}^{\ell} = \frac{1+\varphi_{\overline{Y}(m)}}{1-\varphi_{\overline{Y}(m)}} - \frac{2\varphi_{\overline{Y}(m)}\big[1-\varphi_{\overline{Y}(m)}^{b}\big]}{b\big[1-\varphi_{\overline{Y}(m)}\big]^2}. \tag{46}$$

Combining the right-hand side of (46) with (38) and (39), we obtain

$$\mathrm{E}\big[S^2(m,b)\big] = \frac{b}{b-1}\mathrm{Var}[\,\overline{Y}_i(m)\,]\left[1 - \frac{C(m,b)}{b}\right]; \tag{47}$$

and thus our approximately unbiased estimator of $\mathrm{Var}[\,\overline{Y}_i(m)\,]$ is given by

$$\widehat{\mathrm{Var}}[\,\overline{Y}_i(m)\,] = S^2(m,b)\left[\frac{b-1}{b-\widehat{C}(m,b)}\right], \tag{48}$$

where the estimated correlation correction factor $\widehat{C}(m,b)$ is given by

$$\widehat{C}(m,b) = \frac{1+\widehat{\varphi}_{\overline{X}(m)}}{1-\widehat{\varphi}_{\overline{X}(m)}} - \frac{2\widehat{\varphi}_{\overline{X}(m)}\left[1-\widehat{\varphi}^{\,b}_{\overline{X}(m)}\right]}{b\left[1-\widehat{\varphi}_{\overline{X}(m)}\right]^2}. \qquad \blacksquare \tag{49}$$

APPENDIX B

STANDARD ERRORS FOR ARLS REPORTED IN TABLES 3–5 IN CHAPTER 3

In Tables 15–17 below, we present the standard errors for each estimated ARL respectively reported in Tables 3–5. Each estimated ARL and its associated estimated standard error are based on 4,000 independent replications.

Table 15. Standard errors of two-sided ARLs in Table 3 for an AR(1) process

φ_Y	Shift	R&W	J&B		DFTC-VE		DFTC
	η		Area	QDAR	Area	QDAR	Phase II
		$\overline{m}_1 = 4$	$\overline{m}_1 = 94$	$\overline{m}_1 = 1$	$\overline{m}_1 = 93$	$\overline{m}_1 = 1$	
		$\overline{m}_2 = 4$	$\overline{m}_2 = 1$	$\overline{m}_2 = 1$	$\overline{m}_2 = 1$	$\overline{m}_2 = 1$	$\overline{m}_2 = 1$
0.25	0	166.345	83.952	91.387	208.833	159.847	150.249
	0.25	72.184	2.305	2.267	2.151	2.125	1.742
	0.5	19.130	0.832	0.796	0.536	0.519	0.463
	0.75	5.804	0.460	0.433	0.270	0.255	0.230
	1	2.208	0.306	0.275	0.166	0.156	0.138
	1.5	0.414	0.173	0.152	0.089	0.083	0.074
	2	0.127	0.117	0.099	0.058	0.052	0.047
	2.5	0.048	0.087	0.071	0.042	0.038	0.032
	3	0.019	0.068	0.055	0.033	0.028	0.024
	4	0.002	0.047	0.035	0.022	0.018	0.016
		$\overline{m}_1 = 8$	$\overline{m}_1 = 101$	$\overline{m}_1 = 2$	$\overline{m}_1 = 104$	$\overline{m}_1 = 2$	
		$\overline{m}_2 = 8$	$\overline{m}_2 = 1$	$\overline{m}_2 = 2$	$\overline{m}_2 = 1$	$\overline{m}_2 = 2$	$\overline{m}_2 = 1$
0.5	0	175.413	85.192	92.292	183.336	205.411	158.699
	0.25	71.062	3.676	3.624	3.577	3.228	3.007
	0.5	20.115	1.286	1.217	0.891	0.945	0.788
	0.75	6.228	0.703	0.681	0.449	0.484	0.394
	1	2.251	0.470	0.445	0.289	0.302	0.244
	1.5	0.467	0.261	0.241	0.152	0.166	0.125
	2	0.140	0.171	0.160	0.098	0.111	0.080
	2.5	0.054	0.129	0.112	0.072	0.079	0.057
	3	0.021	0.101	0.089	0.055	0.060	0.042
	4	0.000	0.068	0.058	0.037	0.041	0.026
		$\overline{m}_1 = 19$	$\overline{m}_1 = 188$	$\overline{m}_1 = 8$	$\overline{m}_1 = 184$	$\overline{m}_1 = 7$	
		$\overline{m}_2 = 19$	$\overline{m}_2 = 1$	$\overline{m}_2 = 8$	$\overline{m}_2 = 1$	$\overline{m}_2 = 7$	$\overline{m}_2 = 3$
0.7	0	181.855	84.283	97.556	200.754	141.242	159.422
	0.25	73.120	5.895	5.629	6.266	5.527	4.903
	0.5	18.365	2.115	2.052	1.635	1.763	1.361
	0.75	6.101	1.181	1.122	0.836	0.909	0.683
	1	2.307	0.760	0.737	0.524	0.585	0.438
	1.5	0.538	0.443	0.401	0.278	0.307	0.227
	2	0.184	0.295	0.260	0.180	0.202	0.142
	2.5	0.059	0.218	0.193	0.131	0.147	0.100
	3	0.012	0.172	0.150	0.098	0.113	0.077
	4	0.000	0.121	0.103	0.069	0.079	0.049
		$\overline{m}_1 = 58$	$\overline{m}_1 = 487$	$\overline{m}_1 = 30$	$\overline{m}_1 = 485$	$\overline{m}_1 = 30$	
		$\overline{m}_2 = 58$	$\overline{m}_2 = 1$	$\overline{m}_2 = 30$	$\overline{m}_2 = 1$	$\overline{m}_2 = 30$	$\overline{m}_2 = 7$
0.9	0	221.147	91.991	102.068	232.267	112.648	145.525
	0.25	100.262	16.173	14.682	21.271	15.973	12.922
	0.5	30.853	5.504	5.084	5.123	4.864	3.913
	0.75	11.900	3.102	2.815	2.568	2.644	2.010
	1	4.851	2.003	1.863	1.651	1.692	1.250
	1.5	1.290	1.167	1.009	0.870	0.936	0.655
	2	0.442	0.776	0.675	0.583	0.600	0.407
	2.5	0.166	0.592	0.500	0.431	0.448	0.278
	3	0.041	0.465	0.400	0.337	0.356	0.216
	4	0.000	0.324	0.293	0.237	0.270	0.135

Table 16. Standard errors of two-sided ARLs in Table 4 for an EAR(1) process

φ_Y	Shift	R&W	J&B		DFTC-VE		DFTC
	η		Area	QDAR	Area	QDAR	Phase II
		$\overline{m}_1 = 4$	$\overline{m}_1 = 92$	$\overline{m}_1 = 1$	$\overline{m}_1 = 90$	$\overline{m}_1 = 1$	
		$\overline{m}_2 = 4$	$\overline{m}_2 = 1$	$\overline{m}_2 = 1$	$\overline{m}_2 = 1$	$\overline{m}_2 = 1$	$\overline{m}_2 = 1$
0.25	0	13.714	83.965	83.773	201.646	156.113	163.748
	0.25	6.928	2.302	2.328	2.209	2.180	2.125
	0.5	3.884	0.835	0.797	0.555	0.524	0.534
	0.75	2.282	0.470	0.432	0.269	0.260	0.257
	1	1.239	0.313	0.285	0.173	0.159	0.159
	1.5	0.416	0.176	0.153	0.089	0.083	0.080
	2	0.152	0.118	0.101	0.058	0.052	0.051
	2.5	0.054	0.087	0.072	0.041	0.036	0.036
	3	0.004	0.068	0.054	0.032	0.028	0.027
	4	0.000	0.047	0.036	0.022	0.018	0.018
		$\overline{m}_1 = 8$	$\overline{m}_1 = 98$	$\overline{m}_1 = 2$	$\overline{m}_1 = 98$	$\overline{m}_1 = 2$	
		$\overline{m}_2 = 8$	$\overline{m}_2 = 1$	$\overline{m}_2 = 2$	$\overline{m}_2 = 1$	$\overline{m}_2 = 2$	$\overline{m}_2 = 1$
0.5	0	23.643	86.682	92.125	190.435	196.495	173.514
	0.25	11.675	3.588	3.746	3.802	3.353	3.422
	0.5	5.902	1.291	1.257	0.918	0.992	0.910
	0.75	3.217	0.718	0.695	0.468	0.524	0.442
	1	1.794	0.470	0.451	0.292	0.338	0.270
	1.5	0.531	0.268	0.244	0.149	0.188	0.138
	2	0.172	0.178	0.161	0.098	0.126	0.085
	2.5	0.036	0.131	0.118	0.071	0.096	0.059
	3	0.000	0.104	0.090	0.054	0.077	0.045
	4	0.000	0.070	0.059	0.036	0.057	0.028
		$\overline{m}_1 = 19$	$\overline{m}_1 = 201$	$\overline{m}_1 = 8$	$\overline{m}_1 = 201$	$\overline{m}_1 = 7$	
		$\overline{m}_2 = 19$	$\overline{m}_2 = 1$	$\overline{m}_2 = 8$	$\overline{m}_2 = 1$	$\overline{m}_2 = 7$	$\overline{m}_2 = 3$
0.7	0	43.014	88.042	89.538	223.396	142.122	172.802
	0.25	20.269	6.168	6.240	7.032	5.758	5.776
	0.5	9.062	2.220	2.134	1.730	1.795	1.610
	0.75	4.518	1.205	1.155	0.854	0.932	0.795
	1	2.193	0.784	0.763	0.525	0.593	0.487
	1.5	0.638	0.455	0.407	0.285	0.317	0.246
	2	0.151	0.302	0.263	0.185	0.205	0.156
	2.5	0.007	0.226	0.195	0.137	0.150	0.109
	3	0.000	0.179	0.152	0.106	0.116	0.083
	4	0.000	0.123	0.105	0.072	0.081	0.052
		$\overline{m}_1 = 58$	$\overline{m}_1 = 496$	$\overline{m}_1 = 29$	$\overline{m}_1 = 496$	$\overline{m}_1 = 30$	
		$\overline{m}_2 = 58$	$\overline{m}_2 = 1$	$\overline{m}_2 = 29$	$\overline{m}_2 = 1$	$\overline{m}_2 = 30$	$\overline{m}_2 = 7$
0.9	0	82.836	91.776	103.641	269.238	112.095	172.616
	0.25	38.673	17.403	16.910	23.792	17.244	15.615
	0.5	17.962	5.874	5.525	5.516	5.042	4.494
	0.75	9.641	3.181	2.974	2.692	2.775	2.252
	1	5.046	2.112	1.911	1.633	1.731	1.364
	1.5	1.549	1.198	1.049	0.897	0.935	0.685
	2	0.325	0.812	0.693	0.588	0.638	0.435
	2.5	0.000	0.612	0.515	0.432	0.479	0.301
	3	0.000	0.487	0.406	0.340	0.379	0.227
	4	0.000	0.340	0.296	0.237	0.281	0.147

Table 17. Standard errors of two-sided ARLs in Table 5 for an $M/M/1$ process

τ	Shift	R&W	J&B		DFTC-VE		DFTC
	η		Area	QDAR	Area	QDAR	Phase II
		$\overline{m}_1 = 11$	$\overline{m}_1 = 500$	$\overline{m}_1 = 4$	$\overline{m}_1 = 499$	$\overline{m}_1 = 4$	
		$\overline{m}_2 = 11$	$\overline{m}_2 = 1$	$\overline{m}_2 = 4$	$\overline{m}_2 = 1$	$\overline{m}_2 = 4$	$\overline{m}_2 = 2$
0.3	0	7.335	99.588	97.114	930.824	221.766	130.784
	0.25	4.809	4.973	4.713	5.358	4.353	3.904
	0.5	3.473	1.866	1.683	1.290	1.313	0.959
	0.75	2.079	1.082	0.923	0.655	0.706	0.451
	1	1.338	0.752	0.629	0.436	0.473	0.274
	1.5	0.422	0.457	0.370	0.250	0.282	0.143
	2	0.068	0.327	0.256	0.175	0.193	0.090
	2.5	0.003	0.250	0.187	0.132	0.149	0.063
	3	0.000	0.204	0.152	0.106	0.123	0.046
	4	0.000	0.150	0.110	0.076	0.090	0.030
		$\overline{m}_1 = 55$	$\overline{m}_1 = 499$	$\overline{m}_1 = 22$	$\overline{m}_1 = 499$	$\overline{m}_1 = 24$	
		$\overline{m}_2 = 55$	$\overline{m}_2 = 1$	$\overline{m}_2 = 22$	$\overline{m}_2 = 1$	$\overline{m}_2 = 24$	$\overline{m}_2 = 10$
0.6	0	17.607	101.011	109.678	804.237	142.475	182.401
	0.25	13.644	16.631	14.115	24.984	14.420	13.302
	0.5	11.618	5.750	4.933	5.213	4.726	3.471
	0.75	9.799	3.221	2.723	2.501	2.605	1.660
	1	8.262	2.169	1.806	1.617	1.728	1.023
	1.5	5.932	1.268	1.038	0.930	1.035	0.517
	2	3.572	0.884	0.732	0.630	0.732	0.320
	2.5	2.117	0.683	0.554	0.482	0.566	0.227
	3	1.224	0.557	0.456	0.387	0.477	0.174
	4	0.191	0.396	0.348	0.276	0.365	0.114

APPENDIX C

STANDARD ERRORS FOR ARLS REPORTED IN TABLES 9–14 IN CHAPTER 4

In Tables 18–23 below, we present the standard errors (SEs) for each estimated ARL respectively reported in Tables 9–14. Each estimated ARL and its associated estimated standard error are based on 1,000 independent replications.

Table 18. Standard errors in Table 9 for independent standard normal components

Shift Type	Shift Size	WDFTC $\bar{r}=3$	HTW_n	HTW_p	M^*
In-Control	0	4.777	6.824	6.295	7.108
Global Shift 1	0.25	0.042	0.480	0.051	0.068
	0.5	<0.001	0.015	<0.001	0.007
	0.75	<0.001	0.001	<0.001	0.001
	1	<0.001	0.001	<0.001	<0.001
	2	<0.001	0.001	<0.001	<0.001
Global Shift 2	0.25	0.043	0.505	0.057	0.016
	0.5	<0.001	0.012	<0.001	0.007
	0.75	<0.001	0.001	<0.001	<0.001
	1	<0.001	0.001	<0.001	<0.001
	2	<0.001	0.001	<0.001	<0.001
Local Shift 1	0.25	2.758	5.805	5.115	6.638
	0.5	0.530	4.527	3.035	5.232
	0.75	0.179	3.265	1.386	3.023
	1	0.082	2.124	0.577	1.463
	2	0.013	0.180	0.024	0.080
Local Shift 2	0.25	2.561	6.070	5.164	7.031
	0.5	0.516	4.269	2.532	4.508
	0.75	0.155	2.575	0.942	2.539
	1	0.077	1.514	0.335	0.970
	2	0.010	0.082	0.014	0.059

Table 19. Standard errors in Table 10 for correlated standard normal components

Shift Type	Shift Size	WDFTC $\bar{r}=3$	HTW_n	HTW_p	M^*
In-Control	0	4.774	6.821	6.278	1.541
Global Shift 1	0.25	4.310	6.371	6.274	1.449
	0.5	3.164	5.932	5.580	0.956
	0.75	1.678	5.543	4.561	0.576
	1	0.804	5.398	3.452	0.347
	2	0.123	2.929	0.900	0.055
Global Shift 2	0.25	0.003	0.104	0.009	0.021
	0.5	0.003	<0.001	<0.001	0.015
	0.75	0.003	<0.001	<0.001	0.010
	1	0.003	<0.001	<0.001	0.006
	2	0.003	<0.001	<0.001	0.005
Local Shift 1	0.25	1.460	5.702	4.524	1.588
	0.5	0.210	3.480	1.617	1.577
	0.75	0.077	1.804	0.502	1.576
	1	0.044	0.827	0.137	1.492
	2	0.003	0.028	0.002	1.345
Local Shift 2	0.25	1.341	5.206	3.905	1.618
	0.5	0.187	2.962	1.517	1.603
	0.75	0.069	1.306	0.412	1.451
	1	0.044	0.504	0.118	1.483
	2	0.003	0.011	0.001	1.295

Table 20. Standard errors in Table 11 for correlated general normal components

Shift Type	Shift Size	WDFTC $\bar{r}=8$	HTW_n	HTW_p	M^*
In-Control	0	4.788	6.821	6.563	0.543
Global Shift 1	0.25	0.043	0.078	0.006	0.552
	0.5	0.043	<0.001	<0.001	0.510
	0.75	0.043	<0.001	<0.001	0.374
	1	0.043	<0.001	<0.001	0.260
	2	0.043	<0.001	<0.001	0.084
Global Shift 2	0.25	0.043	0.077	0.005	0.015
	0.5	0.043	<0.001	<0.001	0.006
	0.75	0.043	<0.001	<0.001	0.010
	1	0.043	<0.001	<0.001	0.004
	2	0.043	<0.001	<0.001	<0.001
Local Shift 1	0.25	0.979	5.478	4.441	0.557
	0.5	0.187	3.724	1.931	0.541
	0.75	0.076	1.683	0.673	0.552
	1	0.043	0.785	0.235	0.555
	2	0.043	0.027	0.005	0.536
Local Shift 2	0.25	0.570	4.761	3.629	0.559
	0.5	0.129	2.171	1.252	0.524
	0.75	0.043	0.852	0.322	0.542
	1	0.043	0.266	0.071	0.570
	2	0.043	0.005	0.001	0.532

Table 21. Standard errors in Table 12 for independent standard exponential components

Shift Type	Shift Size	WDFTC $\bar{r}=3$	HTW_n CM_A	HTW_p CM_A	M^* CM_B
In-Control	0	4.800	–	–	0.733
Global Shift 1	0.25	0.047	–	–	0.343
	0.5	<0.001	–	–	0.092
	0.75	<0.001	–	–	0.037
	1	<0.001	–	–	0.019
	2	<0.001	–	–	<0.001
Global Shift 2	0.25	0.047	–	–	0.014
	0.5	<0.001	–	–	0.010
	0.75	<0.001	–	–	0.013
	1	<0.001	–	–	0.002
	2	<0.001	–	–	<0.001
Local Shift 1	0.25	2.982	–	–	0.726
	0.5	0.645	–	–	0.735
	0.75	0.214	–	–	0.727
	1	0.098	–	–	0.681
	2	0.022	–	–	0.485
Local Shift 2	0.25	2.895	–	–	0.724
	0.5	0.537	–	–	0.743
	0.75	0.187	–	–	0.738
	1	0.089	–	–	0.712
	2	0.015	–	–	0.413

Table 22. Standard errors in Table 13 for correlated standard exponential components

Shift Type	Shift Size	WDFTC $\bar{r}=3$	HTW_n CM_A	HTW_p CM_A	M^* CM_B
In-Control	0	4.972	–	–	4.807
Global Shift 1	0.25	4.292	–	–	3.698
	0.5	4.067	–	–	2.201
	0.75	3.510	–	–	1.085
	1	2.891	–	–	0.598
	2	0.615	–	–	0.127
Global Shift 2	0.25	0.031	–	–	0.087
	0.5	0.005	–	–	0.055
	0.75	0.005	–	–	0.045
	1	0.005	–	–	0.036
	2	0.005	–	–	0.022
Local Shift 1	0.25	4.646	–	–	4.605
	0.5	1.207	–	–	4.571
	0.75	0.377	–	–	4.454
	1	0.269	–	–	4.371
	2	0.059	–	–	3.849
Local Shift 2	0.25	4.070	–	–	4.503
	0.5	1.087	–	–	4.590
	0.75	0.382	–	–	4.488
	1	0.137	–	–	4.508
	2	0.048	–	–	3.818

Table 23. Standard errors in Table 14 of WDFTC and profile charts (PC) of Staudhammer [49]

Shift Type	Shift Size	WDFTC $\bar{r} = 5$	PC
MPP	0	8.772	–
	0.01	5.166	–
	0.02	1.575	–
	0.03	1.436	–
	0.04	0.785	–
Taper	0	8.772	–
	0.02	4.860	–
	0.04	1.657	–
	0.06	1.086	–
	0.08	0.625	–
Flare	0	8.772	–
	0.02	6.988	–
	0.04	4.435	–
	0.06	1.952	–
	0.08	1.251	–
Snake	0	8.772	–
	0.02	3.653	–
	0.04	1.384	–
	0.06	0.970	–
	0.08	0.310	–

REFERENCES

[1] ALEXOPOULOS, C., ARGON, N. T., GOLDSMAN, D., STEIGER, N. M., TOKOL, G., and WILSON, J. R., "Efficient computation of overlapping variance estimators for simulation," *INFORMS Journal on Computing*, vol. 19, no. 3, pp. 314–327, 2007.

[2] ALEXOPOULOS, C., ARGON, N. T., GOLDSMAN, D., TOKOL, G., and WILSON, J. R., "Overlapping variance estimators for simulation," *Operations Research*, vol. 55, no. 6, pp. 1090–1103, 2007.

[3] ARADHYE, H. B., BAKSHI, B. R., STRAUSS, R. A., and DAVIS, J. F., "Multiscale statistical process control using wavelets–theoretical analysis and properties," *AIChE Journal*, vol. 49, no. 4, pp. 939–958, 2003.

[4] BAGSHAW, M. and JOHNSON, R. A., "The effect of serial correlation on the performance of cusum tests ii," *Technometrics*, vol. 17, no. 1, pp. 73–80, 1975.

[5] BICKEL, P. J. and LEVINA, E., "Covariance regularization by thresholding," *Annals of Statistics*, vol. 36, no. 6, pp. 2577–2604, 2008.

[6] BILLARD, L. and MOHAMED, F. Y., "Estimation of the parameters of an ear(p) process," *Journal of Time Series Analysis*, vol. 12, no. 3, pp. 179–192, 1991.

[7] BROWNING, R., "Introduction," *Pippa Passes*, lines 13–20, 1841.

[8] CARIO, M. C. and NELSON, B. L., "Modeling and generating random vectors from arbitrary marginal distribution and correlation matrix," tech. rep., Northwestern University, IEMS, Evanston, Illinois, 1997.

[9] CHICKEN, E., PIGNATIELLO, J. J. J., and SIMPSON, J., "Statistical process monitoring of nonlinear profiles using wavelets," *Journal of Quality Technology*, vol. 41, pp. 198–212, 2009.

[10] DANIELS, M. J. and KASS, R. E., "Shrinkage estimators for covariance matrices," *Biometrics*, vol. 57, no. 4, pp. 1173–1184, 2001.

[11] DING, Y., ZENG, L., and ZHOU, S., "Phase I analysis for monitoring nonlinear profiles in manufacturing processes," *Journal of Quality Technology*, vol. 38, no. 3, pp. 199–216, 2006.

[12] DONOHO, D. L. and JOHNSTONE, J. M., "Ideal spatial adaptation by wavelet shrinkage," *Biometrica*, vol. 81, no. 3, pp. 425–455, 1994.

[13] FAN, J., "Test of significance based on wavelet thresholding and neyman's truncation," *Journal of American Statistical Association*, vol. 91, no. 434, pp. 674–688, 1996.

[14] GANESAN, R., DAS, T. K., and VENKATARAMAN, V., "Wavelet-based multiscale statistical process monitoring: a literature review," *IIE Transactions*, vol. 36, no. 9, pp. 787–806, 2004.

[15] GAO, H. Y., "Wavelet shrinkage estimates for heteroscedastic regression models," tech. rep., MathSoft Inc., Seattle, Washington, 1997.

[16] HOFFBECK, J. P. and LANDGREBE, D. A., "Covariance matrix estimation and classification with limited training data," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 18, no. 7, pp. 763–767, 1996.

[17] JENSEN, W. A., JONES-FARMER, L. A., CHAMP, C. W., and WOODALL, W. H., "Effects of parameter estimation on control chart properties: A literature review," *Journal of Quality Technology*, vol. 38, no. 4, pp. 349–364, 2006.

[18] JEONG, M. K., LU, J.-C., and WANG, N., "Wavelet-based spc procedure for complicated functional data," *International Journal of Production Research*, vol. 44, no. 4, pp. 729–744, 2006.

[19] JIN, J. and SHI, J., "Feature-preserving data compression of stamping tonnage information using wavelets," *Technometrics*, vol. 41, no. 4, pp. 327–339, 1999.

[20] JIN, J. and SHI, J., "Automatic feature extraction of waveeform signals for in-process diagnostic performance improvement," *Journal of Intelligent Manufacturing*, vol. 12, pp. 257–268, 2001.

[21] JOHNSON, R. A. and BAGSHAW, M., "The effect of serial correlation on the performance of cusum tests," *Technometrics*, vol. 16, no. 1, pp. 103–112, 1974.

[22] JOLLIFFE, I. T., *Principal Component Analysis*. New York, New York: Springer, 1986.

[23] KANG, L. and ALBIN, S. L., "On-line monitoring when the process yields a linear profile," *Journal of Quality Technology*, vol. 32, no. 4, pp. 418–426, 2000.

[24] KANO, M., NAGAO, K., HASEBE, S., HASHIMOTO, I., OHNO, H., STRAUSS, R., and BAKSHI, B., "Comparison of statistical process monitoring methods: application to the eastman challenge problem," *Computers and Chemical Engineering*, vol. 24, no. 2-7, pp. 175–181, 2002.

[25] KIM, K., MAHMOUD, J. A., and WOODALL, W. H., "On the monitoring of linear profiles," *Journal of Quality Technology*, vol. 35, pp. 317–328, 2003.

[26] KIM, S.-H., ALEXOPOULOS, C., TSUI, K.-L., and WILSON, J. R., "A distribution-free tabular cusum chart for autocorrelated data," *IIE Transactions*, vol. 39, pp. 317–330, 2007.

[27] KIM, S.-H. and WILSON, J. R., "A discussion on 'detection of intrusions in information systems by sequential change-point methods' by tartakovsky et al.," *Statistical Methodology*, vol. 3, no. 3, pp. 315–319, 2006.

[28] LADA, E. K., LU, J.-C., and WILSON, J. R., "A wavelet based procedure for process fault detection," *IEEE Transactions on Semiconductor Manufacturing*, vol. 15, no. 1, pp. 79–90, 2002.

[29] LADA, E. K., STEIGER, N. M., and WILSON, J. R., "Sbatch: A spaced batch means procedure for steady-state simulation analysis," *Journal of Simulation*, vol. 2, no. 3, pp. 170–185, 2008.

[30] LADA, E. K. and WILSON, J. R., "A wavelet-based spectral procedure for steady-state simulation analysis," *European Journal of Operational Research*, vol. 174, no. 3, pp. 1769–1801, 2006.

[31] LEDOIT, O. and WOLF, M., "Some hypothesis tests for the covariance matrix when the dimension is large compared to the sample size," *The Annals of Statistics*, vol. 30, no. 4, pp. 1081–1102, 2002.

[32] LEE, J., ALEXOPOULOS, C., GOLDSMAN, D., KIM, S.-H., TSUI, K.-L., and WILSON, J. R., "Monitoring autocorrelated processes using a distribution-free tabular cusum chart with automated variance estimation," *IIE Transactions*, vol. 41, no. 11, pp. 979–994, 2009.

[33] LU, C.-W. and REYNOLDS, M. R. J., "Control charts for monitoring the mean and variance of autocorrelated processes," *Journal of Quality Technology*, vol. 31, no. 3, pp. 259–274, 1999.

[34] LU, C.-W. and REYNOLDS, M. R. J., "Cusum charts for monitoring an autocorrelated process," *Journal of Quality Technology*, vol. 33, pp. 316–334, 2001.

[35] MALLAT, S. G., *A Wavelet Tour of Signal Processing: The Sparse Way*. San Diego, California: Academic Press, 3rd ed., 2009.

[36] MARAGAH, H. D. and WOODALL, W. H., "The effect of autocorrelation on the retrospective x-chart," *Journal of Statistical Computation and Simulation*, vol. 40, pp. 29–40, 1992.

[37] MONTGOMERY, D. C., *Introduction to Statistical Quality Control*. New York, New York: John Wiley & Sons, 4th ed., 2001.

[38] NOVA, A. M. O. P. and WILSON, J. R., "Estimation of multiresponse simulation metamodels using control variates," *Management Science*, vol. 35, no. 11, pp. 1316–1333, 1989.

[39] OGDEN, T. R., *Essential Wavelets for Statistical Application and Data Analysis*. Boston, Massachusetts: Birkhauser Boston, 1996.

[40] QIU, P., "Distribution-free multivariate process control based on log-linear modeling," *IIE Transactions*, vol. 40, no. 7, pp. 664–677, 2008.

[41] RAMSAY, J. O. and SILVERMAN, B. W., *Functional Data Analysis*. New York, New York: Springer, 1997.

[42] RASMUSSEN, H. K., KOZAK, R. A., and MANESS, T. C., "An analysis of machine caused lumber shape defects in british columbia sawmills," *Forest Products Journal*, vol. 54, no. 6, pp. 47–56, 2004.

[43] ROWLANDS, R. J. and WETHERILL, G. B., *Quality control, in Handbook of Sequential Analysis, B. K. Ghosh and P. K. Sen*. New York, New York: Marcel Dekker, 1991.

[44] RUNGER, G. C. and WILLEMAIN, T. R., "Model-based and model-free control of autocorrelated processes," *Journal of Quality Technology*, vol. 27, no. 4, pp. 283–292, 1995.

[45] SACHS, R. V. and MACGIBBON, B., "Non-parametric curve estimation by wavelet thresholding with locally stationary errors," *Scandinavian Journal of Statistics*, vol. 27, no. 3, pp. 475–499, 2000.

[46] SHAPIRO, S. S. and WILK, M. B., "An analysis of variance test for normality (complete samples)," *Biometrika*, vol. 52, no. 3/4, pp. 591–611, 1965.

[47] SIEGMUND, D., *Sequential Analysis: Tests and Confidence Intervals*. New York, New York: Springer-Verlag, 1985.

[48] STANFIELD, P. M., WILSON, J. R., and KING, R. E., "Flexible modeling of correlated operation times with application in product-resue facilities," *International Journal of Production Research*, vol. 42, pp. 2179–2196, 2004.

[49] STAUDHAMMER, C., *Statistical procedure for development of real-time statistical process control (SPC) in lumber manufacturing (Ph.D thesis)*. Vancouver, British Columbia, Canada: University of British Columbia, 2004.

[50] STAUDHAMMER, C., KOZAK, R. A., and MANESS, T. C., "Spc methods for detecting simple sawing defects using real-time laser range sensor data," *Wood and Fiber Science*, vol. 38, no. 4, pp. 696–716, 2006.

[51] STAUDHAMMER, C., MANESS, T. C., and KOZAK, R. A., "Profile charts for monitoring lumber manufacturing using laser range sensor data," *Journal of Quality Technology*, vol. 39, no. 3, pp. 224–240, 2007.

[52] STEIGER, N. M., LADA, E. K., WILSON, J. R., JOINES, J. A., ALEXOPOULOS, C., and GOLDSMAN, D., "Asap3: A batch means procedure for steady-state simulation analysis," *ACM Transactions on Modeling and Computer Simulation*, vol. 15, no. 1, pp. 39–73, 2005.

[53] STEIGER, N. M. and WILSON, J. R., "Convergence properties of the batch means method for simulation output analysis," *INFORMS Journal on Computing*, vol. 13, no. 4, pp. 277–293, 2001.

[54] VON NEUMANN, J., "Distribution of the ratio of the mean square successive difference to the variance," *The Annals of Mathematical Statistics*, vol. 12, no. 4, pp. 367–395, 1941.

[55] WILLIAMS, J. D., WOODALL, W. H., and BIRCH, J. B., "Statistical monitoring of nonlinear product and process quality profiles," *Quality and Reliability Engineering International*, vol. 23, pp. 925–941, 2007.

[56] WOODALL, W. H., SPITZNER, D. J., MONTGOMERY, D. C., and GUPTA, S., "Using control charts to monitor process and product quality profiles," *Journal of Quality Technology*, vol. 36, pp. 309–320, 2004.

VITA

Joongsup (Jay) Lee was born on March 7, 1980, in Jeonju, South Korea. After serving in the United States Army for three years, he received a B.S. in Mathematical Decision Sciences from the University of North Carolina at Chapel Hill in 2005. He received an M.S. in Operations Research from the H. Milton Stewart School of Industrial and Systems Engineering at the Georgia Institute of Technology in 2008. He currently works as a corporate security analyst at United Parcel Service.