NAVAL POSTGRADUATE SCHOOL

MONTEREY, CALIFORNIA

DISSERTATION

GRAPH-THEORETIC STATISTICAL METHODS FOR DETECTING AND LOCALIZING DISTRIBUTIONAL CHANGE IN MULTIVARIATE DATA

by

Matthew A. Hawks

June 2015

Dissertation Supervisor Robert A. Koyak

Approved for public release; distribution is unlimited

THIS PAGE INTENTIONALLY LEFT BLANK

REPORT DOCUMENTATION PAGE

Form Approved OMB No. 0704–0188

1. AGENCY USE ONLY ***(Leave blank)***	**2. REPORT DATE** June 2015	**3. REPORT TYPE AND DATES COVERED** Doctoral Dissertation

4. TITLE AND SUBTITLE
GRAPH-THEORETIC STATISTICAL METHODS FOR DETECTING AND LOCALIZING DISTRIBUTIONAL CHANGE IN MULTIVARIATE DATA

5. FUNDING NUMBERS

6. AUTHOR(S) Hawks, Matthew A.

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
Naval Postgraduate School
Monterey, CA 93943-5000

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING /MONITORING AGENCY NAME(S) AND ADDRESS(ES)
N/A

10. SPONSORING/MONITORING AGENCY REPORT NUMBER

11. SUPPLEMENTARY NOTES The views expressed in this dissertation are those of the author and do not reflect the official policy or position of the Department of Defense or the U.S. Government. IRB Protocol number ____N/A____.

12a. DISTRIBUTION / AVAILABILITY STATEMENT
Approved for public release; distribution is unlimited

12b. DISTRIBUTION CODE

13. ABSTRACT (maximum 200 words)

This dissertation explores the topic of detecting and localizing change in a series of multivariate data using graph-theoretic statistical criteria. Change-detection methods based on graph theory are emerging due to their ability to detect change of a general nature with desirable power properties. The graph-theoretic structures of minimum non-bipartite matching and nearest neighbors according to distances between observations form the basis of our statistical procedures. We consider the computation time to implement the procedures with the detection power of the derived statistics. In a simulation study, we evaluate the power of our proposed statistical tests in a series of vignettes in which the sampling distribution, dimensionality, change parameter (location or scale), change type (abrupt or gradual), and change magnitude each are allowed to vary. We compare detection power with contemporary parametric and graph-theoretic approaches. Although our tests alone do not provide the information needed to localize a change point, we develop a follow-on procedure that satisfies this objective. We illustrate our proposed statistical tests and change-point localization techniques in an application, which demonstrates how several of the apparent limitations of our approach can be surmounted.

14. SUBJECT TERMS Nonparametric test, non-bipartite matching, nearest neighbor, change point			**15. NUMBER OF PAGES** 147
			16. PRICE CODE
17. SECURITY CLASSIFICATION OF REPORT Unclassified	**18. SECURITY CLASSIFICATION OF THIS PAGE** Unclassified	**19. SECURITY CLASSIFICATION OF ABSTRACT** Unclassified	**20. LIMITATION OF ABSTRACT** UU

NSN 7540–01-280-5500

Standard Form 298 (Rev. 2–89)
Prescribed by ANSI Std. 239–18

THIS PAGE INTENTIONALLY LEFT BLANK

Approved for public release; distribution is unlimited

GRAPH-THEORETIC STATISTICAL METHODS FOR DETECTING AND LOCALIZING DISTRIBUTIONAL CHANGE IN MULTIVARIATE DATA

Matthew A. Hawks
Commander, United States Navy
B.S., United States Naval Academy, 1994
MBA, University of Memphis, 2001
M.S., Massachusetts Institute of Technology, 2006
Naval Engineer, Massachusetts Institute of Technology, 2006

Submitted in partial fulfillment of the
requirements for the degree of

DOCTOR OF PHILOSOPHY IN OPERATIONS RESEARCH

from the

NAVAL POSTGRADUATE SCHOOL
June 2015

Author: Matthew A. Hawks

Approved by:

Robert Koyak
Associate Professor of
Operations Research
Dissertation Supervisor

Emily Craparo
Assistant Professor of
Operations Research

Kyle Lin
Associate Professor of
Operations Research

Lyn Whitaker
Associate Professor of
Operations Research

Craig Rasmussen
Professor and Chair, Department of Applied Mathematics

Approved by: Robert Dell, Chair, Department of Operations Research

Approved by: Douglas Moses, Vice Provost for Academic Affairs

THIS PAGE INTENTIONALLY LEFT BLANK

ABSTRACT

This dissertation explores the topic of detecting and localizing change in a series of multivariate data using graph-theoretic statistical criteria. Change-detection methods based on graph theory are emerging due to their ability to detect change of a general nature with desirable power properties. The graph-theoretic structures of minimum non-bipartite matching and nearest neighbors according to distances between observations form the basis of our statistical procedures. We consider the computation time to implement the procedures with the detection power of the derived statistics. In a simulation study, we evaluate the power of our proposed statistical tests in a series of vignettes in which the sampling distribution, dimensionality, change parameter (location or scale), change type (abrupt or gradual), and change magnitude each are allowed to vary. We compare detection power with contemporary parametric and graph-theoretic approaches. Although our tests alone do not provide the information needed to localize a change point, we develop a follow-on procedure that satisfies this objective. We illustrate our proposed statistical tests and change-point localization techniques in an application, which demonstrates how several of the apparent limitations of our approach can be surmounted.

THIS PAGE INTENTIONALLY LEFT BLANK

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

THIS PAGE INTENTIONALLY LEFT BLANK

LIST OF ACRONYMS AND ABBREVIATIONS

ACF	autocorrelation function
AIC	Akaike information criterion
AR	autoregressive
ARMA	autoregressive moving average
CUSUM	cumulative sum
ESPM	ensemble sum of pair-maxima
EWFD	early warning fire detection
HPC	high performance computing
IQR	interquartile range
JJS	James, James, and Siegmund
MSE	mean squared error
MST	minimum spanning tree
MV	multivariate
MNBM	minimum non-bipartite match, minimum non-bipartite matching
NN	nearest neighbor, nearest neighbors
NRL	Naval Research Laboratory
PACF	partial autocorrelation function
RFI	radio frequency interference
SPD	sum of pair-differences
SPM	sum of pair-maxima

THIS PAGE INTENTIONALLY LEFT BLANK

EXECUTIVE SUMMARY

This dissertation explores the topic of detecting and localizing change in a series of multivariate data using graph-theoretic statistical criteria. A cluster of fire detection sensors produce such a data series; detecting change may mean the onset of a fire or the degradation of a sensor. This topic is of importance to many other fields of practice, including quality assurance, surveillance, remote sensing, and public health. Considering that distributional assumptions become more difficult to verify as the dimensionality of observations increases, nonparametric methods are gaining traction in the study of multivariate change detection. As a subset of the latter, methods based on graph theory are emerging due to their ability to detect change of a general nature with desirable power properties.

Graph-theoretic statistics are derived from a complete graph in which observations are taken as vertices, and either one edge (undirected) or two edges (directed) connect each pair of vertices. An edge weight is assigned, consisting of the dissimilarity or distance between the two adjacent vertices. A subgraph class, such as one consisting of spanning trees or matchings, imposes a particular structure on its vertices and edges. A minimum subgraph is a member of a subgraph class that minimizes the sum of edge weights. For example, a minimum spanning tree is the least-weight acyclic subgraph that contains a path from each vertex to every other vertex, and a minimum non-bipartite matching is a subgraph that pairs all observations so that the total weight of the edges is minimized. Recent statistical literature (Chen & Zhang, 2015; Friedman & Rafsky, 1979; Rosenbaum, 2005; Ruth, 2009; Ruth & Koyak, 2011) establishes that statistics derived from minimum subgraphs are effective at extracting information about the stability of the sampling distributions from which a sequence of multivariate observations is obtained.

Our examination of graph-theoretic change detection techniques is motivated by the need to find new methods that are both computationally efficient and powerful. In his Ph.D. dissertation, Ruth (2009) demonstrates that ensembles of edge-disjoint non-bipartite matchings, which are large collections of minimum subgraphs, allow the

development of statistical tests that match the performance of the best parametric statistical methods when the assumptions justifying use of the latter are true. Finding ensembles, however, requires computation time on the order of the fourth power of the sample size, which greatly restricts their use on large samples. We develop statistical procedures for two less-computationally intensive subgraphs, minimum k-factors and k-nearest neighbors (kNN), to approximate the information obtained from ensembles. A k-factor is a subgraph in which each vertex is incident to exactly k undirected edges. Although an ensemble of minimum non-bipartite matchings generates a sequence of nested factors, our approach is to find a minimum k-factor without sequential generation, which reduces computation time. For a kNN subgraph each vertex is associated with k other vertices that have the least directed edge weights. Efficient algorithms for finding kNN subgraphs are widely available due to their importance in a wide range of applications. For the same number of edges it is possible to obtain a kNN in a small fraction of the computation time required to solve for an ensemble. Additionally, updating a kNN by adding or removing observations is relatively straightforward, an attribute not shared by ensembles and minimum k-factors.

We propose a class of statistical tests for detecting change in the distribution of a sequence of independent multivariate observations based on summing scores assigned to the edges comprising either a minimum k-factor or a kNN subgraph. An edge score is a function of the sequence labels assigned to the two incident vertices. Investigations dating to Wald and Wolfowitz (1940) have noted that the edges of minimum subgraphs tend to associate observations from the same distribution when the sample is composed of observations from two different distributions. Ruth (2009) extends this property in finding that the vertex labels tend to be closer in sequence than expected when a change in distribution is either abrupt or gradual. Our proposed tests not only demonstrate Ruth's findings on two new classes of minimum subgraphs but also consider the best formulation obtained by allowing both the graph complexity k and the edge scoring to vary.

In a simulation study, we evaluate the power of our proposed statistical tests in a series of vignettes in which the sampling distribution, dimensionality, change parameter

(location or scale), change type (abrupt or gradual), and change magnitude each are allowed to vary. Although, as expected, our tests cannot match the performance of ensemble-based tests, we find that the difference is small taking into consideration the increased computational flexibility that our tests afford.

If the null hypothesis of no distributional change is rejected by our tests, it is of interest to estimate the point in the observation sequence at which such change occurred. Although our tests alone do not provide the information needed to localize a change point, we develop a follow-on procedure that satisfies this objective. Our procedure consists of fitting a series of broken-line regressions to standardized counts of edges that associate a given observation to those occurring earlier in the sequence, using hypergeometric distributions that apply if the null hypothesis of no distributional change were true. The break point of the broken-line regression that minimizes the sum of squared errors is taken as the change-point estimator. We evaluate the performance of this estimator in a series of simulation vignettes and find that its accuracy reflects the power of the procedure that was used to conduct the original hypothesis test.

There are limitations to the applicability of the graph-theoretic statistical procedures that we propose. First, these procedures are designed for independently sampled observations, an assumption that frequently is violated in practice. Second, these procedures are powerful in detecting any type of distributional change, including those for which detection may not be desired. And third, these procedures are not designed for real-time implementation. In some cases, it is possible to surmount these limitations, which we demonstrate with an application of our procedures to the fire-detection data of Gottuk et al. (2002). The presence of a set of control data allows the fitting of a time-series model, based only on current and past observations, that removes much of the temporal dependence. Low-level stimuli that affect the sensors but do not justify the setting of an alarm are handled by filtering with a scatterplot smoother that is truncated at a determined threshold value. And, although we process the data off-line in batches, we do so in a manner that near-real time detectability may be possible.

LIST OF REFERENCES

Chen, H., & Zhang, N. (2015). Graph-based change-point detection. *The Annals of Statistics, 43*(1), 139–176.

Friedman, J. H., & Rafsky, L. C. (1979). Multivariate generalizations of the Wald-Wolfowitz and Smirnov two-sample tests. *The Annals of Statistics*, 697–717.

Gottuk, D. T., Wright, M. T., Wong, J. T., Pham, H. V., Rose-Pehrsson, S. L., Hart, S. … Street, T. T. (2002). *Prototype early warning fire detection system: test series 4 results* (No. NRL/MR/6180--02-8602). Washington, DC: Naval Research Lab.

Rosenbaum, P. R. (2005). An exact distribution ‐ free test comparing two multivariate distributions based on adjacency. *Journal of the Royal Statistical Society: Series B (Statistical Methodology), 67*(4), 515–530.

Ruth, D. M. (2009). *Applications of Assignment Algorithms to Nonparametric Tests for Homogeneity* (Ph.D. dissertation, Naval Postgraduate School, Monterey, CA).

Ruth, D. M., & Koyak, R. A. (2011). Nonparametric tests for homogeneity based on non-bipartite matching. *Journal of the American Statistical Association, 106*(496), 1615–1625.

Wald, A., & Wolfowitz, J. (1940). On a test whether two samples are from the same population. *The Annals of Mathematical Statistics, 11*(2), 147–162.

ACKNOWLEDGMENTS

This dissertation may bear my name, but many talented people supported its production.

I am deeply indebted to my advisor, Professor Robert Koyak, for his patience, support, generosity, resourcefulness, innovative spirit, and creative vision. I also thank the committee members—Emily Craparo, Kyle Lin, Craig Rasmussen, and Lyn Whitaker—for their insightful contributions.

We owe a special thanks to Dave Ruth of the Naval Academy for his previous and ongoing work in this research area. Emily Craparo suggested the idea of change detection using nearest neighbors. Susan Rose-Pehrsson and Mark Hammond of the Naval Research Laboratory generously shared their data and reports. The staff of the Naval Postgraduate School's Dudley Knox Library provided superb research support. Camille Rogers and Michele D'Ambrosio helped to clarify our message.

Our research benefitted from the use of IBM's CPLEX software and the Naval Postgraduate School's High Performance Computing (HPC) Center. Professor Sam Buttrey, Tom Halwachs, and HPC team members Brian Andrus, Donna Burych, Jeff Haferman, and Milan Vukcevich provided invaluable assistance.

I would like to thank the many others who did their best to build or rebuild my academic foundations through their diligent instruction and accumulated knowledge: Mike Atkinson, Jerry Brown, David Canright, Matt Carlyle, Tim Chung, Dan Cunningham, Ron Fricker, Jeff Hyink, Pat Jacobs, Wei Kang, Doug MacKinnon, Guillermo Owen, Javier Salmeron, Rachel Silvestrini, and Danny Solow.

My office-mates Christian Klaus, Gary Lazzaro, Dick McGrath, Sofia Miranda, Jesse Pietz, and Austin Wang imparted priceless advice, commiseration, and levity.

The weekly Navigators Bible study group, a rotating cast totaling about 120 men led by Bob Reehm, was a crucial ensemble in this effort. Their comfort, encouragement, and accountability were of inestimable value.

I am thankful for my Faith Lutheran Church family, particularly Pastor Ryan, Pastor Don, Jack, congregational Presidents Deborah and Russ, and my fellow elders Lois, Randy, Harry, Don, and Glenna. All provided unwavering support and wise counsel.

I must acknowledge my childhood family—my older sister, Michelle Clark, for teaching me the joy of honest competition, and my older brother, Harvey Kramer Hawks, for opening my eyes to math. Thanks to my parents, Harv and Ada Hawks, for enduring math contests, inspiring me to excellence, and continuing to encourage my pursuits today.

To my seven children—Danielle, Caleb, Sarah, Mary, David, Joshua, and Hannah—each one of you is a reward; I am truly blessed beyond what I could ever deserve. Now, do your math homework!

Aunt Debby, you have supported this effort more than either of us thought possible. Thank you for following that still, small voice.

I am immeasurably beholden to my wife, Dr. Andra Hawks. My gratitude alone could fill this document, but I will limit myself to a paragraph. Overworked and underappreciated, you supported me even through my stupidity, stubbornness, and stumbling progress. You laid down your abilities, ambitions, and accolades but retained your beauty. I say with sincerity, Andra, "many women have done excellently, but you surpass them all" (Proverbs 31:29 English Standard Version). Thank you. No earthly reward rightly honors you, but I am open to suggestions.

Finally, I would not have completed this document without the mercy and grace of my Lord and Savior Jesus Christ. Throughout this process, He gave me wisdom, knowledge, and understanding.

For we do not have a high priest who is unable to sympathize with our weaknesses,
but one who in every respect has been tempted as we are, yet without sin.
Let us then with confidence draw near to the throne of grace,
that we may receive mercy and find grace to help in time of need.
Hebrews 4:15-16

For the Lord gives wisdom; from his mouth come knowledge and understanding.
Proverbs 2:6

I. INTRODUCTION

Detecting change in the underlying probability distribution of a sequence of observations is an important research area in statistics and other disciplines. The nature of change may be an abrupt shift or a gradual drift in a distributional aspect such as the mean, variance, or a higher moment. The ability to detect change assists in a myriad of situations, for instance diagnosing machinery health, monitoring a manufacturing process, examining social network behavior, considering the possibility of a viral outbreak, or assessing sensor network operation.

JiJi, Hammond, Williams, and Rose-Pehrsson (2003) provide an example of the role change detection plays in monitoring a collection of sensors. The authors describe an experiment in which four types of sensors—photoelectric and ionization smoke detectors, carbon monoxide and carbon dioxide sensors—are used to determine the point at which a combustion event is detectable. The sensors in this experiment are distributed on a decommissioned naval vessel. Twenty-four events, ranging from benign heat-generating activities to smoldering and flaming fires, are staged in various shipboard locations. Of interest is the time that it takes for the sensor network to detect a change that signals a combustion event.

If an analysis detects change, it is natural to ask where in the sequence the change occurred. Localizing the point in the observational sequence at which the change began is broadly known as the change-point problem, or in Eastern literature, the disorder problem (Brodsky & Darkhovsky, 1993). Many methods for change-point localization include an initial test for a distributional change (see Bhattacharya, 1994 and Chen & Gupta, 2001).

Change detection and localization techniques may be categorized as either *static* or *sequential*. Static analysis is done retrospectively; sequential analysis is done concurrent with observation. Static change-detection applications arise out of forensic situations, such as analyzing vehicle death rates before and after passage of a seat belt law (Bhattacharyya & Layton, 1979). Sequential change detection occurs in cases where response time is a priority, such as fire detection. Quality control specialists use

multivariate sequential analysis techniques such as Shewart charts (Hotelling, 1947), cumulative sum (CUSUM) schemes (Woodall & Ncube, 1985; Crosier, 1988) and exponentially weighted smoothing (Lowry, Woodall, Champ, & Rigdon, 1992) to signal degradation in quality. Static analysis allows more computation time compared to sequential analysis. Static change detection is also described in literature as non-sequential, offline, retrospective, or posterior, whereas sequential analysis may be referred to as real-time, dynamic, or streaming. Sequential testing to localize change points in multivariate streaming data is a growing area of research, notably in the machine learning community (see for example, Dries & Rückert, 2009).

Change-detection methods are also characterized as either parametric or non-parametric. Parametric approaches to change detection rely upon distributional assumptions, typically normality. These assumptions are difficult to verify as the number of variables increases, especially if the number of variables is large. Although approaches based on multivariate normality are commonly used, especially in quality control, nonparametric approaches are more attractive for multivariate investigations. There is a need for nonparametric approaches, given that data often fail to meet distributional assumptions.

The purpose of this dissertation is to propose multivariate, nonparametric methods for change detection and change-point localization using graph-theoretic techniques. We limit our investigation to static change analysis, and through an application show that our methods are adaptable for near-sequential analysis.

Graph-theoretic techniques for change detection have their origin in the famous runs test of Wald and Wolfowitz (1940), which provides a univariate approach to the two-sample problem. Friedman and Rafsky (1979) develop a multivariate extension of the runs test using minimum spanning trees. Rosenbaum (2005) uses a test based on matching techniques to detect change in multivariate data. Matching methods are extended by Ruth (2009) and Ruth and Koyak (2011) to include richer information from the graph. They propose a graph-theoretic test for multivariate change detection with power comparable to parametric procedures when parametric assumptions are met. These methods are described in Chapter II.

This dissertation is organized as follows. In Chapter II, we introduce terminology and review the literature on change detection, change-point localization, and graph-theoretic methods in the multivariate setting. Chapter III introduces statistical tests based on three graph-theoretic statistics: the k-factor sum of pair-differences, the k-nearest-neighbors sum of pair-differences, and the nearest-neighbors prior-edge count. We discuss computational considerations for extracting information from graphs, explore the test statistic formation and characteristics, and consider approaches to obtaining a p-value for testing purposes. Through a simulation study, the power of these tests to detect change is estimated. In Chapter IV, we present and evaluate two graph-theoretic change-point localization techniques for an observation sequence containing an underlying distributional change. In Chapter V, we apply our graph-theoretic tests and change-point localization techniques to the fire detection sensor data considered by JiJi et al. (2003). Consideration is given to adapting time-series data to tests that assume data independence. Chapter VI summarizes our findings and discusses areas for further research in this field.

THIS PAGE INTENTIONALLY LEFT BLANK

II. REVIEW OF GRAPH-THEORETIC METHODS FOR CHANGE DETECTION

In this chapter, we introduce terminology and review recent advances in graph-theoretic change detection and change-point localization that are relevant to our investigation.

A. A BIVARIATE EXAMPLE

We use the two data sets listed in Table 1 to illustrate change-detection methods. The first data set (no-change case) consists of 20 observations $\mathbf{y}_1,\ldots,\mathbf{y}_i,\ldots,\mathbf{y}_{20}$ sampled from a standard bivariate normal distribution. The second data set (change case) has 10 observations sampled from a standard bivariate normal distribution, followed by 10 observations sampled from a bivariate normal distribution with identity covariance matrix and mean vector (3, 0).

Table 1. Bivariate example data. The no-change case is to the left, with 20 observations drawn from a standard bivariate normal distribution. The change case is to the right, with observations 11–20 drawn from a bivariate normal distribution having mean vector (3, 0).

Obs. (i)	y_{i1}	y_{i2}	Obs. (i)	y_{i1}	y_{i2}
1	−0.435	0.957	1	0.433	0.523
2	−0.258	−0.340	2	0.675	−0.111
3	0.386	0.034	3	−0.740	−0.917
4	−0.954	−0.224	4	−0.412	−0.367
5	−1.678	0.490	5	0.583	0.065
6	−0.621	1.168	6	−1.068	−0.699
7	1.573	0.971	7	0.196	0.070
8	0.173	−0.787	8	1.086	0.309
9	0.881	−0.527	9	1.690	0.337
10	−0.591	−0.802	10	−0.105	0.126
11	1.007	−0.045	11	2.884	−0.384
12	−0.839	−2.092	12	3.324	−0.320
13	−0.315	−1.238	13	2.342	−0.168
14	−0.515	−1.186	14	3.584	0.448
15	−0.458	0.104	15	3.794	0.559
16	−0.488	−1.359	16	1.920	0.937
17	−1.610	0.282	17	4.437	1.363
18	−0.786	0.488	18	2.948	−0.445
19	−0.076	−1.044	19	1.920	−0.323
20	2.180	0.008	20	3.165	−2.158

To use the fire detection example we explore in Chapter V, consider the observation label (i) as a time index and the two variables (y_{i1} and y_{i2}) as standardized measurements from an ionization smoke detector and a photoelectric smoke detector, respectively. The change case records an event occurring between observations 10 and 11 that only triggers a response from the ionization smoke detector. Figure 1 depicts the data graphically.

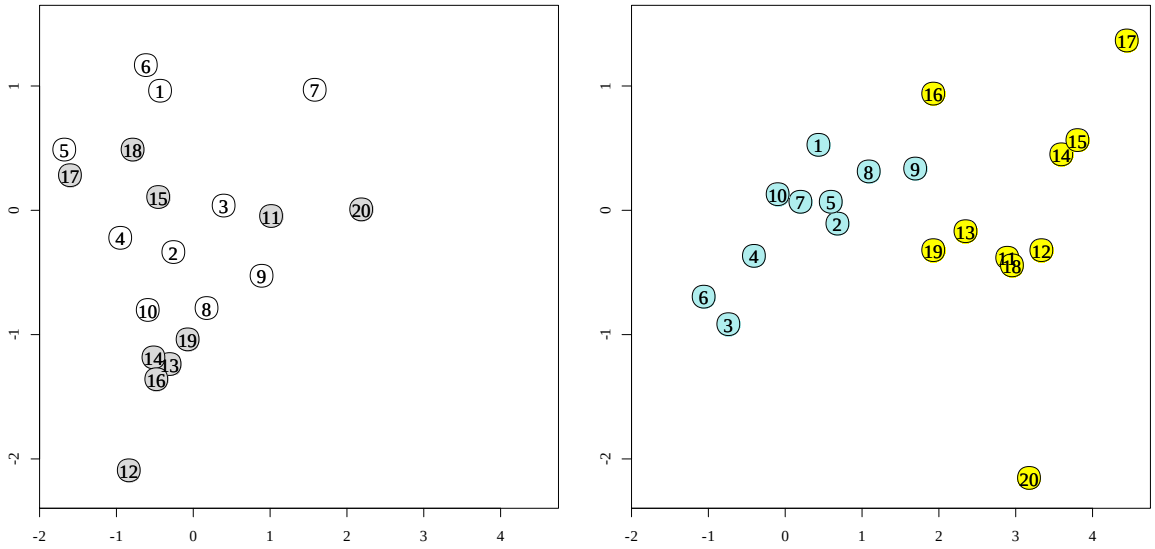

Figure 1. Bivariate example data, labeled by observation index. The no-change case is to the left: 20 observations drawn from a standard bivariate normal distribution. Observations 11–20 are colored gray to facilitate later comparisons. The change case is to the right: 20 observations drawn from a bivariate normal distribution with identity covariance matrix and an abrupt change in the mean vector from (0, 0) to (3, 0) starting at observation 11. Observations 1–10 are shown in blue, observations 11–20 are shown in yellow. The circles marking observations 11 and 18 overlap.

B. PARAMETRIC APPROACHES TO CHANGE DETECTION

Much of the change detection and localization literature concerns parametric methods. A recent bibliography on the change-point problem (change detection and change-point localization) is compiled by Lee (2010), stressing the relative dearth of nonparametric methods. Khodadadi and Asgharian (2008) provide a bibliography concerning the change-point problem in the context of regression. Tran, Gaber, and Sattler (2014) describe recent change-detection efforts as applied to streaming data.

James, James, and Siegmund (1992) present a likelihood ratio test approach for detecting a change in the mean vector of a multivariate normal distribution. The James, James, Siegmund (JJS) method remains a pillar in parametric change detection (Chen & Gupta, 2011), and we use this test as an example of a parametric approach to compare with the graph-theoretic methods presented later. JJS formulate the test statistic for detecting a single change in the mean vector of a multivariate normal distribution as follows. For independent p-dimensional multivariate normal observations $\mathbf{Y}_1,\ldots,\mathbf{Y}_i,\ldots,\mathbf{Y}_N$ with a common covariance matrix and a change-point candidate τ in $\{\tau_0,\ldots,\tau_1\}$, the modified likelihood ratio test statistic is

$$T_{\mathrm{JJS}} = \max_{\tau_0\le\tau\le\tau_1} \frac{N}{\tau(N-\tau)}\left(S_\tau - \frac{\tau}{N}S_N\right)'\left(U_N - \frac{1}{N}S_N S_N'\right)^{-1}\left(S_\tau - \frac{\tau}{N}S_N\right)$$
$$S_\tau = \sum_{i=1}^{\tau}\mathbf{Y}_i, \quad U_N = \sum_{i=1}^{N}\mathbf{Y}_i\mathbf{Y}_i'. \tag{1}$$

JJS also present an approximation to the tail probabilities as follows:

$$\Pr[T_{\mathrm{JJS}} \ge x] \cong$$
$$\frac{1}{\Gamma(p/2)}\left(\frac{Nx}{2}\right)^{p/2}(1-x)^{(N-p-3)/2}\int_{t_0}^{t_1}\frac{1}{t(1-t)}v\left(\left(\frac{x}{t(1-t)(1-x)}\right)^{1/2}\right)dt,\ 0<x<1, \tag{2}$$

where $\tau_0/N \to t_0$ and $\tau_1/N \to t_1$ as $N\to\infty$, $0<t_0\le t_1<1$, and

$$v(t) = \frac{2}{t^2}\exp\left\{-2\sum_{\tau=1}^{\infty}\frac{1}{\tau}\Phi\left(-\frac{t\tau^{1/2}}{2}\right)\right\}, \tag{3}$$

where $\Phi(t)$ is the standard normal cumulative distribution function. The JJS test detects change in an interval of possible change points, but does not attempt to localize it. For the example data sets of Section A, we obtain JJS p-values of 0.17 for the no-change case, and p-value 0.000029 for the change case. The conclusion is that there is a distributional change in the change case.

Parametric methods are developed for specific distributions and may lack robustness when applied to other underlying data distributions, as can be the case when

the sampling distribution is unknown. This was explicitly shown by Ruth (2009) for JJS. We next describe nonparametric approaches to change detection in multivariate data.

C. NONPARAMETRIC APPROACHES TO CHANGE DETECTION

An overview of both static and sequential nonparametric change-detection methods is contained in Chapter 2 of Brodsky and Darkhovsky (2000). A nonparametric approach to detect and localize mean change in multivariate data is introduced in Darkhovsky (1994) using the family of Kolmogorov-Smirnoff type statistics. Qui and Hawkins (2003) present a nonparametric cumulative sum (CUSUM) procedure for detecting shifts in the mean vector of a multivariate measurement. Their distribution-free procedure uses order information from both the measurements and their null hypothesis means. Lung-Yut-Fong, Lévy-Leduc, and Cappé (2012) contribute a nonparametric two-sample homogeneity test for multivariate data based on the Wilcoxon rank statistic. They apply their test to static change detection to estimate multiple change points. Li, Zhang, and Jeske (2013) propose two multivariate CUSUM procedures as nonparametric counterparts to the parametric multivariate CUSUM of Crosier (1988). The two procedures, based on spatial sign and data depth, show power equivalent to Crosier's multivariate CUSUM in detecting location and scale changes when applied to normally distributed data, and higher power than Crosier's test when applied to non-normally distributed data. Holland and Hawkins (2014) provide a nonparametric CUSUM method of detecting location changes in multivariate data without requiring significant data history. Their procedure is based on an approximately distribution-free generalization of the Wilcoxon-Mann-Whitney test. Preuss, Puchstein, and Dette (2014) pursue spectral distribution estimation to detect multiple covariance structural changes of a multivariate process. Matteson and James (2014) use nonparametric hierarchical clustering to conduct static change detection and localization in multivariate data. Their procedure allows for detection of distributional change of any nature and includes testing for estimating multiple change points.

D. GRAPH-THEORETIC APPROACHES TO CHANGE DETECTION

An area of nonparametric change analysis that has shown promise is graph-theoretic statistics. We review graph-theoretic procedures for change analysis, starting with methods developed for the classic two-sample problem (which we specify below), and proceed to current change detection and localization techniques.

1. Graph Theory Terminology

We introduce a few graph theory concepts, following conventions in West (2001). These terms are summarized in Table 2 at the end of this chapter. A *graph G* consists of a *vertex* set *V*, an *edge* set *E*, and a relationship between vertices and edges. We convey the vertex-edge relationship in a drawing with vertices as points and edges as lines, or by naming each edge by its two incident vertices, e.g., edge $\{i, j\}$. The two vertex endpoints of an edge are *neighbors* and *adjacent*. A *simple* graph has no duplicate edges or edges with the same starting and ending vertex. This dissertation is restricted to simple graphs. The *degree* of a vertex is the number of incident edges. All vertices of a *regular* graph have the same degree, say *k*; such a graph is *k*-regular. A *directed* graph specifies an ordered pair of vertices for each edge: (*tail*, *head*). For example, directed edge (i, j) goes from vertex *i* to vertex *j*. For a directed graph, the *outdegree* of vertex *i* is the number of edges with tail *i*; the *indegree* of vertex *i* is the number of edges with head *i*. Edges may be assigned *edge weights* according to some nonnegative measure of distance or cost. A *complete* graph has edges between all possible vertex pairs.

Having defined the basic elements of a graph, we continue with additional concepts. A *subgraph* is a subset of a graph; said graph *contains* the subgraph. A *spanning subgraph* of a graph is a subgraph including all vertices of the graph. A *factor* is a spanning subgraph. A *k-factor* is a spanning *k*-regular subgraph. A *minimum* spanning subgraph with property *P* is the spanning subgraph of *G* with property *P* that minimizes the sum of the included edge weights. Subgraphs having no vertices in common are *disjoint*, while subgraphs having no edges in common are *edge-disjoint*. Two or more edge-disjoint subgraphs with identical vertex sets are an *ensemble*. A *matching* is a set of edges with no shared endpoints. A *perfect* matching is a matching

incident to every vertex. The edge set of a 1-factor is a perfect matching. An undirected *nearest-neighbor* subgraph includes only those edges connecting each observation to its nearest (minimum edge weight) neighbor.

Continuing our introduction of graph theory concepts, a *path* is a graph whose vertices are adjacent if and only if they are consecutive in an ordered list. A *Hamiltonian* path is a spanning path. A *cycle* is a closed path where the first and last vertices are the same. A *connected* graph contains at least one path from each vertex to every other vertex. A *tree* is an acyclic connected graph. A *spanning tree* is a spanning subgraph that is a tree. The edge cardinality $|E|$ of a spanning tree is one less than the vertex cardinality $|E|$. A minimum spanning tree (MST) is the spanning tree that minimizes the sum of the included edge weights. A *degree-constrained* MST (*k*MST) has the additional constraint that the maximum degree of each vertex is *k*. A 2MST is a shortest Hamiltonian path.

Now we apply graph theory to change detection. We represent the data under analysis as a complete graph. Let *N* be the number of observations. Each observation is a vertex, labeled sequentially from 1 to *N*, while undirected edges represent connections between observations (we consider directed edges later). A complete graph of *N* observations has edge cardinality $|E| = N(N-1)/2$, and each of the *N* vertices is of degree $N-1$. Edge weights (d_{ij}) reflect a standard distance metric such as Euclidean, Manhattan (L_1 distance), Chebyshev (L_∞ metric), or Mahalanobis (scale-invariant) distance. To detect a change in the underlying distribution of the data, we analyze the behavior of minimum spanning subgraphs.

2. The Two-Sample Problem

The two-sample problem is posed as follows. Let $\mathbf{Y}_1,\ldots,\mathbf{Y}_i,\ldots,\mathbf{Y}_N$ represent *p*-dimensional observations, such that m_1 observations are independently sampled from distribution F_1, and m_2 observations are independently sampled from distribution F_2 $(N = m_1 + m_2)$. Two distributions are equal ($F_1 = F_2$) if and only if their cumulative distribution functions $F_1(t)$ and $F_2(t)$ and equal for all *t*. Based on the samples, is there

enough evidence to reject the null hypothesis that the distributions F_1 and F_2 are identical? H_0: $F_1 = F_2$; H_1: $F_1 \neq F_2$.

For univariate observations (p = 1), Wald and Wolfowitz (1940) propose a statistic derived by sorting the data in non-decreasing order regardless of the sample source (e.g., $Y_1 \leq \ldots \leq Y_j \leq \ldots \leq Y_N$, $\forall j \in \{2,\ldots,N-1\}$). We count the total number of runs, a run being one or more consecutive observations (as sorted) from one source. A relatively small count of runs is indicative of different underlying distributions. Subsequent authors extend the two-sample problem to change analysis where F_1 and F_2 are multivariate.

3. A Test Using the Minimum Spanning Tree

Friedman and Rafsky (1979) propose a MST-based test to generalize the Wald-Wolfowitz runs test to multivariate data. The context again is the two-sample problem, in which there are two samples of known size (m_1, m_2) drawn from two unknown multivariate distributions (F_1, F_2). The observations are represented as a complete graph, with the vertices (observations) identified by sample source, edges linking all pairs of vertices, and edge weights d_{ij} given by Euclidean distance or some other measure of vertex dissimilarity. The MST has edge cardinality $|E_{\text{MST}}| = N-1$. An edge indicator x_e is assigned, which is equal to one if edge e links vertices from different samples, and x_e = 0 otherwise. All edges with indicator x_e = 1 are deleted, and the resulting number of disjoint subtrees R is obtained. This results in a statistic similar to a runs test:

$$
\begin{aligned}
d_{ij} &= \left|\left|\mathbf{Y}_i - \mathbf{Y}_j\right|\right| = \sqrt{\sum_{r=1}^{p}\left(Y_{ir} - Y_{jr}\right)^2}, \quad \forall\{i,j\} \in E \\
x_e &= \begin{cases} 1, & \text{if edge } e \text{ links vertices from different samples} \\ 0, & \text{otherwise} \end{cases}, \quad \forall e \in \{1,\ldots,N-1\} \qquad (4) \\
R &= \sum_{e=1}^{N-1} x_e + 1.
\end{aligned}
$$

This statistic is nonparametric, but it is not distribution-free. Friedman and Rafsky show the variance of R depends on the topology of the MST; specifically, the variance depends

on number of MST edges that share a common vertex. Standardizing the R statistic based on the null hypothesis expected value and conditional variance provides a Z score that can be compared to the standard normal distribution to obtain a p-value.

A relatively small disjoint subtree count indicates heterogeneity between the two samples. The extreme cases give insight as to why this test is able to detect heterogeneity. The minimum possible value of R is 2, indicating that the MST includes only a single edge incident to observations in both samples. The maximum possible value of R is N, indicating that the MST consists solely of edges incident to observations in both samples. Figure 2 shows the MSTs for the example data sets presented above in Section A.

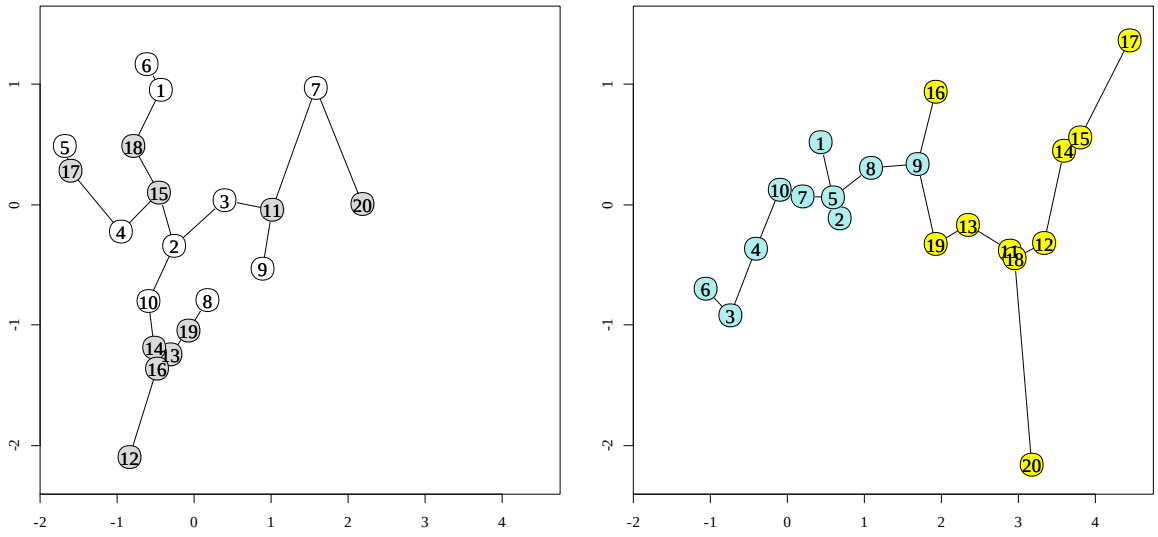

Figure 2. Minimum spanning trees for example data sets. The no-change case is to the left; the change case is to the right.

Using $m_1 = m_2 = 10$ for both cases, the no-change case and the change case yield R values of 12 and 3, respectively. The resulting p-values from Friedman and Rafsky's test are 0.68 and 0.000093, respectively. Thus, the null hypothesis that $F_1 = F_2$ is not rejected in the no-change case, and it is rejected in the change case. Zhou, Zi, Geng, and Li (2014) apply the MST in a statistical process control setting to detect change.

As suggested by John Tukey, Friedman and Rafsky (1979) show that including a second or third edge-disjoint MST generally improves the heterogeneity-detection power

of the test. Unfortunately, there is no guarantee that one may extract an additional edge-disjoint MST from the samples. Constructing a MST requires $O\left(|E|\log N\right)$ computation time (West, 2001); for a complete graph, $|E|$ is $O\left(N^2\right)$.

4. Tests Using the Minimum Matching

Rosenbaum (2005) extends the work of Friedman and Rafsky by using a minimum-matching methodology and producing the cross-match test, which, unlike the Friedman and Rafsky test, is distribution-free. Again, the data in question are known to be from two samples. For the sake of discussion, we assume N is even, although an odd value of N may also be accommodated. Instead of minimum spanning trees, the cross-match test uses the edges of the minimum non-bipartite matching (MNBM). *Non-bipartite* matching refers to pairing observations without regard to their sample identity. The MNBM has cardinality $|E_{\text{MNBM}}| = N/2$.

The context again is the two-sample problem, in which there are two p-dimensional samples of known size (m_1, m_2) drawn from two unknown distributions (F_1, F_2). We represent the observations as a complete, undirected graph (V, E), with the vertex set V referenced as i or j and the edge set E referenced by its endpoints, edge $\{i, j\}$. Using, for example, the Euclidean distance metric, we define the edge weights d_{ij} as the distances between observations. A set of binary edge indicator variables x_{ij} indicates MNBM membership, with $x_{ij} = x_{ji}$:

$$\begin{aligned} d_{ij} &= \left\|\mathbf{Y}_i - \mathbf{Y}_j\right\| = \sqrt{\sum_{r=1}^{p}\left(Y_{ir} - Y_{jr}\right)^2}, \quad \forall\{i, j\} \in E \\ x_{ij} &= \begin{cases} 1, & \text{if } \{i, j\} \in \text{MNBM} \\ 0, & \text{otherwise} \end{cases}, \quad \forall\{i, j\} \in E. \end{aligned} \tag{5}$$

In matrix form, **X** is also known as the MNBM *adjacency matrix*. We solve the following integer linear program for **X**, the edge indicators of the MNBM:

$$\min_{x} \sum_{\{i,j\}\in E} d_{ij} x_{ij}$$
$$\text{s.t.} \quad \sum_{j\in V} x_{ij} = 1, \ \forall i \in V \tag{6}$$
$$x_{ij} \in \{0,1\}, \ \forall \{i,j\} \in E.$$

In a MNBM, observations are more likely to be adjacent to observations drawn from their same distribution rather than observations drawn from the other, as explained in Rosenbaum (2005). The cross-match statistic *A* counts those edges in the MNBM that include an endpoint from each sample. A smaller value of the statistic, i.e., more edges within samples as opposed to across samples, provides evidence against identical source distributions. Under the null hypothesis, the probability of exactly *a* edges connecting observations from the two samples is

$$P(A=a) = \frac{2^a (N/2)!}{\binom{N}{m_2}((m_1-a)/2)!\,a!\,((m_2-a)/2)!}, \ 0 \le a \le \min(m_1, m_2). \tag{7}$$

Figure 3 shows the MNBMs for the example data sets of Section A.

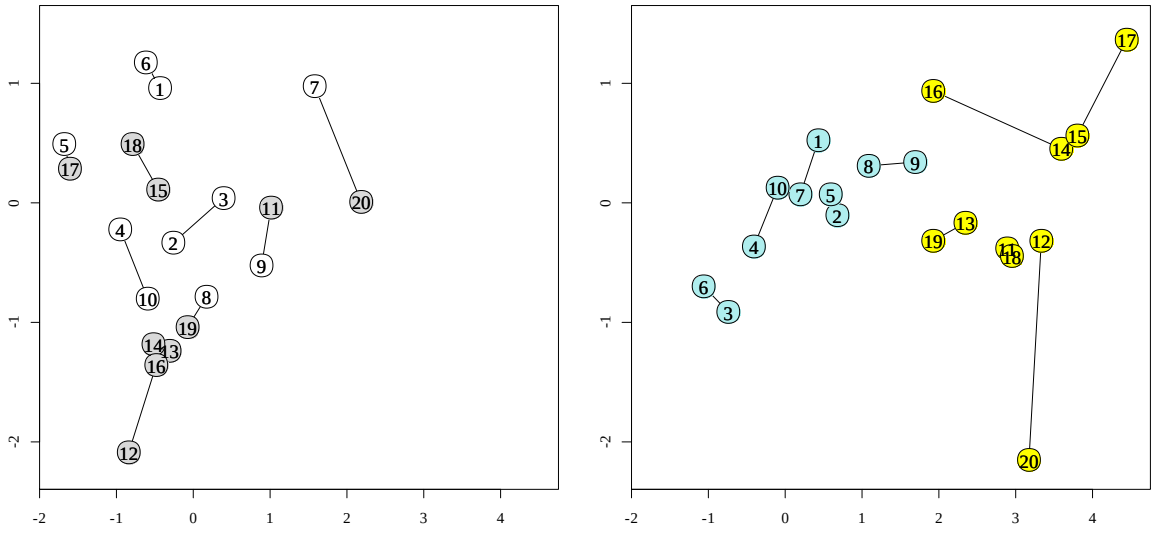

Figure 3. Minimum non-bipartite matchings for example data sets. The no-change case is to the left; the change case is to the right.

Using 10 for both sample sizes m_1 and m_2, the no-change case and the change case yield A values of 4 and 0, respectively. The resulting p-values from Rosenbaum's test are 0.43 and 0.0014, respectively, failing to reject the null hypothesis in the no-change case, and rejecting it in the change case. The empirically observed computation time required to extract a MNBM with the Kolmogorov (2009) implementation of the Blossom algorithm is on the order of $N^2 \log N$.

Other minimum subgraphs have been considered for change detection as well. Biswas, Mukhopadhyay, and Ghosh (2014) propose a two-sample statistic based on the approximate shortest Hamiltonian path as defined earlier in this section. The shortest Hamiltonian path problem is non-deterministic polynomial-time (NP) hard; the authors employ a suboptimal heuristic algorithm. In higher dimensions, their test shows power superior to both Friedman and Rafsky's MST test and Rosenbaum's cross-match test.

Recent work extends these two-sample tests into the realm of change detection and localization. Ruth (2009) and Ruth and Koyak (2011) generalize Rosenbaum's cross-match test by considering a sequence of observations instead of two samples. Their exact test iteratively divides sequentially ordered observations into two subsets, the first m_1 observations and the last $m_2 = N - m_1$ observations, $\forall m_1 \in \{2,3,\ldots,N-1\}$. Each two-subset sequence is evaluated using the edges of the same MNBM to yield a set of cross-match statistics, which detect change in an interval of possible change points.

Under the alternative hypothesis of distributional change in an observation sequence, a MNBM tends to pair observations that are closer in sequence than would be expected under the null hypothesis. The sum of pair-differences across all pairs should accordingly be smaller in situations when the null hypothesis should be rejected. Ruth (2009) considers an equivalent statistic, the sum of pair-maxima (SPM):

$$T_N = \sum_{i=2}^{N} \sum_{j=1}^{i-1} i x_{ij}. \tag{8}$$

Ruth establishes a normal approximation rejection criterion for T_N that is valid under the null hypothesis that there is no change. For the example data of Section A, the no-change

case and the change case yield T_{20} values of 134 and 127, respectively. From Ruth (2009), the critical value for T_{20} at the 0.05 test level is 129, so we reject the null hypothesis at the 0.05 test level for the change case.

Chen and Zhang (2015) propose a statistic based on the number of cross-matches in a MST, MNBM, or undirected nearest-neighbor subgraph by allowing the first sample to consist of vertices $\{1, \ldots, j\}$, $0 < j < N$. They use their statistic to determine single or multiple change points, in both low- and high-dimensional data. Like Friedman and Rafsky (1979), Chen and Zhang demonstrate increased power for richer graph structures. They consider three edge-disjoint MSTs, MNBMs, and undirected nearest-neighbors subgraphs.

5. Tests Using Ensembles and Factors

Graph-theoretic tests derive their power from the idea that when observations are drawn from two different distributions, the edges of a minimum subgraph such as a MST or MNBM tend to associate observations from the same distribution more often than those from different distributions. Edge-disjoint ensembles further exploit this idea, as the use of second and third edge-disjoint MSTs show in Friedman and Rafsky (1979) and Chen and Zhang (2015). Ruth (2009) and Ruth and Koyak (2011) show that, compared to a single MNBM, ensembles of higher order provide substantial increases in power.

Ensembles are obtained recursively as follows:

(1) Start with the complete graph.

(2) Extract the MNBM using the Blossom algorithm of Kolmogorov (2009).

(3) Remove the MNBM edges from the graph; they are part of the ensemble.

(4) Repeat steps (2) and (3) until the desired ensemble order is achieved.

Let k denote the number of edge-disjoint MNBMs in an ensemble. Let kMNBM refer to an ensemble of k different MNBMs. The edge cardinality of a kMNBM is $\left|E_{k\text{MNBM}}\right| = kN/2$. If the number of observations N is even, than $N/2$ distinct edge-disjoint subgraphs are guaranteed to be obtainable (Anderson, 1972).

Ruth (2009) introduces the ensemble sum of pair-maxima (ESPM) test based on forming $k = N/2$ edge-disjoint MNBMs and the previously discussed SPM statistic. The cumulative sum of pair-maxima over the first k MNBM is

$$S_k = \sum_{j=1}^{k} T_j, \quad k = 1, \ldots, N/2, \tag{9}$$

where T_j is the SPM statistic for the jth MNBM of the ensemble. Ruth expresses the ESPM statistic as follows:

$$B = \max_{k \in \{1,2,K\ N/2\}} \left(\frac{E[S_k] - S_k}{c} \right), \quad \text{where } c = (N-1)\sqrt{\frac{N(N+1)}{180}}. \tag{10}$$

The similarity of sequential values of S_k to a Brownian bridge motivates the scaling constant c and the formulation for the ESPM statistic, as shown in Ruth (2009). The ESPM statistic is not distribution-free under the null hypothesis that all observations are from the same distribution.

The ESPM statistic achieves desirable power—essentially matching the change-detection power of JJS under JJS-optimal conditions, and exceeding parametric methods when applied to non-normally-distributed data. Ruth and Koyak (2011) demonstrate that the power of an ensemble-based test increases as the number of matches increases. The empirically observed computation time for recursive matching is on the order of N^4, a consequential drawback for larger data sets.

Bodgan (2014) applies the ESPM to sequential testing, insofar as proposing appropriate methods to maintain test level and achieve reasonable power during multiple testing. This approach shows promise to adapt other static methods to sequential scenarios, while recognizing the need for more computationally efficient graph-theoretic approaches.

Is there a statistic that achieves the detection power of the ESPM, yet comes from an alternative subgraph extracted in a way that does not require recursive matching? We

propose advancing directly to a minimum *k*-regular spanning subgraph. Such a subgraph, referred to as the minimum *k*-factor, shares the same edge cardinality as the *k*MNBM. Unlike the *k*MNBM, the minimum *k*-factor is not necessarily decomposable into a series of *k* edge-disjoint MNBMs and is not recursively determined. Using the set of binary edge indicator variables x_{ij}, we obtain the minimum *k*-factor adjacency matrix (**X**) by making the following slight modification to the optimization problem of (6):

$$\begin{aligned} \min_{\mathbf{x}} \; & \sum_{\{i,j\}\in E} d_{ij} x_{ij} \\ \text{s.t.} \quad & \sum_{j\in V} x_{ij} = k, \;\; \forall i \in V \\ & x_{ij} \in \{0,1\}, \;\; \forall \{i,j\} \in E. \end{aligned} \tag{11}$$

In Chapter III, we demonstrate that by using optimization software such as CPLEX (IBM, 2013), obtaining a minimum *k*-factor has an empirically observed computation time on the order of $N^{2.5}$.

As discussed in more detail in Chapter III, we develop a statistic based on an edge scoring function $h(i, j)$ applied to the edge indicator x_{ij}:

$$T = \sum_{\{i,j\}\in E} h(i,j)\, x_{ij}. \tag{12}$$

Some example candidate scoring functions already discussed include a cross-match indicator and the pair-maximum:

$$\begin{aligned} h(i,j) &= \begin{cases} 1, & \text{if vertices } i \text{ and } j \text{ are from different samples} \\ 0, & \text{otherwise} \end{cases} \\ h(i,j) &= \max\{i,j\}. \end{aligned} \tag{13}$$

6. Tests Using Nearest-Neighbor Graphs

Statistics derived from nearest neighbors (NN) have also been used in change-detection problems. NN is a subgraph of the complete directed graph with $|V|$ directed edges, one from each vertex, such that each edge (i, j) means vertex j is the NN of vertex i. A NN graph is not symmetric; $(i,j) \in E_{\text{NN}}$ does not imply $(j,i) \in E_{\text{NN}}$. NN graph

vertices all have outdegree one, but their indegree may take any integer value on $[0, N-1]$. Figure 4 shows the NN graphs for the example data sets of Section A.

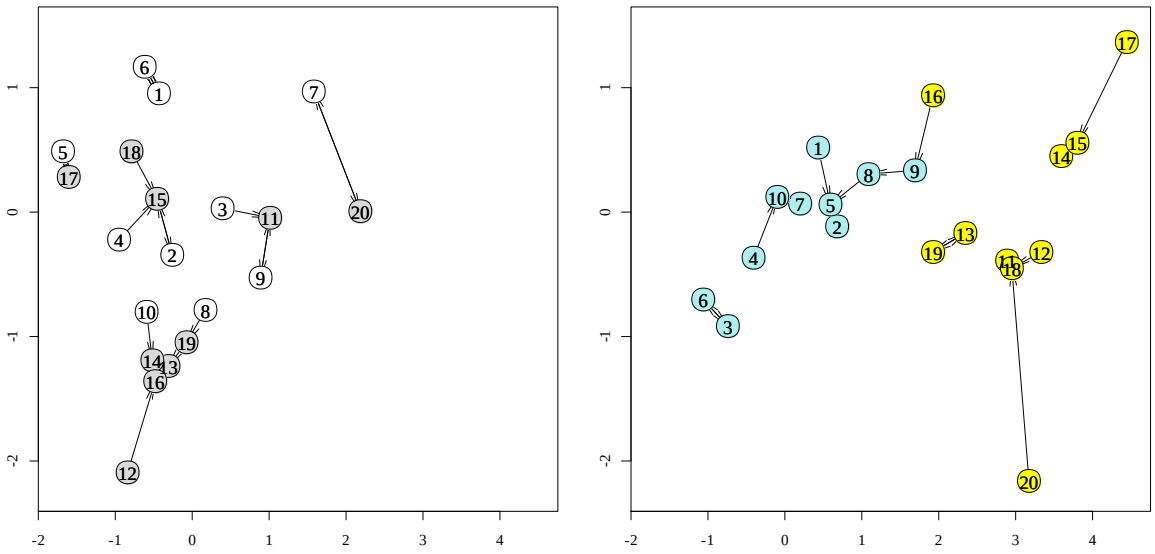

Figure 4. Nearest-neighbor graphs for example data sets. The no-change case is to the left; the change case is to the right.

The undirected NN graph is a subgraph of the MST. A *k*NN graph includes edges to the *k*-nearest neighbors of each vertex and has edge cardinality $|E_{kNN}| = kN$.

Most applications of NN fall generally into the areas of classification and regression. The use of NN in classification, often associated with machine learning and pattern recognition, involves assigning an observation to a group based on the group identification of its NN (Kotsiantis, 2007). NN regression typically involves assigning a value to an observation based on the value of its NN (Altman, 1992).

Schilling (1986) and Henze (1988) both provide similar statistics for multivariate two-sample testing using NN. Let $NN_i(r)$ be the *r*th NN of observation *i*. We define indicator variable $I_i(r)$ by

$$I_i(r) = \begin{cases} 1, & \text{if } NN_i(r) \text{ belongs to the same sample as observation } i \\ 0, & \text{otherwise.} \end{cases} \tag{14}$$

Now the statistic for testing the null hypothesis of identical sampling distributions is given by (Henze, 1988)

$$T_{k,N} = \sum_{i=1}^{N} \sum_{r=1}^{k} I_i(r). \tag{15}$$

This quantity counts all edges in which an observation and its neighbor are from the same sample. Large values of $T_{k,N}$ signal rejection of the null hypothesis. Henze shows that under the null hypothesis, the statistic is distribution-free and asymptotically normal, although permutation testing is recommended for smaller sample sizes. Using the single NN as shown in Figure 4, and a sample size of 10 for both m_1 and m_2, the no-change case and the change case yield $T_{1,20}$ values of 8 and 19, respectively. The resulting permutation *p*-values from the test are 0.76 and 0.0002, respectively. We fail to reject the null hypothesis in the no-change case, and reject it in the change case. Schilling conducts a simulation study of the detection power of $T_{k,N}$ for location and scale changes, demonstrating an increase in change-detection power as *k* increases from one to three.

Hall and Tajvidi (2002) propose a NN-based approach to the multivariate two-sample problem, calibrated by permutation testing. Consider a pooled sample $\mathscr{Y}$ of size N composed of two independent random samples Y_1 and Y_2 of size m_1 and m_2 from distributions F_1 and F_2, respectively. Define quantity $M_{1,i}(k)$ as the number of Y_2 observations that are kNN of observation $Y_{1,i}$, $1 \le i \le m_1$, $1 \le k \le N-1$ (and similarly define $M_{2,i}(k)$). Under the null hypothesis, conditional on the pooled sample, $M_{1,i}(k)$ has a hypergeometric distribution (as does $M_{2,i}(k)$):

$$\Pr\left[M_{1,i}(k) = j \middle| \mathscr{Y} \right] = \frac{\binom{m_1 - 1}{k - j}\binom{m_2}{j}}{\binom{N-1}{k}}, \; 0 \le j \le k, \tag{16}$$

which implies that, under the null hypothesis:

$$E\left[M_{1,i}(k) \middle| \mathscr{Y} \right] = \frac{m_2 k}{N-1}. \tag{17}$$

The conditional probability and expected value of $M_{2,i}(k)$ are analogous, interchanging m_1 and m_2. The test statistic T is based on the absolute deviations of M_1 and M_2 from their expected value under the null hypothesis, denoted as DM_1 and DM_2 respectively. Hall and Tajvidi include $\gamma > 0$ and nonnegative weights w_1 and w_2 as adjustable parameters for T:

$$T = \frac{1}{m_1}\sum_{i=1}^{m_1}\sum_{k=1}^{m_2} DM_{1,i}(k)^{\gamma} w_1(k) + \frac{1}{m_2}\sum_{i=1}^{m_2}\sum_{k=1}^{m_1} DM_{2,i}(k)^{\gamma} w_2(k). \tag{18}$$

Hall and Tajvidi use permutation testing to obtain quantiles for the test statistic, conditional on the set of pairwise distances in $\mathscr{Y}$. In a simulation study, they determine that the simplest case of $\gamma = 1$ and constant weight functions can be used.

In the two-sample context, NN directed edges, contrasted with MNBM undirected edges, present additional opportunities for constructing a cross-match test. Extracting a MNBM results in two types of undirected edges: within-group $(F_1 \to F_1)$ and across-group $(F_1 \to F_2$ or $F_2 \to F_1)$. Extracting a NN subgraph results in within-group directed edges $(F_1 \to F_1)$ as well as two distinct types of across-group directed edges: $F_1 \to F_2$ and $F_2 \to F_1$. This additional information, obtained with relatively low computation time, may result in cross-match tests that compare favorably with Rosenbaum's test in terms of power and computational efficiency. Vaidya (1989) demonstrates finding the exact kNN of each of N observations in $O(kN \log N)$ computation time. Approximate algorithms provide a kNN computation time of $O((N+k)\, p \log N)$, which is a reduction unless $p > kN/(N + k)$ (Arya, Mount, Netanyahu, Silverman, & Wu, 1998).

7. Change Detection in Higher Dimensions

Graph-theoretic tests are subject to the curse of dimensionality (Bellman, 1961, see Appendix A). The sparseness of multivariate samples grows in higher dimensions, which reduces the power of Friedman and Rafsky's MST test and Rosenbaum's cross-match test. Chen and Friedman (2015) derive additional statistics from the MST and MNBM to compensate for this aspect of the curse of dimensionality. Ruth (2009) finds a

reduction in the change-detection power of the ESPM statistic when moving from five-dimensional to 20-dimensional observations. Beyer, Goldstein, Ramakrishnan, and Shaft (1999) discuss the utility of NN as dimensionality increases. They find that for sample sizes of one million observations, the distance to the NN approaches the distance to the farthest neighbor in as few as 10 to 15 dimensions.

Table 2. Graph-theory terminology.

Term	Definition
graph	a vertex set, an edge set, and a relationship between vertices and edges
neighbors; adjacent	two vertices connected by an edge
vertex degree	the number of incident edges
regular graph	a graph in which all vertices have the same degree
directed graph	a graph with directed edges, specifying an ordered pair of vertices for each edge: (tail, head)
vertex *i* outdegree	the number of edges with tail *i*
vertex *i* indegree	the number of edges with head *i*
edge weight	a nonnegative measure of distance or cost assigned to each edge
complete graph	a graph with edges between all possible vertex pairs
subgraph	subset of a graph; the original graph contains the subgraph
spanning	including all vertices of the original graph
factor	spanning subgraph
k-factor	spanning *k*-regular subgraph
minimum spanning subgraph	spanning subgraph that minimizes the sum of the included edge weights
disjoint	subgraphs having no vertices in common
edge-disjoint	subgraphs having no edges in common
ensemble	two or more edge-disjoint subgraphs with identical vertex sets
matching	a set of edges with no shared endpoints
perfect matching	a matching incident to every vertex
nearest neighbor subgraph	a subgraph including only those edges connecting each observation to its nearest (minimum edge weight) neighbor.
path	a graph whose vertices are adjacent if and only if they are consecutive in an ordered list
Hamiltonian path	a spanning path
cycle	a closed path where the first and last vertices are the same
connected graph	a graph containing at least one path from each vertex to every other vertex
tree	an acyclic connected graph
spanning tree	a spanning subgraph that is a tree
minimum spanning tree	the spanning tree that minimizes the sum of the included edge weights

III. NEW APPROACHES TO GRAPH-THEORETIC CHANGE DETECTION

In this chapter, we introduce new statistical tests for change detection based on the subgraphs introduced in Chapter II. We also consider the effects of varying the match degree, how best to score edges in the subgraph, and how to obtain a *p*-value for testing purposes. Finally, using simulated data, we estimate the change-detection power of our proposed test statistics, comparing them to parametric and nonparametric alternatives.

A. MOTIVATION

In Chapter II, we establish that graph-theoretic methods have the ability to detect change, but that the most powerful methods (as demonstrated by Ruth, 2009 and Ruth & Koyak, 2011) are also computationally intensive. For example, Ruth (2009) found the Ensemble Sum of Pair-Maxima (ESPM) test statistic to have an empirically observed computation time on the order of N^4. We ask the questions: instead of recursively forming k edge-disjoint matchings, do we maintain power and reduce computational expense by directly forming a minimum graph factor of degree k, and deriving a single statistic for that? If we are interested in computational ease, does a statistic based on a k-nearest-neighbors (kNN) subgraph provide power comparable to a k-factor?

B. ALTERNATIVE MULTIPLE-MATCHING SUBGRAPHS

We propose and evaluate two new families of graph-theoretic statistics based on the undirected edges of the minimum k-factor, where k is the common degree of the vertices, and the directed edges of the kNN subgraph.

1. Minimum *k*-Factor

Consider the number of observations N to be even. As established in Chapter II Section D, at least $N/2$ edge-disjoint non-bipartite matchings may be recursively obtained from a complete graph. Instead of serial extraction of an ensemble, we proceed directly to a minimum k-factor. For $k = N/2$, this subgraph is denoted the minimum *half factor*, based on the minimum-distance pairing for the data. The distance between observations i

and j, using Euclidean, Mahalanobis, or another distance metric, is represented as d_{ij}. Using the set of binary edge indicator variables x_{ij}, we obtain the general minimum k-factor adjacency matrix (**X**) by solving an integer linear program:

$$\begin{aligned} \min_{\mathbf{x}} & \sum_{\{i,j\}\in E} d_{ij}x_{ij} \\ \text{s.t.} \quad & \sum_{j\in V} x_{ij} = k, \ \ \forall i \in V \\ & x_{ij} \in \{0,1\}, \ \forall \{i,j\} \in E. \end{aligned} \tag{19}$$

The edge cardinality of the minimum k-factor is $Nk/2$. Figure 5 shows the minimum half factor for the example data sets from Chapter II Section A. The change case is distinguished by an uneven concentration of edges compared to the no-change case. Table 3 lists the 10 matches for each observation. The no-change case has 56 of its 100 edges matching vertices from different samples, while the change case has 16 such edges.

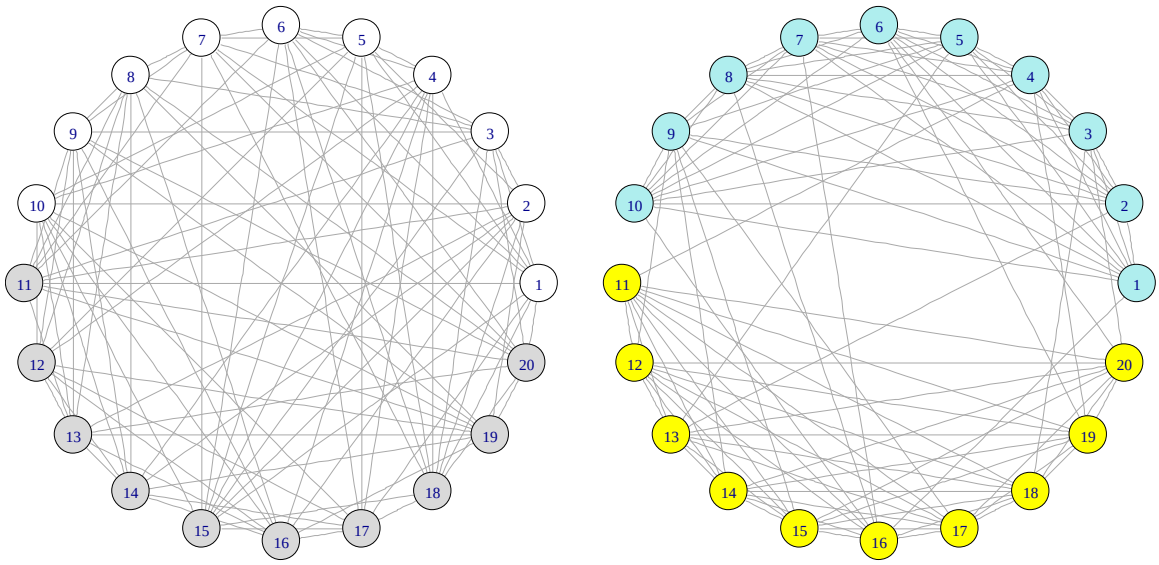

Figure 5. Minimum half factor for example data sets, plotted in a circular layout for clarity. The no-change case is to the left; the change case is to the right.

Table 3. Minimum half factor matches by vertex (i) for example data sets. The no-change case is to the left; the change case is to the right.

i	Vertices Matched to i									
1	3	4	5	6	7	11	15	17	18	20
2	3	5	10	11	13	14	15	16	18	19
3	1	2	6	7	8	9	11	15	18	20
4	1	5	6	10	12	14	15	16	17	18
5	1	2	4	6	7	10	12	15	17	18
6	1	3	4	5	7	11	15	17	18	20
7	1	3	5	6	8	9	11	15	18	20
8	3	7	9	11	12	13	14	16	19	20
9	3	7	8	11	12	13	14	16	19	20
10	2	4	5	12	13	14	15	16	17	19
11	1	2	3	6	7	8	9	13	19	20
12	4	5	8	9	10	13	14	16	17	19
13	2	8	9	10	11	12	14	16	19	20
14	2	4	8	9	10	12	13	16	17	19
15	1	2	3	4	5	6	7	10	17	18
16	2	4	8	9	10	12	13	14	17	19
17	1	4	5	6	10	12	14	15	16	18
18	1	2	3	4	5	6	7	15	17	20
19	2	8	9	10	11	12	13	14	16	20
20	1	3	6	7	8	9	11	13	18	19

i	Vertices Matched to i									
1	2	3	4	5	6	7	8	9	10	16
2	1	3	4	5	6	7	8	9	10	13
3	1	2	4	5	6	7	8	10	18	20
4	1	2	3	5	6	7	8	10	11	19
5	1	2	3	4	6	7	8	9	10	13
6	1	2	3	4	5	7	8	10	19	20
7	1	2	3	4	5	6	8	9	10	16
8	1	2	3	4	5	6	7	9	10	16
9	1	2	5	7	8	10	12	14	15	17
10	1	2	3	4	5	6	7	8	9	16
11	4	12	13	14	15	16	17	18	19	20
12	9	11	13	14	15	16	17	18	19	20
13	2	5	11	12	14	15	17	18	19	20
14	9	11	12	13	15	16	17	18	19	20
15	9	11	12	13	14	16	17	18	19	20
16	1	7	8	10	11	12	14	15	17	18
17	9	11	12	13	14	15	16	18	19	20
18	3	11	12	13	14	15	16	17	19	20
19	4	6	11	12	13	14	15	17	18	20
20	3	6	11	12	13	14	15	17	18	19

2. Multiple Nearest Neighbors

In addition to statistics based on a minimum k-factor, we also explore statistics based on a nearest neighbor (NN) subgraph. NN subgraphs have two advantages over minimum k-factors. First, there is additional information to exploit as the edges are directed, and second, finding NN is known to have lower computational costs compared to minimum k-factors as discussed earlier. We define the kNN edge set as the directed edges from the entire edge set E that are indicated by

$$x_{ij} = \begin{cases} 1, & \text{if } j \text{ is a } k\text{-nearest neighbor of } i \\ 0, & \text{otherwise} \end{cases}, \quad \forall (i, j) \in E. \tag{20}$$

The edge cardinality of the kNN subgraph is Nk. Figure 6 shows the $k = 10$ NN graph for the bivariate example from Chapter II Section A. Again, the difference between the change and no-change cases is apparent, but not as sharp as for the minimum half factor. Table 4 lists the 10 NN for each observation.

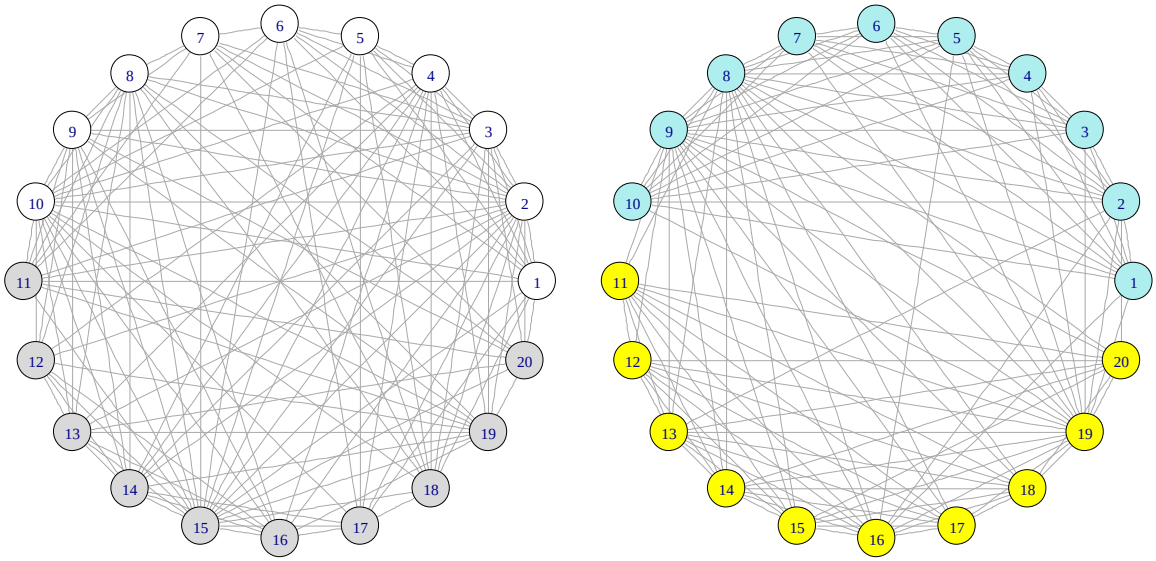

Figure 6. Nearest-neighbors graphs with $k = 10$ for example data sets, plotted in a circular layout for clarity. The no-change case is to the left; the change case is to the right.

Table 4. The 10-nearest neighbors of each vertex (i) for example data sets. The no-change case is to the left; the change case is to the right.

i	10 Nearest Neighbors of i									
1	2	3	4	5	6	10	11	15	17	18
2	3	4	8	10	13	14	15	16	18	19
3	1	2	4	8	9	10	11	15	18	19
4	2	5	10	13	14	15	16	17	18	19
5	1	2	3	4	6	10	14	15	17	18
6	1	2	3	4	5	10	11	15	17	18
7	1	2	3	6	8	9	11	15	18	20
8	2	3	9	10	11	13	14	15	16	19
9	2	3	8	10	11	13	14	15	19	20
10	2	3	4	8	13	14	15	16	18	19
11	1	2	3	7	8	9	10	15	19	20
12	2	4	8	9	10	13	14	15	16	19
13	2	4	8	9	10	12	14	15	16	19
14	2	3	4	8	10	12	13	15	16	19
15	1	2	3	4	6	8	10	17	18	19
16	2	4	8	9	10	12	13	14	15	19
17	1	2	4	5	6	10	14	15	16	18
18	1	2	3	4	5	6	8	10	15	17
19	2	3	4	8	9	10	13	14	15	16
20	1	2	3	7	8	9	11	13	15	19

i	10 Nearest Neighbors of i									
1	2	3	4	5	7	8	9	10	16	19
2	1	3	4	5	7	8	9	10	16	19
3	1	2	4	5	6	7	8	9	10	19
4	1	2	3	5	6	7	8	9	10	19
5	1	2	3	4	7	8	9	10	16	19
6	1	2	3	4	5	7	8	9	10	19
7	1	2	3	4	5	6	8	9	10	19
8	1	2	4	5	7	9	10	13	16	19
9	1	2	5	7	8	11	13	16	18	19
10	1	2	3	4	5	6	7	8	9	19
11	8	9	12	13	14	15	16	18	19	20
12	9	11	13	14	15	16	17	18	19	20
13	2	8	9	11	12	14	15	16	18	19
14	8	9	11	12	13	15	16	17	18	19
15	8	9	11	12	13	14	16	17	18	19
16	1	2	5	8	9	11	13	14	18	19
17	8	9	11	12	13	14	15	16	18	19
18	8	9	11	12	13	14	15	16	19	20
19	1	2	5	8	9	11	12	13	16	18
20	2	8	9	11	12	13	14	15	18	19

The no-change case has 111 of its 200 edges matching vertices from different samples, while the change case has 46 such edges. Centrally located observations, such as 2 and 15 in the no-change case, are 10-NN of more than ten other observations; this high NN indegree is not indicative of a violation of homogeneity.

3. Computation Time Considerations

Before developing statistics, we establish the relative computation effort for generating minimum subgraphs. For comparison, our computations were completed on a 64-bit Dell OptiPlex 990 Windows 7 platform with an Intel Core i5-2400 3.1 GHz processor, 8 GB RAM, using R (R Core Team, 2013) version 3.0.1 with packages cplexAPI (Gelius-Dietrich, 2013) and FNN (Beygelzimer, Kakadet, Langford, Arya, Mount, & Li, 2013) and using CPLEX version 12.5.

It is computationally less expensive to directly generate a minimum k-factor than to sequentially generate k edge-disjoint minimum non-bipartite matchings (kMNBM). The Blossom algorithm (Kolmogorov, 2009) is more efficient at extracting a single MNBM than using a general-purpose integer linear program solver such as CPLEX. As the value of k increases, the CPLEX solver is much more effective at obtaining the minimum k-factor than recursive use of the Blossom algorithm. We demonstrate this in Figure 7. The times are the medians observed from 100 simulations at each value of $k \in \{1,2,3,4,5,10,25,50,\ldots,250\}$, in each instance using 500, 10-dimensional independent standard normal observations. The CPLEX computation times are relatively constant with respect to k, while the computation times of recursive use of the Blossom algorithm increase in an almost linear manner. Although this is based on a sample size of 500, we have observed essentially the same result using different values of N.

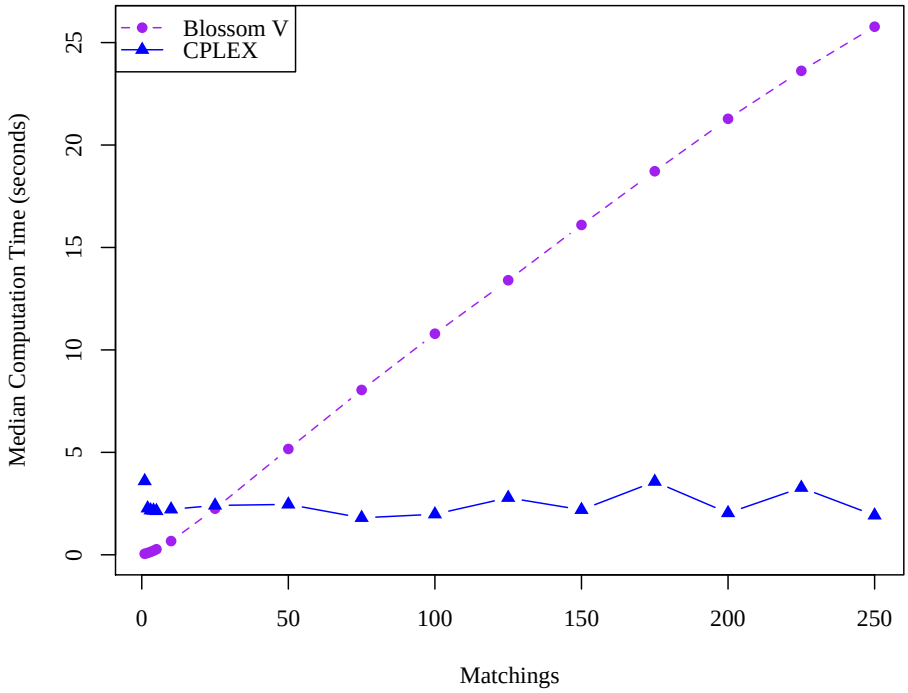

Figure 7. Computation time to obtain a minimum k-factor for $N = 500$, 10-dimensional observations sampled from a multivariate normal distribution, for the CPLEX algorithm and the recursive (edge-disjoint) Blossom V algorithm. The times are medians observed from 100 simulations at each value of k, $k \in \{1,2,3,4,5,10,25,50,\ldots,250\}$.

For the datasets considered, we empirically observe that the computation time for an edge-disjoint ensemble (kMNBM) is on the order of N^4, and that the computation times for minimum half factor and kNN subgraphs are both on the order of $N^{2.5}$. Figure 8 shows the median computation times required to extract the minimum subgraph for different sample sizes, plotted on logarithmic scales. Again, we use 100 simulations of 10-dimensional data at each dataset size. Although the linear regression lines for the minimum half factor and kNN series are nearly parallel, the ratio of their computation times is about 40, revealing a substantial computational advantage in favor of kNN. For example, with N = 1,000 observations and $k = N/2$, solving for the minimum half factor takes about 12 seconds while solving for the kNN takes about 0.25 seconds. In contrast, the kMNBM takes almost 11 minutes to solve.

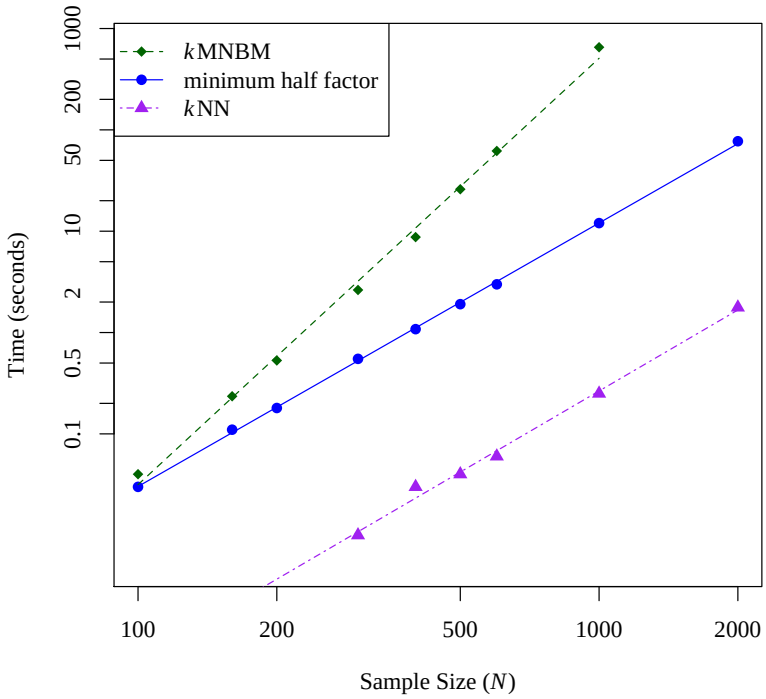

Figure 8. Comparison of median computation times for *k*MNBM, minimum half factor and the *k*NN vs. sample size, plotted on a log-log scale with fitted regression lines, for $k = N/2$, with 100 simulations per sample size. The regression line slopes for the *k*MNBM, minimum half factor and *k*NN are 4.21, 2.60, and 2.66, respectively.

Table 5 shows in more detail the computational effort for the *k*MNBM, minimum half factor, and *k*NN for $k = N/2$ as N increases. The times in Table 5 are the quantiles observed from 100 simulations using 10-dimensional data. The branch-and-bound exact algorithm for the minimum half factor leads to considerable variability in computational effort.

Table 5. Quantiles of observed computation times (seconds) for k edge-disjoint minimum non-bipartite matchings (kMNBM), minimum k-factor (kF), and k-nearest neighbors (kNN) for $k = N/2$ as data size (N) varies, observed in 100 simulations of 10-dimensional data for each value of N.

Data Size (N)	Quantiles of observed computation times (seconds)																				
	Maximum			0.90			0.75			0.50 (Median)			0.25			0.10			Minimum		
	*k*MNBM	*k*F	*k*NN	*k*MNBM	*k*F	*k*NN	*k*MNBM	*k*F	*k*NN	*k*MNBM	*k*F	*k*NN	*k*MNBM	*k*F	*k*NN	*k*MNBM	*k*F	*k*NN	*k*MNBM	*k*F	*k*NN
100	0.07	0.13	0.02	0.05	0.08	*	0.05	0.07	*	0.04	0.03	*	0.04	0.03	*	0.03	0.03	*	0.03	0.03	*
160	0.33	0.23	0.02	0.25	0.21	0.02	0.25	0.20	*	0.24	0.11	*	0.23	0.09	*	0.22	0.09	*	0.20	0.09	*
200	0.63	0.36	0.02	0.59	0.33	0.02	0.57	0.22	0.01	0.55	0.18	*	0.53	0.17	*	0.51	0.17	*	0.49	0.15	*
300	2.86	0.97	0.02	2.77	0.91	0.02	2.73	0.84	0.02	2.63	0.55	0.01	2.56	0.50	0.01	2.48	0.48	*	2.35	0.45	*
400	9.88	1.94	0.04	9.25	1.76	0.03	8.98	1.63	0.03	8.75	1.08	0.03	8.46	0.99	0.02	8.26	0.96	0.01	7.72	0.89	0.01
500	28.38	3.26	0.06	26.99	3.01	0.05	26.48	2.78	0.05	25.95	1.91	0.04	25.34	1.73	0.03	24.82	1.68	0.03	23.26	1.59	0.03
600	68.92	8.21	0.08	64.05	4.76	0.08	63.08	4.51	0.07	61.92	2.99	0.06	60.38	2.81	0.06	59.03	2.73	0.06	56.56	2.66	0.06
1,000	703.8	24.46	0.29	675.0	16.67	0.28	663.05	15.86	0.27	653.6	12.02	0.25	635.8	11.05	0.25	623.2	10.50	0.25	610.7	10.09	0.23
2,000	#	114.7	1.87	#	98.51	1.81	#	92.87	1.79	#	77.21	1.77	#	70.40	1.74	#	66.65	1.71	#	61.67	1.67

* Some kNN computations were less than one-hundredth of a second.
# Computations of kMNBM were not conducted for N = 2,000.

C. TEST STATISTICS: FORMATION AND PROPERTIES

In this section, we introduce test statistics based on the minimum *k*-factor and *k*NN subgraphs and discuss general properties. Our test statistics take the form of using information about the vertex labels of the minimum subgraph edges. While we derive unimodal statistics, they are not necessarily normally distributed.

1. General Form of Test Statistics

In Chapter II, we introduce Rosenbaum's cross-match test (Rosenbaum, 2005). Following extraction of the MNBM, the edges are scored according to the sample identity of their endpoints:

$$h(i,j) = \begin{cases} 1, & \text{if vertices } i \text{ and } j \text{ are from different samples} \\ 0, & \text{otherwise.} \end{cases} \tag{21}$$

In the case of directed graphs, (i, j) implies "from i to j." Other score functions may also be considered. In the sum of pair-maxima test of Ruth (2009), the score function used is $h(i, j) = \max\{i, j\}$. For a general score function $h(i, j)$ we define the following test statistic:

$$T = \frac{1}{2}\sum_{i=1}^{N}\sum_{j=1,\, j\neq i}^{N} h(i,j)\, x_{ij}. \tag{22}$$

2. Properties of Test Statistics

Properties of a test statistic taking the form of (22) depend on whether the minimum subgraph is a minimum *k*-factor (undirected edges) or a *k*NN subgraph (directed edges).

a. Minimum k-Factor Test Statistic Characteristics

For a minimum *k*-factor and unspecified score function, T has a known expected value and variance, which we derive here (R. A. Koyak, personal communication, January 8, 2014). Recognizing the symmetric nature of the adjacency matrix (**X**) for the minimum *k*-factor:

$$T = \frac{1}{2}\sum_{i=1}^{N}\sum_{j=1,\, j\neq i}^{N} h(i,j)\, x_{ij} = \sum_{i=2}^{N}\sum_{j=1}^{i-1} h(i,j)\, x_{ij}. \tag{23}$$

In addition to terms defined above, we let V_i represent the set of vertices that in the minimum *k*-factor are adjacent to vertex *i*. Now $E\left[x_{ij}\right] = \Pr\left[j \in V_i\right]$. By our formulation, $\sum_{j\in V} x_{ij} = k \;\; \forall i \in V,$ so then $\sum_{j\in V} \Pr\left[j \in V_i\right] = k.$ It follows that $\Pr\left[j \in V_i\right] = k/(N-1),$ $i \neq j.$ We now obtain the following expression:

$$\begin{aligned} \bar{h} &= \frac{2\sum_{i=2}^{N}\sum_{j=1}^{i-1} h(i,j)}{N(N-1)} \\ E[T] &= \frac{kN(N-1)\bar{h}}{2(N-1)} = \frac{kN\bar{h}}{2}. \end{aligned} \tag{24}$$

We expand terms in the variance as follows:

$$\begin{aligned} V[T] = V\left[\sum_{i=2}^{N}\sum_{j=1}^{i-1} h(i,j)\, x_{ij}\right] &= \sum_{i=2}^{N}\sum_{j=1}^{i-1} h^2(i,j)\, V\left[x_{ij}\right] + \sum_{r=3}^{N}\sum_{j=1}^{r-1}\sum_{\substack{s=1 \\ s\neq j}}^{r-1} h(r,j)\, h(r,s)\, Cov\left[x_{rj}, x_{rs}\right] \\ &+ \sum_{i=2}^{N}\sum_{\substack{r=2 \\ r\neq i}}^{N}\sum_{s=1}^{\min(i,r)-1} h(i,s)\, h(r,s)\, Cov\left[x_{is}, x_{rs}\right] + \sum_{i=3}^{N}\sum_{r=2}^{i-1}\sum_{s=1}^{r-1} h(i,r)\, h(r,s)\, Cov\left[x_{ir}, x_{rs}\right] \\ &+ \sum_{r=3}^{N}\sum_{s=2}^{r-1}\sum_{j=1}^{s-1} h(s,j)\, h(r,s)\, Cov\left[x_{sj}, x_{rs}\right] + \sum_{i=2}^{N}\sum_{j=1}^{i-1}\sum_{\substack{r=2 \\ r\neq i \\ r\neq j}}^{N}\sum_{\substack{s=1 \\ s\neq i \\ s\neq j}}^{r-1} h(i,j)\, h(r,s)\, Cov\left[x_{ij}, x_{rs}\right]. \end{aligned} \tag{25}$$

The variance of the edge indicators is expressed as follows:

$$V\left[x_{ij}\right] = \Pr\left[j \in V_i\right]\Pr\left[j \notin V_i\right] = \left(\frac{k}{N-1}\right)\left(1 - \frac{k}{N-1}\right), \;\; \forall \{i,j\} \in E. \tag{26}$$

Note that $V\left[x_{ij}\right]$ depends on neither *i* nor *j*; we instead use τ^2. Each of the next four terms of (25) involves shared vertices, and each covariance expression is the same:

$$
\begin{aligned}
E\left[x_{rj}x_{rs}\right] &= \Pr\left[j,s\in V_r \,\middle|\, j\neq s\right] = \left(\frac{k}{N-1}\right)\left(\frac{k-1}{N-2}\right) \\
Cov\left[x_{rj},x_{rs}\right] &= Cov\left[x_{is},x_{rs}\right] = Cov\left[x_{ir},x_{rs}\right] = Cov\left[x_{sj},x_{rs}\right] \\
&= \left(\frac{k}{N-1}\right)\left(\frac{k-1}{N-2}\right) - \left(\frac{k}{N-1}\right)^2 = -\frac{\tau^2}{N-2}.
\end{aligned} \tag{27}
$$

Let *i*, *j*, *r*, *s* denote distinct vertices. Now $E\left[x_{ij}x_{rs}\right] = \Pr\left[j\in V_i \,\middle|\, s\in V_r\right]\Pr\left[s\in V_r\right]$. We find $\Pr\left[j\in V_i \,\middle|\, s\in V_r\right]$ as follows:

$$
\begin{aligned}
k &= (N-1)\Pr\left[j\in V_i\right] \\
&= \Pr\left[s\in V_i \,\middle|\, s\in V_r\right] + \Pr\left[r\in V_i \,\middle|\, s\in V_r\right] + (N-3)\Pr\left[j\in V_i \,\middle|\, s\in V_r\right] \\
&= \frac{2k(N-1)}{N-2} + (N-3)\Pr\left[j\in V_i \,\middle|\, s\in V_r\right]; \\
\Pr\left[j\in V_i \,\middle|\, s\in V_r\right] &= \frac{(N-4)k+2}{(N-2)(N-3)}.
\end{aligned} \tag{28}
$$

This leads to our final covariance term:

$$
\begin{aligned}
E\left[x_{ij}x_{rs}\right] &= \frac{(N-4)k^2+2k}{(N-1)(N-2)(N-3)} \\
Cov\left[x_{ij},x_{rs}\right] &= \frac{(N-4)k^2+2k}{(N-1)(N-2)(N-3)} - \left(\frac{k}{N-1}\right)^2 = \frac{2\tau^2}{(N-2)(N-3)}.
\end{aligned} \tag{29}
$$

We now substitute the results of (27) and (29) into (25):

$$
V\left[T\right] = \tau^2\left(\begin{array}{l}
\displaystyle\sum_{i=2}^{N}\sum_{j=1}^{i-1} h^2(i,j) - \frac{1}{N-2}\sum_{i=2}^{N}\sum_{j=1}^{i-1}\sum_{r=2}^{N}\sum_{s=1}^{r-1} h(i,j)h(r,s) \\
\qquad\qquad \text{Card}\left(\{i,j\}\cap\{r,s\}\right)=1 \\
\displaystyle + \frac{2}{(N-2)(N-3)}\sum_{i=2}^{N}\sum_{j=1}^{i-1}\sum_{r=2}^{N}\sum_{s=1}^{r-1} h(i,j)h(r,s) \\
\qquad\qquad \text{Card}\left(\{i,j\}\cap\{r,s\}\right)=0
\end{array}\right). \tag{30}
$$

To simplify the second term, we define the following two sums:

$$H(\bullet, j) = \sum_{i=j+1}^{N} h(i, j), \quad 1 \le j \le N-1, \quad H(\bullet, N) \equiv 0$$
$$H(i, \bullet) = \sum_{j=1}^{i-1} h(i, j), \quad 2 \le i \le N, \qquad H(1, \bullet) \equiv 0. \tag{31}$$

We now obtain the result:

$$V[T] = \left(\frac{k}{N-1}\right)\left(1 - \frac{k}{N-1}\right) v_N^2,$$
$$\text{where } v_N^2 = \frac{N-1}{N-3} \sum_{i=2}^{N} \sum_{j=1}^{i-1} h^2(i, j) - \frac{N-1}{(N-2)(N-3)} \sum_{i=1}^{N} \left(H(i, \bullet) + H(\bullet, i)\right)^2 \tag{32}$$
$$+ \frac{2}{(N-2)(N-3)} \left(\sum_{i=2}^{N} \sum_{j=1}^{i-1} h(i, j) \right)^2.$$

Although we have exact expressions for the expected value and variance of *T*, there is no indication that *T* is approximately normally distributed in large samples. To assess the normality of *T*, we analyze the results of simulations both graphically and numerically. Using $h(i, j) = |i - j|$ (pair-differences) and choosing $k = N/2$, we simulate 100,000 instances of $N = 200$ five-dimensional multivariate normal observations under the null hypothesis.

Figure 9 presents a normal quantile-quantile plot for simulated values of *T* scored by pair-differences with $k = N/2$. The quantile-quantile plot has a concave down shape indicating left-skewness. A Shapiro-Wilk test for normality on a 500-observation sample of the statistic gives a *p*-value of 5.3×10^{-12}, evidence that *T* is unlikely to be approximately normally distributed, even in large samples.

Due to the evident skewness and more generally the lack of a normal distribution for *T*, it is inappropriate to determine critical values using a normal approximation with the known expected value and variance. We discuss methods to obtain a *p*-value for testing purposes in Section D of this chapter.

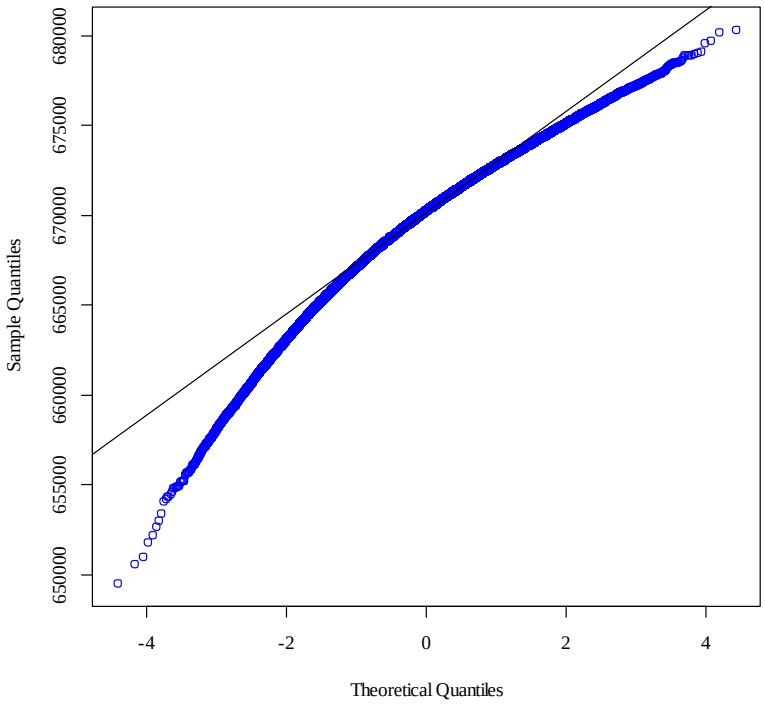

Figure 9. Normal quantile-quantile plot for statistic T based on 100,000 null hypothesis simulations of $N = 200$ five-dimensional multivariate normal observations.

b. Test Based on Nearest Neighbors

Our kNN test statistic has a known expected value. The expected value takes the same form as the minimum k-factor test statistic, from (24):

$$E[T] = \frac{kN(N-1)\bar{h}}{2(N-1)} = \frac{kN\bar{h}}{2}. \tag{33}$$

The variance has no neatly expressible closed form due to its dependence on the conditional probability of observation i being a NN of observation j, given that j is a NN of i. This conditional probability is in turn dependent on the unknown sampling distribution of the data under the null hypothesis. We discuss methods to obtain a p-value for testing purposes in Section D of this chapter.

3. Best Formulation for Test Statistics

We have three unanswered questions regarding our proposed statistics:

(1) Is there a particular score function that provides the best detection power?

(2) How can we exploit the directional information from *k*NN tests?

(3) Is there a particular value of *k* that provides the best detection power for tests using either the *k*-factor or *k*NN?

We use a variety of simulation scenarios to estimate change-detection power, similar to the scenarios of Ruth (2009). Detection power is the probability of rejecting the null hypothesis when the null hypothesis is false. The critical value for a test is based on the appropriate quantile from 100,000 null hypothesis simulations.

In all cases, we illustrate the power to detect change under scenarios in which $N = 200$ observations are taken from our sample space $\mathbb{R}^p$, where p is the observation dimension. For change scenarios, we place a change point in the middle of the observation sequence. Conducting 10,000 simulations per scenario, we estimate detection power as the fraction of times that the test correctly rejects the null hypothesis when it is false, using a test level of $\alpha = 0.05$. We also vary the following factors:

(1) Sampling distribution. We consider multivariate normal (F_{MVN}), multivariate normal mixture (F_{mix}), and multivariate Weibull (F_{Weib}) probability distribution families.

(2) Dimension. We analyze scenarios in two dimensions, $p = 5$ and $p = 20$.

(3) Change parameter. We examine changes in mean and scale.

(4) Type of change. We consider cases of abrupt (jump) and gradual (drift) change in the parameter undertaking change.

(5) Magnitude of change (Δ). We consider changes of varying magnitudes.

Our distributions are generated as follows. The multivariate distribution F_{MVN} is based on the density function:

$$f_{\text{MVN}}\left(\mathbf{y};\boldsymbol{\mu},\Sigma\right) = \frac{1}{\left(2\pi\right)^{N/2}\left|\Sigma\right|^{1/2}}\exp\left\{-\frac{1}{2}\left(\mathbf{y}-\boldsymbol{\mu}\right)'\Sigma^{-1}\left(\mathbf{y}-\boldsymbol{\mu}\right)\right\} \quad \forall \mathbf{y} \in \mathbb{R}^p, \tag{34}$$

where $\boldsymbol{\mu} \in \mathbb{R}^p$ is the mean vector, and $\Sigma \in \mathbb{R}^{p \times p}$ is the covariance matrix. Under the null hypothesis, $\boldsymbol{\mu} = \mathbf{0}$ and $\Sigma = I_p$ (I_p is the $p \times p$ identity matrix) so that $\mathbf{y} \sim N_p\left(\mathbf{0}, I_p\right)$. For changes in mean, the first element of the mean vector μ_1 is a value $\Delta \neq 0$. For changes in scale, $\Sigma = \left(1+\Delta\right)^2 I_p$.

The multivariate normal mixture distribution F_{mix} is more outlier-prone than the multivariate normal distribution. It is formed by mixing two multivariate normal distributions $N_p\left(\mathbf{0}, I_p\right)$ and $N_p\left(\mathbf{0}, \psi^2 I_p\right)$, where $\psi > 1$. We take

$$\mathbf{Y} = \left[1 + \left(\psi - 1\right) B\right] \mathbf{Z}, \tag{35}$$

where B is a Bernoulli random variable with success probability q. For our simulations, $\psi = 4$ and $q = 0.1$. For changes in mean, the first element of the mean vector is equal to Δ; for changes in scale, we substitute $\left(1+\Delta\right)^2 I_p$ for I_p in the covariance matrix of the two multivariate normal distributions we are mixing.

We define F_{Weib} to be the distribution on $\mathbb{R}^p$ associated with p independent identically distributed univariate Weibull random variables:

$$F_{\text{Weib}}\left(\mathbf{y}; \eta, \boldsymbol{\beta}\right) = \begin{cases} \displaystyle\prod_{i=1}^{p} \left[1 - \exp\left\{-\left(\frac{y_i}{\beta_i}\right)^{\eta}\right\}\right], & \mathbf{y} \in \mathbb{R}_+^p, \\ 0, \quad \text{otherwise} \end{cases} \tag{36}$$

where η and β_i are the univariate Weibull shape and scale parameters, respectively, $\mathbf{y} = \left(y_1, \ldots, y_p\right)$, and $\boldsymbol{\beta} = \left(\beta_1, \ldots, \beta_p\right)$. We set $\eta = 1.5$, $\beta_i = 1 + \Delta$, and $\beta_j = 1$, $j \in \left\{2, \ldots, p\right\}$.

Our change magnitudes (Δ) range from 0 to 1.0 in increments of 0.25. The fraction of simulations rejecting the null hypothesis when the null hypothesis is true ($\Delta = 0$) is representative of the false alarm rate for the test. Our abrupt changes are of magnitude Δ and occur at observation τ. The drift changes begin at observation τ, increasing linearly until reaching the full value of Δ at observation N. For changes in

distribution mean and changes in the scale of the multivariate Weibull distribution, we simulate the change in only the first component of each observation. For changes in scale for the multivariate normal and multivariate normal mixture distributions, we simulate the change in all components.

a. Selection of Score Function

Is there a particular score function that provides the best detection power? We consider the sum of pair-differences score function as discussed above. To magnify or suppress large label differences, raising the pair-differences to a power (other than 1) and summing the result produces a family of possible score functions, in the manner of a Box-Cox transformation (Box & Cox, 1964). We investigate the following score functions with the minimum *k*-factor and the *k*NN:

$$h(i,j) = \begin{cases} \dfrac{|i-j|^{\lambda}-1}{\lambda}, & (\lambda \neq 0) \\ \log|i-j|, & (\lambda = 0) \end{cases}, \quad \forall (i,j) \in E, \tag{37}$$

for $\lambda \in \{-2.0, -1.5, -1.0, -0.9, \ldots, 0.9, 1.0, 1.5, 2.0\}$.

We set $k = N/2$ and evaluate the change-detection power of the resulting subgraphs scored by the transformation of (37). For each sampling distribution and dimension, we conduct 100,000 null hypothesis simulations of the test statistic under the null hypothesis, using the 0.05 quantile of the simulation values as a critical value for an $\alpha = 0.05$-level test. For each sampling distribution, dimension, level of change magnitude (Δ), and family parameter (λ), we conduct 10,000 simulations of $N = 200$ observations with a jump in the first mean vector component at observation $\tau = 101$. As shown in Table 6, for all scenarios examined, the value of $\lambda = 1$ provides at or near the maximum detection power for the minimum half factor. We naturally prefer no transformation, and in any case prefer to use an interpretable value (Faraway, 2005). There is a narrower interval for the *k*NN-based test; the 95 percent confidence limits for the score function parameter providing the maximum power in the vignettes examined intersect at [−0.4, −0.3]. We choose $\lambda = -0.3$, as the value −0.4 is at the lower end of several vignette

confidence intervals. The negative parameter value indicates that there is better *k*NN change-detection performance when the label differences are suppressed than otherwise.

Table 6. Score function parameter (λ) 95 percent confidence interval providing the maximum detection power at the 0.05 level of test. Based on 10,000 simulations per vignette of a jump change of magnitude Δ in the first component of the mean vector of $N = 200$ p-dimensional simulated observations sampled from the specified distribution.

		k-factor ($k = 100$) 95% Confidence Interval for λ			*k*NN ($k = 100$) 95% Confidence Interval for λ		
p	Δ	F_{MVN}	F_{mix}	F_{Weib}	F_{MVN}	F_{mix}	F_{Weib}
5	0.25	[0.2, 2.0]	[0.0, 2.0]	[0.3, 2.0]	[−0.7, 0.7]	[−0.9, 0.3]	[−0.5, 0.3]
	0.50	[0.4, 1.5]	[0.4, 1.5]	[0.3, 1.5]	[−0.4, 0.1]	[−0.7, 0.0]	[−0.4, 0.0]
	0.75	[0.3, 1.0]	[0.4, 1.5]	[0.2, 1.5]	[−0.4, 0.0]	[−0.6, −0.2]	[−0.4, 0.0]
	1.00	[0.0, 1.5]	[0.3, 1.5]	[−0.1, 2.0]	[−0.4, 0.1]	[−0.5, −0.2]	[−0.5, 0.0]
20	0.25	[0.0, 2.0]	[−0.4, 2.0]	[0.2, 2.0]	[−1.0, 2.0]	[−2.0, 2.0]	[−0.7, 0.3]
	0.50	[0.4, 2.0]	[0.3, 2.0]	[0.4, 1.5]	[−0.5, 0.2]	[−1.0, 0.7]	[−0.5, 0.0]
	0.75	[0.5, 1.5]	[0.5, 1.5]	[0.4, 1.5]	[−0.4, 0.1]	[−1.0, −0.2]	[−0.5, −0.1]
	1.00	[0.4, 1.5]	[0.5, 1.5]	[0.1, 1.5]	[−0.4, 0.0]	[−0.9, −0.3]	[−0.5, −0.1]

Figure 10 shows the typical responses when evaluating the score function with the *k*-factor and with the *k*NN, with 95 percent confidence bounds on the detection power. The detection power shown is for the second row of Table 6: a five-dimensional multivariate normal distribution with an abrupt change of magnitude 0.5 in the first mean vector component at observation 101.

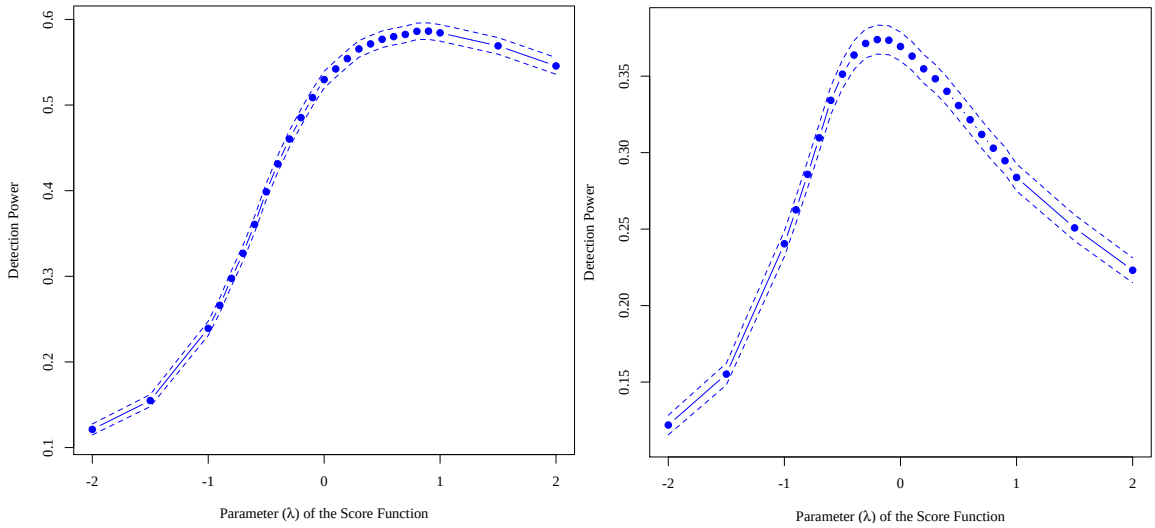

Figure 10. Estimated detection power as a function of the score function parameter λ, for the $N = 200$ multivariate normal observations with an abrupt change of magnitude 0.5 in the first component of the mean vector at observation $\tau = 101$, with 10,000 simulations per value of λ with a 0.05 test level. The minimum half factor test is to the left, the *k*NN ($k = 100$) test is to the right. The dashed lines are the 95 percent confidence limits for the detection power.

b. Use of Directional Information

We now consider the use of the directed edges of the *k*NN subgraph to construct a test statistic that is sensitive to distributional change. Our second statistic takes advantage of the directed nature of *k*NN subgraph edges.

We obtain our statistic as follows. We take the count of those *k*NN whose index is less than the current index, for all indices. For ease of discussion, a *prior edge* is an edge of the minimum subgraph whose destination observation label is less than the source observation label. An equivalent score function is $h(i, j) = I\{i > j\}$, where $I\{\}$ is the indicator function such that:

$$I\{A\} = \begin{cases} 1, & \text{if } A \text{ is true,} \\ 0, & \text{if } A \text{ is not true.} \end{cases} \tag{38}$$

We call this statistic the NN prior-edge count (K):

$$K = \frac{1}{2}\sum_{i=1}^{N}\sum_{j=1,\, j\neq i}^{N} I\{i > j\}\, x_{ij} = \frac{1}{2}\sum_{i=2}^{N}\sum_{j=1}^{i-1} x_{ij}. \tag{39}$$

The *k*NN subgraph edge cardinality is *kN*, and under the null hypothesis, *K* is halving a count of half of the edges.

$$E[K] = \frac{kN}{4}. \tag{40}$$

As explained earlier with *T* based on *k*NN, there is no neatly expressible closed form for the variance of *K*.

We now have two statistics based on the *k*NN subgraph. Table 7 shows that under the alternative hypothesis, the attained change-detection powers of NN statistics *T* and *K* vary greatly for the multivariate normal case. For illustration, we use $N = 200$ and $k = 100$ for the *k*NN subgraph, and use *T* of (22).

Table 7. Estimated power of nearest-neighbor statistics *T* and *K* to detect a change of magnitude Δ at observation 101 of 200 drawn from a five-dimensional multivariate normal sampling distribution, with 10,000 simulations per vignette. Jump in the first component of the mean vector is to the left; jump in all components of the scale is to the right.

F_{MVN}; $p = 5$	Mean Jump		Scale Jump	
Δ	*T*	*K*	*T*	*K*
0.00	0.05	0.05	0.05	0.05
0.25	0.11	0.05	0.22	0.99
0.50	0.38	0.05	0.82	1.00
0.75	0.80	0.05	1.00	1.00
1.00	0.98	0.05	1.00	1.00

The *T* statistic shows power in detecting both location and scale changes. The *K* statistic is seen to be effective for detecting scale changes, while exhibiting no power to detect location changes. Combining *T* and *K* allows for both location and scale detection. We call this combined test the composite *k*NN test. To maintain an overall test level of $\alpha = 0.05$, we use the Bonferroni adjustment based on Boole's inequality (Seneta, n.d.). Setting a test level of $\alpha = 0.025$ for each of the two tests (by using the 0.025 quantile or

by multiplying the minimum obtained *p*-value by two and comparing the product to 0.05) conservatively allows for maintaining a simultaneous test level of $\alpha = 0.05$. The composite *k*NN test decision rule is as follows:

$$\text{Test Decision} = \begin{cases} \text{reject } H_0, \text{ if either } T \text{ or } K \text{ reject at a } \alpha = 0.025 \text{ level of test;} \\ \text{fail to reject } H_0, \text{ otherwise.} \end{cases} \quad (41)$$

Applying the composite *k*NN test to the simulation scenario of Table 7 results in the change-detection powers determined from simulation shown in Table 8. In Section E of this chapter, the composite *k*NN test and the *k*-factor sum of pair-differences test are compared.

Table 8. Simulation calculations of the power of the composite *k*NN test to detect a change of magnitude Δ at observation 101 of 200 drawn from a five-dimensional multivariate normal sampling distribution, with 10,000 simulations per vignette. Jump in the first component of the mean vector is to the left; jump in all components of the scale is to the right.

F_{MVN}; p = 5	Mean Jump	Scale Jump
Δ		
0.00	0.05	0.05
0.25	0.09	0.98
0.50	0.30	1.00
0.75	0.73	1.00
1.00	0.97	1.00

c. Selection of k

For statistics derived from either the minimum *k*-factor or the *k*NN subgraphs, based on $k = N/2$, the k is motivated by the desire to include a rich subgraph and to provide a test comparable in power to the Ensemble Sum of Pair-Maxima test proposed by Ruth (2009). There is justification for considering different k values, which we now explore. The detection power of our statistics may depend on how many matches or neighbors included in the subgraph. We examine the detection power as a function of k and recommend values of k. We start by examining the behavior of the ESPM test statistic.

(1) Ensemble Sum of Pair-Maxima Behavior as *k* Increases

The ESPM test statistic is the maximum of a process considering all values of *k* up to $k = N/2$. The process is constructed such that the null hypothesis expected value of the statistic is zero. Any critical value exceedance occurring from $k = 1$ to $k = N/2$ results in change detection (rejection of the null hypothesis). In many cases, however, it was observed that exceedance of the critical value at any point in the sequence usually was met with exceedance at the $k = N/2$ point as well. Understandably, using the $N/2$-point of the sequence should result in a loss of power, which we suspect will be minimal under most circumstances. Figure 11 shows the ESPM responses as the number of ensemble matches increases for two of the change detection simulation scenarios described earlier: an abrupt jump change in the first component of the mean vector, and a gradual drift change in scale in all five dimensions, respectively, of 200 five-dimensional observations drawn from a multivariate normal distribution. Figures depicting ESPM response in the multivariate normal null hypothesis case, mean drift case, and scale jump case are given in Appendix B.

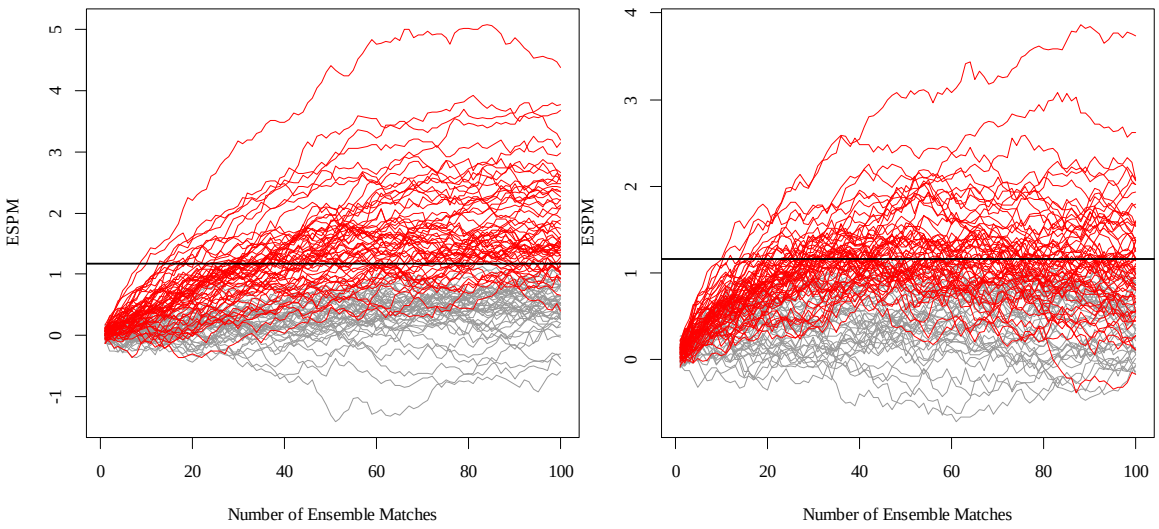

Figure 11. The Ensemble Sum of Pair-Maxima test statistic as the number of ensemble matches increases from 1 to $N/2$ for two change detection scenarios, with 100 simulations per scenario. To the left, a 0.5 magnitude jump inserted at observation $\tau = 101$ in the first component of the mean vector of a five-dimensional multivariate normal distribution. To the right, a 0.5 magnitude drift change started at observation $\tau = 101$ in the standard deviation of a five-dimensional multivariate normal distribution. The thick black line is the critical value. Red traces indicate the ESPM statistic exceeding the critical value based on an $\alpha = 0.05$ level of test. Gray traces do not exceed the critical value.

For the change in the mean, shown on the left in Figure 11, although the ESPM traces appear to be generally increasing as k increases, the ESPM value first exceeds and then drops below the critical value in 11 of the 100 simulations. These cases show as the red lines that end below the thick black line. When we consider only the trace values at k = 100, the detection power falls from 0.57 to 0.46. The variance drift case, shown on the right in Figure 11, is even more pronounced. The ESPM value exceeds and then drops below the critical value in 33 of the 100 simulations. Considering only the trace values at k = 100, the detection power falls from 0.59 to 0.26.

(2) Detection Power as k Varies – Minimum k-Factor Sum of Pair-Differences

To evaluate the influence of k on the detection power of the sum of pair-differences (SPD) based on the minimum k-factor subgraph, we investigate values of k from $0.05N$ to $0.5N$ for N = 200. For some of the simulation scenarios outlined earlier in

this section, we determine the change-detection power of the minimum *k*-factor SPD. We present the results in Table 9, highlighting the maximum detection power for each case.

Table 9. Test power to detect a change based on pair-differences determined from a the minimum *k*-factor for a range of values of *k* on a subset of sampling distributions, dimensions, change parameters, change types, and change magnitudes, using change point $\tau = 101$ based on $N = 200$, $\alpha = 0.05$, and 10,000 simulations per vignette. Dimensions are $p = 5$ unless otherwise noted. The maximum powers for each case are shaded.

	Detection Power as *k* Varies – minimum *k*-factor									
Simulation Scenario	10	20	30	40	50	60	70	80	90	100
F_{MVN} mean jump 0.75	0.71	0.83	0.88	0.90	0.92	0.92	0.93	0.93	0.94	0.94
F_{MVN} mean drift 0.75	0.31	0.40	0.46	0.50	0.53	0.55	0.54	0.55	0.57	0.56
F_{MVN} mean jump 0.75 $p = 20$	0.44	0.57	0.61	0.65	0.70	0.71	0.73	0.75	0.72	0.74
F_{MVN} mean drift 0.75 $p = 20$	0.17	0.22	0.24	0.26	0.29	0.30	0.31	0.33	0.30	0.31
F_{MVN} scale jump 0.50	0.96	0.98	0.97	0.97	0.97	0.96	0.94	0.94	0.91	0.90
F_{MVN} scale drift 0.75	0.94	0.96	0.96	0.95	0.94	0.92	0.90	0.89	0.85	0.81
F_{mix} mean jump 0.75	0.50	0.62	0.68	0.73	0.75	0.77	0.79	0.79	0.80	0.80
F_{mix} mean drift 0.75	0.21	0.26	0.30	0.33	0.37	0.38	0.38	0.40	0.40	0.39
F_{mix} mean jump 0.75 $p = 20$	0.20	0.28	0.31	0.35	0.37	0.38	0.40	0.41	0.39	0.39
F_{mix} scale jump 0.50	0.79	0.87	0.89	0.90	0.89	0.89	0.87	0.86	0.83	0.80
F_{mix} scale drift 0.75	0.74	0.83	0.84	0.85	0.85	0.84	0.82	0.80	0.77	0.74
F_{Weib} β jump 0.75	0.82	0.93	0.96	0.96	0.97	0.97	0.97	0.97	0.97	0.96
F_{Weib} β drift 0.75	0.70	0.84	0.89	0.91	0.92	0.93	0.93	0.93	0.93	0.92
F_{Weib} β jump 0.75 $p = 20$	0.51	0.64	0.68	0.71	0.72	0.72	0.71	0.71	0.70	0.68

For the *k*-factor SPD, the value of *k* corresponding to the highest detection power is seen to depend on the change parameter (mean or scale) more than on the sampling distribution, change type (jump or drift), or dimension. The detection power for $k = N/2$ is at or near the highest detection power for changes in mean and changes in the β parameter of the multivariate Weibull (F_{Weib}). Cases with a change in scale see a eight to 15 percentage points drop in detection power from a peak occurring between $k = N/10$ to $k = N/5$. For the change of scale cases there is still considerable power at $k = N/2$, but it is not the peak power.

(3) Recommended Value for k – Minimum k-Factor

We do lose power in the scale change cases, but at $k = N/2$, the k-factor SPD is near enough to the maximum power in most of the multivariate normal, heavy-tailed, and skewed cases. The empirically observed computation time to obtain the minimum k-factor is approximately invariant with respect to k as shown in Section B of this chapter.

(4) Detection Power as k Varies – k-Nearest Neighbors

We similarly examine the composite kNN test, with statistics from (22) and (39) and the Bonferroni adjustment of (41). We present the results in Table 10, highlighting the maximum detection power for each case.

Table 10. Test power to detect a change based on the composite kNN test for a range of values of k on a subset of sampling distributions, dimensions, change parameters, types, and magnitudes, using change point $\tau = 101$ based on $N = 200$, $\alpha = 0.05$, and 10,000 simulations per vignette. Dimensions are $p = 5$ unless otherwise noted. The maximum powers for each case are shaded.

Detection Power as k Varies – kNN										
Simulation Scenario	10	20	30	40	50	60	70	80	90	100
F_{MVN} mean jump 0.75	0.41	0.54	0.61	0.65	0.68	0.69	0.71	0.71	0.73	0.73
F_{MVN} mean drift 0.75	0.16	0.21	0.24	0.27	0.28	0.29	0.30	0.31	0.30	0.31
F_{MVN} mean jump 0.75 $p = 20$	0.19	0.26	0.29	0.32	0.33	0.36	0.37	0.37	0.37	0.37
F_{MVN} mean drift 0.75 $p = 20$	0.10	0.12	0.12	0.13	0.13	0.14	0.14	0.14	0.15	0.15
F_{MVN} scale jump 0.25	0.92	0.96	0.96	0.96	0.96	0.96	0.97	0.97	0.97	0.98
F_{MVN} scale drift 0.25	0.67	0.71	0.73	0.72	0.74	0.75	0.76	0.77	0.79	0.80
F_{mix} mean jump 0.75	0.29	0.39	0.43	0.47	0.49	0.49	0.48	0.47	0.45	0.44
F_{mix} mean drift 0.75	0.12	0.16	0.17	0.19	0.19	0.19	0.19	0.19	0.18	0.18
F_{mix} mean jump 0.75 $p = 20$	0.12	0.13	0.13	0.14	0.13	0.13	0.11	0.12	0.10	0.10
F_{mix} scale jump 0.25	0.67	0.74	0.75	0.77	0.76	0.77	0.77	0.77	0.77	0.75
F_{mix} scale drift 0.25	0.40	0.46	0.47	0.47	0.48	0.50	0.50	0.51	0.51	0.50
F_{Weib} β jump 0.75	0.58	0.74	0.80	0.83	0.85	0.86	0.87	0.88	0.89	0.89
F_{Weib} β drift 0.75	0.33	0.42	0.47	0.51	0.53	0.56	0.56	0.60	0.61	0.62
F_{Weib} β jump 0.75 $p = 20$	0.41	0.52	0.59	0.63	0.65	0.66	0.68	0.68	0.68	0.67

For all of the multivariate normal (F_{MVN}) and multivariate Weibull (F_{Weib}) cases examined, as well as the multivariate normal mixture (F_{mix}) scale change cases, the

highest detection power is achieved at or near $k = N/2$. For the F_{mix} location change cases, the peak power is obtained when using fewer neighbors in the subgraph.

(5) Recommended Value for Composite *k*NN Test

We recommend the value $k = N/2$. There is evidence this is not the best choice for heavy-tailed distributions, but using $k = N/2$ produces a test that detects change reasonably well across the distributions and scenarios tested.

Our statistics now incorporate these explorations into the score function transformation, directional information, and value of *k*. A more thorough understanding of the interaction between the score function transformation and the value of *k* is left to future research.

D. OBTAINING A CRITICAL OR *p*-VALUE

In this section, we compare and contrast four methods of obtaining a critical or *p*-value for testing purposes. We consider accuracy and computational implications of the methods. To demonstrate concepts, we use the minimum half factor sum of pair-differences (SPD) statistic. The distributions of our statistics do not have closed form solutions, although the first two moments of the minimum half factor SPD statistic are available.

1. Critical Values from Probability Inequalities

Knowing the expected value and variance, we could generate a critical value at an appropriate level of test based on Chebyshev's inequality as shown in (42), but this is known to be overly conservative (Wackerly, Mendenhall, & Scheaffer, 2008).

$$Z = \frac{T - E[T]}{\sqrt{V[T]}}$$
$$\Pr[|Z| \geq \kappa] \leq \frac{1}{\kappa^2}. \tag{42}$$

We slightly sharpen this inequality to a one-sided Chebyshev bound, also known as Cantelli's inequality (Ion, 2001). The lower-tail version of the bound is shown in (43):

$$\Pr\left[Z \le -\kappa\right] \le \frac{1}{\kappa^2 + 1}. \tag{43}$$

If the distribution of Z is unimodal, which our experience through simulations suggests is a reasonable assumption for the minimum half factor SPD statistic, an even sharper inequality is possible (Ion, 2001; Bickel & Krieger, 1992):

$$\Pr\left[Z \le -\kappa\right] \le \frac{4}{9\left(\kappa^2 + 1\right)}, \quad \kappa \ge \sqrt{3}. \tag{44}$$

For example, at a 0.05 level of test, the one-sided Chebyshev inequality of (43) gives critical value $\kappa = -4.36$, while the unimodal bound of (44) gives $\kappa = -2.81$. Were our statistic known to be normally distributed, the bound for a 0.05 level of test would be the familiar $\kappa = -1.645$. Using the 100,000 null hypothesis simulations discussed in Section C of this chapter, the unimodal bound ($\kappa = -2.81$) results in an approximate level of test of 0.010. It also is possible to derive a conservative *p*-value for a test by allowing κ to be the standardized value of the observed test statistic. For the example data sets of Chapter II Section A, the half factor SPD test applied to the no-change case and the change case gives standardized values of $Z = -0.21$ and $Z = -8.07$, respectively. The no-change case fails to reject the null hypothesis, while the half factor SPD test for the change case provides sufficient evidence to reject the null hypothesis of no change.

2. Critical Values from Control Data

A second method to obtain a critical or *p*-value is the use of control data to establish the distribution of our statistic. By having ample control data, or by simulating enough null hypothesis scenarios (say, 100,000), we obtain an appropriate percentile of the desired statistic under the null hypothesis, or we can assess the strength of evidence represented by the observed value of the statistic against it.

There are at least one advantage and two drawbacks to estimating a critical value from data. The advantage is that, once it is established, it is used indefinitely to signal change, assuming future data follow the same distribution when not undergoing change. A significant drawback is the need for the control data, which are not always available.

Lacking control data, simulation is used to approximate the critical or *p*-value. This method also assumes that the null hypothesis is true for the control data. If the alternative hypothesis is true in our control data, the underlying graphical structure reflects that; we then attain an inaccurate critical value.

3. Critical or *p*-Values from Permutation Testing

In lieu of solving for the minimum half factor a large number of times, we use the principles outlined in Good (2005) to obtain a critical value. Under the null hypothesis, we solve for the minimum half factor once. We then permute the vertex labels, say 100,000 times, maintaining the underlying minimum subgraph structure. Computing the desired statistic for each permutation, we establish the null hypothesis permutation distribution of the test statistic. Finally, we select the desired quantile. This method requires just one extraction of the minimum subgraph, from just one set of control data.

In a similar manner we derive a permutation *p*-value. The observed value of the statistic is compared to the permutation distribution to obtain the *p*-value:

$$p\text{-value} = \frac{\#\{\text{permutation value} \leq \text{observed value}\}}{\#\text{ of permutations}}. \tag{45}$$

Consider the example data sets from Chapter II Section A. For the no-change case and the change case, respectively, the minimum half factor SPD test results in permutation *p*-values of 0.38 and 0.00, and the composite *k*NN test results in permutation *p*-values of 0.52 and 0.00. For both tests, the null hypothesis is not rejected in the no-change case, but is rejected in the change case.

4. Numerical Experiments

Using two sets of simulations, we use the same sampling distributions described in Section C of this chapter to compare the various methods of obtaining a *p*-value. Results of these experiments are summarized in Table 11. As expected, a test based on the unimodal probability inequality (44) is notably conservative. The quantiles obtained by the second and third methods are nearly identical. These findings suggest that we may

forgo many minimum subgraph calculations and instead use one minimum subgraph, permuting the observation labels, as is usually done in a permutation test (Good, 2005).

Table 11. Quantiles for the standardized minimum half factor sum of pair-differences at the $\alpha = 0.05$ level of test, determined by 100,000 null hypothesis simulations (Simulation), and determined by a single null hypothesis simulation and 100,000 permutations of the observation labels (Permutation), for $N = 200$ simulated null hypothesis observations. The null hypothesis minimum half factor sum of pair-differences is not normally distributed.

Dimension	Method	F_{MVN} 0.05 quantile	F_{mix} 0.05 quantile	F_{Weib} 0.05 quantile
$p = 5$	Simulation	−1.790	−1.779	−1.798
	Permutation	−1.791	−1.774	−1.791
$p = 20$	Simulation	−1.719	−1.717	−1.713
	Permutation	−1.728	−1.720	−1.718

We compare the change-detection power of the minimum half factor sum of pair-differences using the above methods (using null hypothesis data to establish a critical value vs. using a permutation test to directly obtain the *p*-value) for a subset of the cases from Section C of this chapter. As shown in Tables 12 through 14, the change-detection power is virtually unchanged.

Table 12. Test power to detect a location change of magnitude Δ in the first component of the mean vector at observation 101 of 200 drawn from a multivariate normal sampling distribution, with 10,000 simulations per vignette. Dimension $p = 5$ case is to the left; dimension $p = 20$ case is to the right.

F_{MVN}; mean change	$p = 5$		$p = 20$	
Δ	HF CV	HF P	HF CV	HF P
0.00	0.05	0.05	0.05	0.05
0.25	0.16	0.16	0.10	0.09
0.50	0.58	0.56	0.32	0.33
0.75	0.94	0.94	0.73	0.74
1.00	1.00	1.00	0.97	0.97

HF: minimum Half Factor sum of pair-differences
CV: power obtained via comparison to critical value
P: power obtained via permuting index labels

Table 13. Test power to detect a change of magnitude Δ in the first component of the mean vector at observation 101 of 200 drawn from a multivariate normal mixture sampling distribution, with 10,000 simulations per vignette. Dimension $p = 5$ case is to the left; dimension $p = 20$ case is to the right.

F_{mix}; mean change	$p = 5$		$p = 20$	
Δ	HF CV	HF P	HF CV	HF P
0.00	0.05	0.05	0.05	0.05
0.25	0.12	0.14	0.07	0.06
0.50	0.42	0.40	0.17	0.19
0.75	0.80	0.81	0.40	0.38
1.00	0.97	0.98	0.71	0.68

HF: minimum Half Factor sum of pair-differences
CV: power obtained via comparison to critical value
P: power obtained via permuting index labels

Table 14. Test power to detect a change of magnitude Δ in the first component of the scale parameter β of observation 101 of 200 drawn from a multivariate Weibull sampling distribution, with 10,000 simulations per vignette. Dimension $p = 5$ case is to the left; dimension $p = 20$ case is to the right.

F_{Weib}; β change	$p = 5$		$p = 20$	
Δ	HF CV	HF P	HF CV	HF P
0.00	0.05	0.05	0.05	0.05
0.25	0.25	0.26	0.16	0.15
0.50	0.73	0.73	0.59	0.60
0.75	0.96	0.96	0.92	0.92
1.00	1.00	1.00	0.99	0.99

HF: minimum Half Factor sum of pair-differences
CV: power obtained via comparison to critical value
P: power obtained via permuting index labels

5. Recommendations on Obtaining *p*-Values

Our brief exploration yields some findings. We see the sharpest inequality is conservative. We determine that the power to detect change is independent of whether one uses a null hypothesis simulation-derived critical value, a null hypothesis permutation-test critical value, or a *p*-value directly from permutation testing on observations. This result gives the investigator liberty to choose the method that works best for the situation.

In many situations, sufficient control data will not be available. Even if control data are available, assuming the null hypothesis carries the risk that the alternative is actually true. For these reasons, and in light of attaining equivalent detection powers using the latter three methods, our recommendation is to use permutation testing to directly obtain the *p*-value for testing purposes.

E. CHANGE-DETECTION POWER COMPARISONS

We present computer simulation results that compare the change-detection power of the *k*-factor sum of pair-differences (SPD) test and the composite *k*NN test to

parametric and nonparametric alternatives, the James, James, and Siegmund (JJS) (1992) test and Ensemble Sum of Pair-Maxima (ESPM) test of Ruth (2009). As determined in Section C of this chapter, we use the minimum half factor SPD test ($k = N/2$) with score function parameter $\lambda = 1$, and the $k = N/2$ composite kNN test with $\lambda = -0.3$. We use the simulation scenarios as outlined in Section C of this chapter.

We extract our minimum subgraphs based on the Euclidean distance between observations. Ruth (2009) finds that using the Euclidean distance results in higher detection powers than using a covariance-scaled metric such as the Mahalanobis distance. Detection powers are based on 10,000 simulations per vignette. We vary the sampling distribution (multivariate normal, multivariate normal mixture, multivariate Weibull), the dimension ($p = 5$ or $p = 20$), the change parameter (location or scale), the change type (jump or drift), and the change magnitude (0.0 to 1.0 in increments of 0.25).

1. Multivariate Normal Vignettes

Tables 15 through 17 show the detection powers for the multivariate normal cases. Discussion follows the tables.

Table 15. Test power to detect a location change of magnitude Δ in the first component of the mean vector at observation 101 of 200 drawn from a five-dimensional multivariate normal sampling distribution, with 10,000 simulations per vignette. Jump case is to the left; drift case is to the right (after Ruth, 2009).

F_{MVN}; $p = 5$; mean change	Jump				Drift			
Δ	ESPM	JJS	HF	kNN	ESPM	JJS	HF	kNN
0.00	0.06	0.05	0.05	0.05	0.05	0.05	0.05	0.05
0.25	0.17	0.12	0.16	0.08	0.10	0.09	0.09	0.06
0.50	0.60	0.53	0.58	0.31	0.26	0.23	0.25	0.13
0.75	0.94	0.93	0.94	0.73	0.57	0.54	0.56	0.30
1.00	1.00	1.00	1.00	0.97	0.86	0.86	0.85	0.60

ESPM: Ensemble Sum of Pair-Maxima JJS: James, James, and Siegmund test
HF: Half Factor sum of pair-differences kNN: composite kNN test

Table 16. Test power to detect a location change of magnitude Δ in the first component of the mean vector at observation 101 of 200 drawn from a 20-dimensional multivariate normal sampling distribution, with 10,000 simulations per vignette. Jump case is to the left; drift case is to the right (after Ruth, 2009).

F_{MVN}; p = 20; mean change	Jump				Drift			
Δ	ESPM	JJS	HF	*k*NN	ESPM	JJS	HF	*k*NN
0.00	0.05	0.04	0.05	0.05	0.05	0.04	0.05	0.05
0.25	0.10	0.06	0.10	0.07	0.07	0.05	0.07	0.06
0.50	0.34	0.21	0.32	0.14	0.14	0.09	0.15	0.09
0.75	0.75	0.64	0.73	0.38	0.33	0.23	0.31	0.15
1.00	0.98	0.96	0.97	0.73	0.59	0.49	0.57	0.29

ESPM: Ensemble Sum of Pair-Maxima JJS: James, James, and Siegmund test
HF: Half Factor sum of pair-differences *k*NN: composite *k*NN test

Table 17. Test power to detect a scale change of magnitude Δ in all components at observation 101 of 200 drawn from a five-dimensional multivariate normal sampling distribution, with 10,000 simulations per vignette. Jump case is to the left; drift case is to the right (after Ruth, 2009).

F_{MVN}; p = 5; scale change	Jump				Drift			
Δ	ESPM	JJS	HF	*k*NN	ESPM	JJS	HF	*k*NN
0.00	0.05	0.06	0.05	0.05	0.05	0.05	0.05	0.05
0.25	0.32	0.09	0.25	0.98	0.13	0.11	0.11	0.81
0.50	0.96	0.15	0.90	1.00	0.53	0.26	0.38	1.00
0.75	1.00	0.21	1.00	1.00	0.93	0.42	0.81	1.00
1.00	1.00	0.25	1.00	1.00	1.00	0.55	0.98	1.00

ESPM: Ensemble Sum of Pair-Maxima JJS: James, James, and Siegmund test
HF: Half Factor sum of pair-differences *k*NN: composite *k*NN test

A parametric change-detection method such as JJS is expected to work only with the distribution for which it was designed. We find the graph-theoretic ESPM and minimum half factor SPD tests achieve detection powers comparable with the parametric JJS test in the multivariate normal mean change cases, consistent with Ruth (2009). The

minimum half factor SPD detection power holds closely to that of the ESPM test for mean changes, but its power lags somewhat in the scale drift case. This lag is further explored in Section C of this chapter. The composite *k*NN test shows dominance in scale changes, but falls short of the other three tests in detecting mean changes. Compared to the other tests, there is a greater decrease in detection power for the NN test going from lower to higher dimensions.

2. Multivariate Normal Mixture Vignettes

Tables 18 through 20 show the detection power for the multivariate normal mixture cases as described in Section C of this chapter.

Table 18. Test power to detect a location change of magnitude Δ in the first component of the mean vector at observation 101 of 200 drawn from a five-dimensional multivariate normal mixture sampling distribution, with 10,000 simulations per vignette. Jump case is to the left; drift case is to the right. Shading indicates excessive false alarm rate (after Ruth, 2009).

F_{mix} ; p = 5; mean change	Jump				Drift			
Δ	ESPM	JJS	HF	*k*NN	ESPM	JJS	HF	*k*NN
0.00	0.05	0.20	0.05	0.05	0.05	0.20	0.05	0.05
0.25	0.12	0.22	0.12	0.07	0.08	0.21	0.08	0.06
0.50	0.42	0.32	0.41	0.16	0.18	0.25	0.18	0.09
0.75	0.81	0.56	0.80	0.43	0.39	0.35	0.39	0.17
1.00	0.97	0.82	0.97	0.78	0.67	0.48	0.66	0.34

ESPM: Ensemble Sum of Pair-Maxima JJS: James, James, and Siegmund test
HF: Half Factor sum of pair-differences *k*NN: composite *k*NN test

Table 19. Test power to detect a location change of magnitude Δ in the first component of the mean vector at observation 101 of 200 drawn from a 20-dimensional multivariate normal mixture sampling distribution, with 10,000 simulations per vignette. Shading indicates excessive false alarm rate.

F_{mix} ; p = 20; mean change	Jump			
Δ	ESPM	JJS	HF	*k*NN
0.00	0.05	0.22	0.05	0.05
0.25	0.08	0.23	0.07	0.05
0.50	0.18	0.26	0.18	0.06
0.75	0.41	0.34	0.39	0.10
1.00	0.72	0.54	0.70	0.17

ESPM: Ensemble Sum of Pair-Maxima JJS: James, James, and Siegmund test
HF: Half Factor sum of pair-differences *k*NN: composite *k*NN test

Table 20. Test power to detect a scale change of magnitude Δ in all components at observation 101 of 200 drawn from a five-dimensional multivariate normal mixture sampling distribution, with 10,000 simulations per vignette. Jump case is to the left; drift case is to the right. Shading indicates excessive false alarm rate (after Ruth, 2009).

F_{mix} ; p = 5; scale change	Jump				Drift			
Δ	ESPM	JJS	HF	*k*NN	ESPM	JJS	HF	*k*NN
0.00	0.05	0.20	0.05	0.05	0.05	0.21	0.05	0.05
0.25	0.31	0.22	0.24	0.74	0.14	0.27	0.11	0.50
0.50	0.89	0.25	0.80	1.00	0.47	0.37	0.37	0.94
0.75	1.00	0.30	0.99	1.00	0.85	0.45	0.74	1.00
1.00	1.00	0.33	1.00	1.00	0.99	0.53	0.95	1.00

ESPM: Ensemble Sum of Pair-Maxima JJS: James, James, and Siegmund test
HF: Half Factor sum of pair-differences *k*NN: composite *k*NN test

For the multivariate normal mixture distribution with mean changes, the ESPM test demonstrates nearly identical change-detection power as the minimum half factor SPD test. For scale changes, the ESPM test shows detection power greater than the minimum half factor SPD test. The composite *k*NN test again demonstrates relatively low

detection power for mean changes but shows superior detection power in scale changes. Comparing the lower dimensional and higher dimensional mean jump cases, we see a greater decrease in detection power for the composite *k*NN test compared to the other tests. The JJS test fails to hold its level for this heavy-tailed distribution.

3. Multivariate Weibull Vignettes

Tables 21 and 22 show the detection power for the multivariate Weibull cases as defined in Section C of this chapter.

Table 21. Test power to detect a scale parameter change of magnitude Δ in the first component of the β vector at observation 101 of 200 drawn from a five-dimensional multivariate Weibull sampling distribution, with 10,000 simulations per vignette. Jump case is to the left; drift case is to the right. Shading indicates excessive false alarm rate (after Ruth, 2009).

F_{Weib} ; p = 5; β change	Jump				Drift			
Δ	ESPM	JJS	HF	*k*NN	ESPM	JJS	HF	*k*NN
0.00	0.05	0.09	0.05	0.05	0.05	0.09	0.05	0.05
0.25	0.27	0.23	0.25	0.17	0.13	0.18	0.12	0.13
0.50	0.78	0.70	0.73	0.55	0.42	0.47	0.38	0.33
0.75	0.98	0.97	0.96	0.88	0.74	0.79	0.68	0.61
1.00	1.00	1.00	1.00	0.99	0.93	0.95	0.89	0.84

ESPM: Ensemble Sum of Pair-Maxima JJS: James, James, and Siegmund test
HF: Half Factor sum of pair-differences *k*NN: composite *k*NN test

Table 22. Test power to detect a scale parameter change of magnitude Δ in the first component of the β vector at observation 101 of 200 drawn from a 20-dimensional multivariate Weibull sampling distribution, with 10,000 simulations per vignette. Shading indicates excessive false alarm rate.

F_{Weib} ; p = 20; β change	Jump			
Δ	ESPM	JJS	HF	*k*NN
0.00	0.05	0.07	0.05	0.05
0.25	0.17	0.12	0.16	0.11
0.50	0.62	0.33	0.59	0.32
0.75	0.94	0.70	0.92	0.68
1.00	1.00	0.92	0.99	0.93

ESPM: Ensemble Sum of Pair-Maxima JJS: James, James, and Siegmund test
HF: Half Factor sum of pair-differences *k*NN: composite *k*NN test

For the multivariate Weibull distribution, the ESPM test exceeds the detection power of the minimum half factor SPD test and the composite *k*NN test across the two dimensions (p = 5 and p = 20) considered. The composite *k*NN test nearly matches the performance of the minimum half factor SPD test for the β drift scenario. The composite *k*NN test shows a greater decrease in power from the lower dimensional β jump case to the higher dimensional β jump case than the ESPM and minimum half factor SPD tests. The JJS test fails to hold its level for this skewed distribution.

4. Summary of Change-Detection Power Comparisons

Strengths and limitations of the minimum half factor and *k*NN statistics come through these simulations. The minimum half factor SPD test shows power to detect change across all 14 simulated scenarios, at or near the power of the ESPM test with the notable exceptions of the multivariate normal and heavy-tailed scale changes. The composite *k*NN test shows superior power in scale change cases, but trails considerably in all other cases.

IV. GRAPH-THEORETIC CHANGE-POINT LOCALIZATION

In the event that change is detected in the sampling distribution of a series of observations, the natural question is, where does the onset of change occur? In this chapter, we propose methods for change-point localization that are applied if the null hypothesis of homogeneity is rejected. We limit our localization efforts to a search for a single abrupt change.

A. CHANGE-POINT ESTIMATION USING EDGE COUNT DATA

We derive change-localization statistics from the same minimum half factor and *k*NN subgraphs that detected change using the tests described in Chapter III. These statistics reveal the property that, when moving along the observation sequence, the edges of the minimum subgraph exhibit a pattern that would not be expected under the null hypothesis. By counting the edges going to or coming from observations with a lower index value than the current observation, we can localize the change point with reasonable accuracy. Recall from Chapter III Section C, a *prior edge* is an edge of the minimum subgraph whose destination observation label is less than the source (reference) observation label.

Under the alternative hypothesis, there is a tendency to form minimum subgraph edges among observations on the same side of the change point (Friedman & Rafsky, 1979). Observations just before the change point should have more prior edges than observations just after a change point. In a minimum subgraph, observations after the change point are more likely to be adjacent to observations like themselves – observations identified with a higher label value. For a single abrupt change, we expect observations just after the change point to have relatively low prior edge counts. By finding a point along the observation sequence where the prior edge count drops abruptly, we estimate the change point. To formalize this reasoning, we use piecewise regression based on the prior edge counts.

1. Statistical Properties

We use L_i to denote the prior edge count for observation i based on the minimum half factor or the *outgoing* edges of the kNN subgraph:

$$\begin{aligned} x_{ij} &= \begin{cases} 1, \text{ if edge } (i,j) \text{ is in the minimum subgraph} \\ 0, \text{ otherwise} \end{cases}, \quad \forall (i,j) \in E \\ L_i &= \sum_{j=1}^{i-1} x_{ij}, \ \forall i \in \{2,\ldots,N\}. \end{aligned} \tag{46}$$

Note that $L_N = N/2$.

We use (47) to obtain the prior *incoming* edge count, L_i'. The irregular indegree of the kNN graph edges requires different treatment for the prior incoming edges than the prior outgoing edges.

$$\begin{aligned} x_{ji} &= \begin{cases} 1, \text{ if edge } (j,i) \text{ is in the minimum subgraph} \\ 0, \text{ otherwise} \end{cases}, \ \forall (j,i) \in E \\ L_i' &= \sum_{j=1}^{i-1} x_{ji}, \ \forall i \in \{2,\ldots,N\}. \end{aligned} \tag{47}$$

Note that L_N' may take any value in $\{0, 1, \ldots, N-1\}$.

Under the null hypothesis, L_i and L_i' have a hypergeometric distribution (Wackerly, Mendenhall, & Scheaffer, 2008). For the minimum half factor and the *outgoing* edges of the kNN subgraph, the probability of L_i taking on any value from zero to k is given by

$$\Pr[L_i = l] = \frac{\binom{i-1}{l}\binom{N-i}{N-1-l}}{\binom{N-1}{k}}, \tag{48}$$

where $N-1$: population size (all indices except itself)
k: sample size (number of edges incident to i)
$i-1$: total prior edges available
l: possible values for the prior edge count $\{0,1,\ldots k\}$
$l \le i-1$
$k-l \le N-i.$

We know the null hypothesis expected value and variance of the prior edge count for each observation. Later, these parameters are used for standardization:

$$\begin{aligned} E\left[L_i\right] &= k\left(\frac{i-1}{N-1}\right) \\ V\left[L_i\right] &= k\left(\frac{i-1}{N-1}\right)\left(\frac{N-i}{N-1}\right)\left(\frac{N-1-k}{N-2}\right). \end{aligned} \tag{49}$$

The null hypothesis parameters for the *k*NN prior *incoming* edge count need a slight modification. For the *k*NN subgraph, the indegree of observation i $\left(k_i'\right)$ may take any value in $\{0, 1, \ldots, N-1\}$. The conditional distribution of L_i', given the total number of incoming edges for vertex *i*, is hypergeometric:

$$x_{ji} = \begin{cases} 1, & \text{if edge } \left(j,i\right) \text{ is in the subgraph} \\ 0, & \text{otherwise} \end{cases}, \ \forall \left(j,i\right) \in E$$

$$\begin{aligned} \Pr\left[L_i' = l \,\middle|\, \sum_{j=1, j\neq i}^{N} x_{ji} = k_i'\right] &= \frac{\binom{i-1}{l}\binom{N-i}{N-1-l}}{\binom{N-1}{k_i'}} \\ E\left[L_i'\right] &= k_i'\left(\frac{i-1}{N-1}\right) \\ V\left[L_i'\right] &= k_i'\left(\frac{i-1}{N-1}\right)\left(\frac{N-i}{N-1}\right)\left(\frac{N-1-k_i'}{N-2}\right). \end{aligned} \tag{50}$$

For all subgraphs, the hypergeometric random variables describing the prior edge count have the indicated expected value and variance under the null hypothesis, but the random variables are not independent.

We search for a change point by optimizing the fit of a two-piece regression line to the standardized prior edge counts. To avoid endpoint effects, we do not consider indices below some τ_0 and above $N - \tau_0$. Each of the remaining observations is interrogated as a possible break point, and the break point providing the best fit for the piecewise regression based on minimizing the residual mean squared error (MSE) is chosen.

Piecewise regression may allow discontinuities in the value of the conditional mean of the dependent variable or vector, and/or only allow discontinuities in the slope of the regression line (Koul & Qian, 2002). For our application, we choose a model that allows discontinuity continuity in the value of the conditional mean of the dependent variable. Our algorithm is a variation of one from Crawley (2007). Our localization method is specific to estimating a single change point.

Our regression is of the following form:

$$\begin{aligned} &i: \text{observation index, } i \in \{2,\ldots,N\} \\ &\tau: \text{interrogation index, } \tau \in \{\tau_0,\ldots,N-\tau_0\}; \\ &\quad \text{usually } \tau_0 = \lfloor N/10 \rfloor, \text{ where } \lfloor A \rfloor \text{ is the largest integer} \le A \\ &Z_i: \text{standardized prior edge count for index } i. \end{aligned} \tag{51}$$

For each $\tau \in \{\tau_0, \text{K}\ , N-\tau_0\}$ we calculate:

$$MSE(\tau) = \sum_{i=2}^{N} \left(Z_i - \hat{\beta}_{0,\tau} - \hat{\beta}_{1,\tau} i - \hat{\beta}_{2,\tau} I\{i > \tau\} - \hat{\beta}_{3,\tau} i I\{i > \tau\} \right)^2. \tag{52}$$

Let $\hat{\tau}$ denote the index value that minimizes the MSE(τ). We take $\hat{\tau}$ as our estimator of the change point.

We have two available estimators from the *k*NN subgraph, namely the prior incoming edges estimate $\hat{\tau}_{\text{IN}}$ and the prior outgoing edges estimate $\hat{\tau}_{\text{OUT}}$. We take $\hat{\tau}$ to be $\hat{\tau}_{\text{IN}}$ or $\hat{\tau}_{\text{OUT}}$ depending on which produces the smaller value of

$$\begin{aligned} &\frac{MSE_{\text{IN}}\left(\hat{\tau}_{\text{IN}}\right)}{SST_{\text{IN}}} \text{ and } \frac{MSE_{\text{OUT}}\left(\hat{\tau}_{\text{OUT}}\right)}{SST_{\text{OUT}}}, \\ &SST_{\text{IN}} = \sum_{i=2}^{N} \left(Z_{i,\text{IN}} - \bar{Z}_{\text{IN}} \right)^2, \; SST_{\text{OUT}} = \sum_{i=2}^{N} \left(Z_{i,\text{OUT}} - \bar{Z}_{\text{OUT}} \right)^2. \end{aligned} \tag{53}$$

2. A Change-Point Estimation Procedure

Including the change-detection process, we proceed along these steps:

(1) Extract the applicable subgraph (minimum half factor or *k*NN).

(2) Use the procedures of Chapter III to test the hypothesis of no change.

(3) If the no-change hypothesis is not rejected, stop here and do NOT conduct change-point localization.

(4) If the null hypothesis is rejected, obtain the change-point estimator $\hat{\tau}$ using the process described above.

B. CHANGE-POINT LOCALIZATION: ILLUSTRATIONS

Returning to our example data sets from Chapter II Section A, Figures 12 and 13 provide the minimum half factor and 10-NN subgraphs, respectively, for the no-change case and the change case. In both figures, the no-change case is to the left and the change case is to the right. Observations are listed along the bottom and top of each figure as the reference vertices and matching vertices, respectively. The prior edges for each reference vertex are red; other minimum subgraph edges are grey. Figure 13 places the source observations as the reference vertices to highlight *k*NN prior outgoing edges. The same figure inverted would highlight *k*NN prior incoming edges; these edges are grey in Figure 13.

As seen in the plots on the left of Figures 12 and 13, in the no-change case, some prior edges cross from higher reference vertices to lower matching vertices. Red is seen throughout the plots. From the plots to the right, it is evident that not as many prior edges cross from higher reference vertices to lower matching vertices in the change case as in the no-change case. Figure 12 demonstrates that the minimum half factor is a 10-regular subgraph, while Figure 13 shows that the vertices in the *k*NN subgraph have constant outdegree and fluctuating indegree.

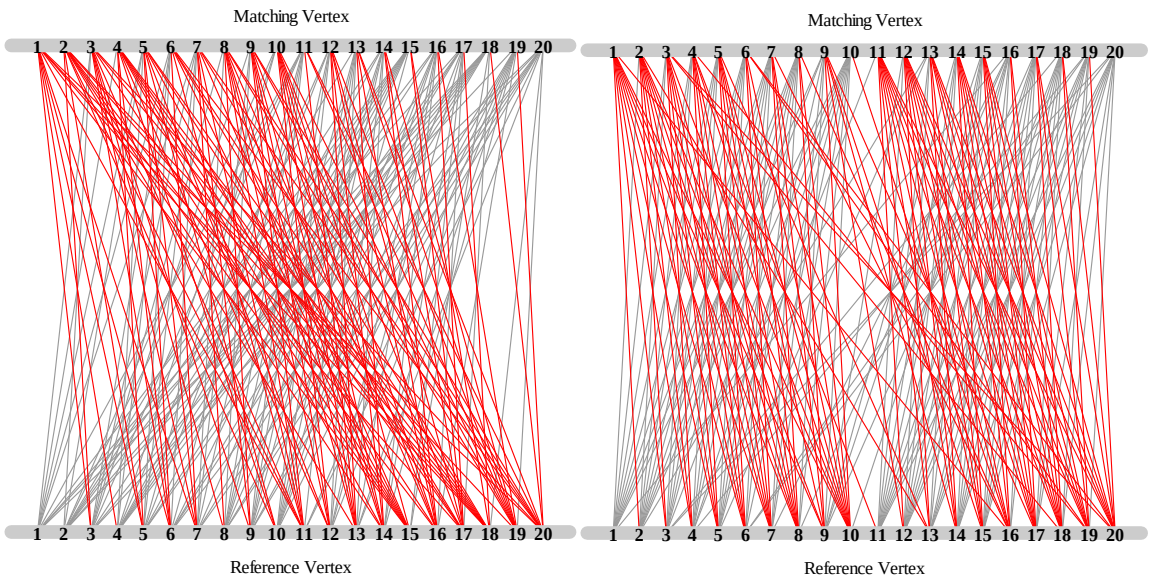

Figure 12. Minimum half factor for the example data sets. Prior edges of reference vertices are shown in red. The no-change case is to the left, the change case is to the right.

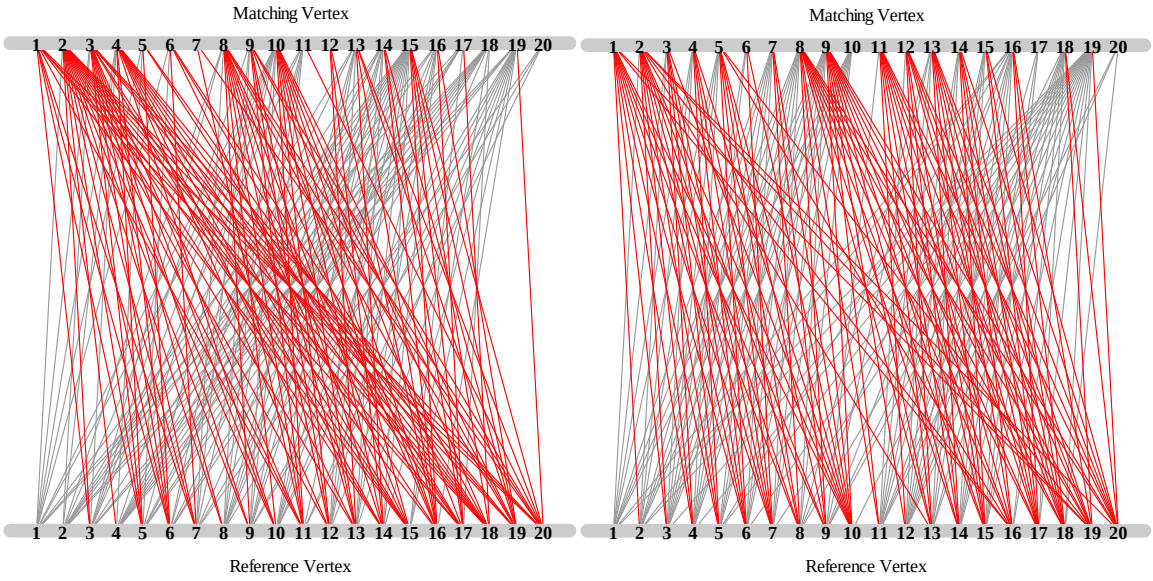

Figure 13. The 10-nearest neighbors for the example data sets. Prior outgoing edges of reference vertices are shown in red. The no-change case is to the left, the change case is to the right.

Table 23 amplifies Figure 13, showing the 10-NN of each vertex as well as the prior outgoing edge count, its expected value, standard deviation, and standardized value, for the no-change case to the left and the change case to the right. The standardized prior

edge count on the left of Table 23 shows no discernible pattern for the no-change sample. As seen on the right of Table 23, the standardized prior edge count (z_i) for the change case shows a clear pattern. All values of z_i, $i < 11$ are positive, while almost all values of z_i, $i > 10$ are negative. Analogous tables for minimum half factor and prior incoming edge count are given in Appendix C.

Table 23. The 10-nearest neighbors prior outgoing edge count (l_i), expected value, standard deviation, and standardized value for example data sets. The no-change case is to the left, the change case is to the right.

Standard Bivariate Normal

Vertex i	l_i	μ	σ	z_i
1	0	0.00	0.00	--
2	0	0.53	0.50	−1.05
3	2	1.05	0.69	1.38
4	1	1.58	0.82	−0.71
5	4	2.11	0.91	2.08
6	5	2.63	0.98	2.41
7	4	3.16	1.04	0.81
8	2	3.68	1.08	−1.56
9	3	4.21	1.10	−1.10
10	4	4.74	1.12	−0.66
11	7	5.26	1.12	1.56
12	5	5.79	1.10	−0.72
13	6	6.32	1.08	−0.29
14	7	6.84	1.04	0.15
15	7	7.37	0.98	−0.37
16	9	7.89	0.91	1.21
17	9	8.42	0.82	0.71
18	10	8.95	0.69	1.53
19	10	9.47	0.50	1.05
20	10	10.00	0.00	--

Change in Mean Vector from (1, 0) to (3, 0) at Observation $i = 11$

Vertex i	l_i	μ	σ	z_i
1	0	0.00	0.00	--
2	1	0.53	0.50	0.95
3	2	1.05	0.69	1.38
4	3	1.58	0.82	1.74
5	4	2.11	0.91	2.08
6	5	2.63	0.98	2.41
7	6	3.16	1.04	2.73
8	5	3.68	1.08	1.22
9	5	4.21	1.10	0.72
10	9	4.74	1.12	3.82
11	2	5.26	1.12	−2.92
12	2	5.79	1.10	−3.43
13	5	6.32	1.08	−1.22
14	5	6.84	1.04	−1.77
15	6	7.37	0.98	−1.39
16	8	7.89	0.91	0.12
17	8	8.42	0.82	−0.52
18	8	8.95	0.69	−1.38
19	10	9.47	0.50	1.05
20	10	10.00	0.00	--

As shown previously, using the tests from Chapter III, we reject the null hypothesis for the change case. Both the prior incoming and prior outgoing edge count regressions result in $\hat{\tau} = 11$, as shown in Figure 14.

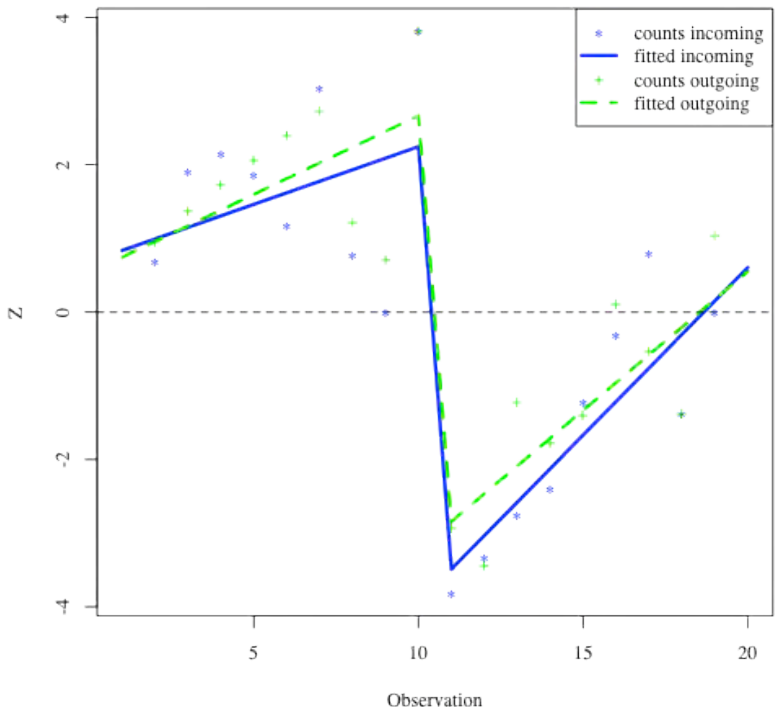

Figure 14. Standardized count of nearest-neighbor prior edges and optimized piecewise regression lines, for the 10-nearest neighbor subgraph formed from the example data sets. The prior outgoing edge counts are in green, the prior incoming edge counts are in blue.

Next, we illustrate our proposed change localization procedure on a larger data set. Starting with $N = 200$ five-dimensional multivariate normal simulated observations, we insert an abrupt jump in the first component of the mean vector at observation $\tau = 101$ by adding 0.5 to those values. Figure 15 shows the changed component of our data, plotted by observation label.

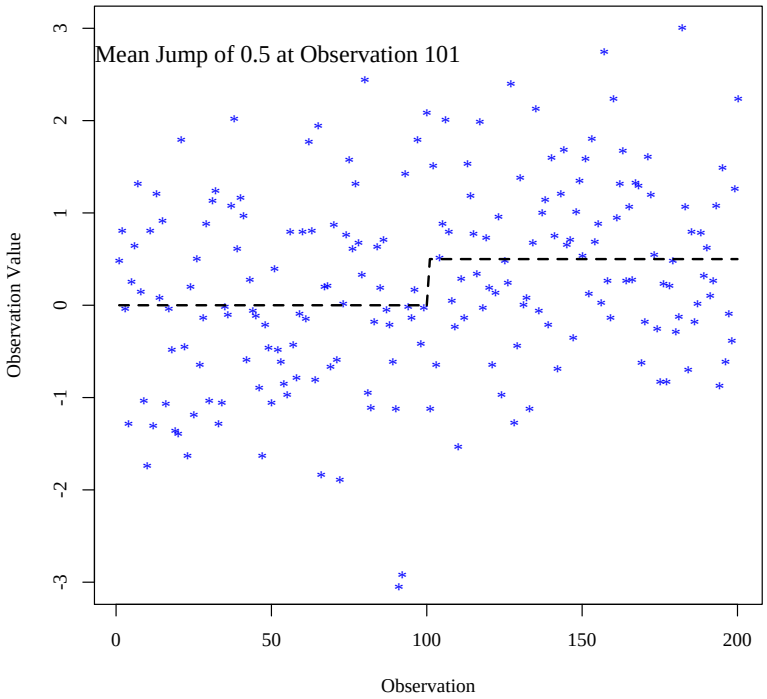

Figure 15. The changed component of our five-dimensional example data, plotted against observation label (N = 200). Data are randomly generated from a multivariate normal distribution. The first component of the mean vector is the dotted black line, showing the inserted change of magnitude 0.5 at observation 101. The remaining four dimensions are unchanged.

Using these observations, we obtain the minimum half factor and calculate the minimum half factor sum of pair-differences (SPD). The standardized SPD statistic for these data is −2.97, which is less than the critical value −2.81 based on the unimodal bound in (44) of Chapter III Section D. By permuting the observation labels 10,000 times we obtain a p-value of 0.0075 leading to a rejection of the null hypothesis of homogeneity at the $\alpha = 0.05$ test level, and thus conclude that the observations exhibit a change in the sampling distribution.

Now for all indices $i \in \{2,\ldots,N\}$ we determine the prior edge counts l_i as described in (46). For example, observation 15 has its 100 edges connecting to the following indices that are less than 15 {2, 3, 4, 5, 6, 7, 8, 11, 13}, so $l_{15} = 9$. Under the null hypothesis, $E[l_{15}] = 7.04$ and $SD[l_{15}] = 1.81$, which gives a standardized count of $z_{15} = 1.09$ (54).

$$z_{15} = \frac{l_{15} - E[l_{15}]}{\sqrt{V[l_{15}]}} = \frac{9 - 7.04}{1.81} = 1.09 \tag{54}$$

Using the standardized counts, we search for the optimal point to break a piecewise regression. We iterate through the indices $i = \{20, 21, \ldots, 180\}$, splitting the data into two subsets just past each considered index. We fit the linear regression as described in (51) and (52). The index that minimizes the MSE is our change-point estimate.

Figure 16 shows the optimized, least-squares piecewise regression lines overlaid on the standardized prior edge count for which the change point is estimated at $\hat{\tau} = 111$.

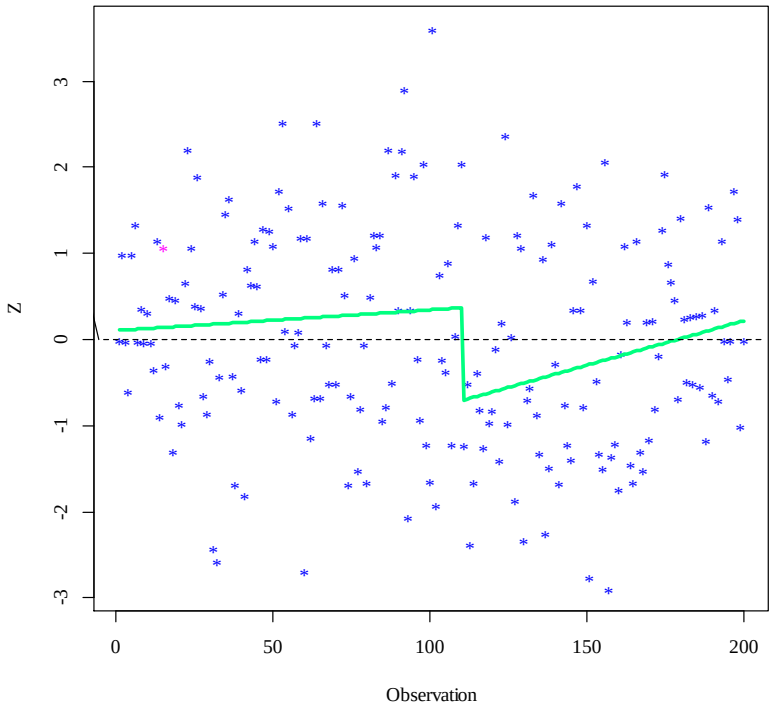

Figure 16. Standardized count of prior edges in blue, with the optimized piecewise regression lines in green, for the minimum half factor formed from the larger data set. The standardized count of prior edges for observation 15 is in magenta ($z_{15} = 1.09$)

The slopes of the fitted regression lines show that for indices less than the proposed change point, prior edges were more likely than to be expected under the null hypothesis ($z > 0$). Just after the proposed change point, prior edges were less likely than to be expected under the null hypothesis ($z < 0$). Recall that our example is for a mean jump of 0.5 in the first of five dimensions at $\tau = 101$. Finally, Figure 17 shows the estimated change point plotted with the changed component of our data.

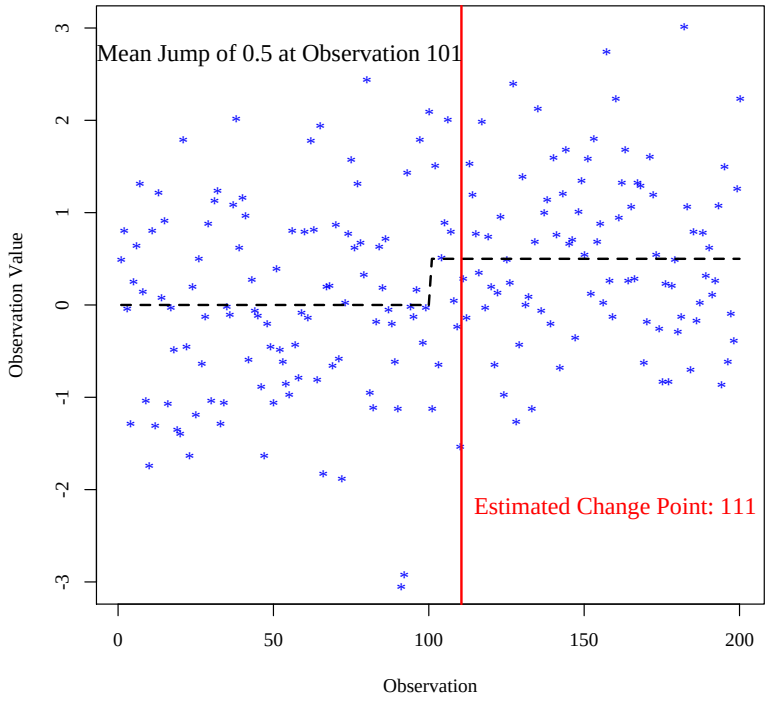

Figure 17. The estimated change point and the changed component of our five-dimensional example data plotted against observation label. The first component of the mean vector is the dotted black line, showing the inserted change of magnitude 0.5 at observation 101. The estimated change point is the red line. The remaining four dimensions are unchanged.

In our evaluation of change-point estimators, we regard estimator $\hat{\tau}$ as close to the truth if falls in a narrow window centered on the truth with a width $0.05N$. For $N = 200$, the window width is 10, so the estimator must fall within $\tau \pm 5$. Despite change being detected and localized, the estimated change point of $\hat{\tau} = 111$ fails this test and is judged not to be close to the actual change point of $\tau = 101$.

C. LOCATION ESTIMATOR COMPARISONS

We use a series of simulations to evaluate the location estimators. The simulations expose the estimators to different distributions, change types, and change magnitudes. Location estimator accuracy is assessed in two ways. We primarily report a success rate, which we define as the fraction of times the change point was localized within the $0.05N$-wide window centered on true change point. Localization occurs only if change is detected – if no change is detected, that is counted as a localization failure. We also report the interquartile range (IQR) of the estimated change points.

We test our location estimators under a variety of scenarios, similar to the scenarios considered in Chapter III. For our location estimator simulations, we simulate five-dimensional observations. We place an abrupt (jump) change in the middle of the observation sequence. We conduct 1,000 simulations per scenario, and display the success rate as the fraction of times out of 1,000 that the change was detected and localized within plus or minus five indices of the actual change point. Tables 24 through 26 present the results of a simulation study of success rates and IQRs for change-point estimations using the minimum half factor prior edge count (HF) and the best of the *k*NN prior incoming edge count and the *k*NN prior outgoing edge count as described in (53). Each table identifies the distribution $(F_{\text{MVN}}, F_{\text{mix}}, F_{\text{Weib}})$, the parameter (mean, scale) under change, and the change magnitudes (0.25, 0.50, 0.75, and 1.00). A sample size of $N = 200$ and actual change point of $\tau = 101$ is used in all vignettes. For changes in distribution mean, and for the change in the scale of the multivariate Weibull distribution, we simulate the change in only the first component of each observation. For scale changes in the multivariate normal and multivariate normal mixture distributions, we simulate the change in all components.

1. Multivariate Normal Vignettes

Table 24 shows the location estimator success rate and IQR for the multivariate normal cases.

Table 24. Change-point localization success rate and IQR for the five-dimensional multivariate normal simulated observations with 1,000 simulations per vignette. A change of magnitude Δ is inserted at observation $\tau = 101$ of $N = 200$. Mean vector change is to the right; scale change is to the left. Success rate is defined as the fraction of times the change is detected at a 0.05 test level and change-point estimator equal to $\tau \pm 5$. IQR is in parentheses.

$F_{MVN}; p = 5$	Mean Jump		Scale Jump	
Δ	HF	*k*NN	HF	*k*NN
0.25	0.02 (55)	0.01 (71)	0.04 (65)	0.23 (28)
0.50	0.17 (25)	0.08 (40)	0.45 (10)	0.58 (11)
0.75	0.52 (8)	0.38 (10)	0.80 (3)	0.85 (2)
1.00	0.76 (3)	0.73 (4)	0.94 (1)	0.96 (1)

HF: SPD test and edges from the minimum Half Factor
*k*NN: Composite test and edges from the *k*-Nearest Neighbors

For each of the two change-point estimators, in both the location and scale change vignettes, the success rate increases and the IQR narrows as the change magnitude increases. For the location change vignettes, the minimum half factor success rate exceeds that of the *k*NN estimator. In the multivariate normal scale change scenarios, the success rate of the *k*NN estimators exceeds that of the minimum half factor. Both distinctions narrow as the change magnitude increases.

2. Multivariate Normal Mixture Vignettes

Table 25 shows the location estimator success rate and IQR for the multivariate normal mixture cases.

Table 25. Change-point localization success rate and IQR for the five-dimensional multivariate normal mixture simulated observations with 1,000 simulations per vignette. A change of magnitude Δ is inserted at observation $\tau = 101$ of $N = 200$. Mean vector change is to the right; scale change is to the left. Success rate is defined as the fraction of times the change is detected at a 0.05 test level and change-point estimator equal to $\tau \pm 5$. IQR is in parentheses.

F_{mix}; $p = 5$	Mean Jump		Scale Jump	
Δ	HF	*k*NN	HF	*k*NN
0.25	0.01 (70)	0.01 (69)	0.05 (48)	0.13 (42)
0.50	0.11 (36)	0.03 (57)	0.36 (14)	0.35 (20)
0.75	0.34 (16)	0.15 (22)	0.70 (5)	0.58 (12)
1.00	0.59 (7)	0.42 (9)	0.85 (3)	0.74 (4)

HF: SPD test and edges from the minimum Half Factor
*k*NN: Composite test and edges from the *k*-Nearest Neighbors

Similar to the multivariate normal location change case, the minimum half factor success rate consistently exceeds the success rate of the *k*NN estimator in the multivariate normal mixture location change case. The results are not as pronounced for the multivariate normal mixture scale change case; the minimum half factor estimator success rate exceeds that of the *k*NN estimator for the larger change magnitudes examined.

3. Multivariate Weibull Vignettes

Table 26 shows the location estimator success rate and IQR power for the multivariate Weibull cases.

Table 26. Change-point localization success rate and IQR for the five-dimensional multivariate Weibull simulated observations with 1,000 simulations per vignette. A change of magnitude Δ is inserted in the scale parameter β at observation $\tau = 101$ of $N = 200$. Success rate is defined as the fraction of times the change is detected at a 0.05 test level and change-point estimator equal to $\tau \pm 5$. IQR is in parentheses.

F_{Weib}; $p = 5$	β Jump	
Δ	HF	*k*NN
0.25	0.05 (53)	0.02 (56)
0.50	0.28 (18)	0.21 (22)
0.75	0.56 (8)	0.51 (9)
1.00	0.71 (6)	0.69 (5)

HF: SPD test and edges from the minimum Half Factor
*k*NN: Composite test and edges from the *k*-Nearest Neighbors

The minimum half factor achieves more success than the *k*NN edges for the multivariate Weibull with a scale parameter change, although it is a close distinction.

4. Summary of Change-Point Estimator Comparisons

For location changes, estimators based on prior edge counts derived from the minimum half factor have a higher success rate than those of the *k*NN, but for scale change, the success rate for *k*NN prior edge counts is generally higher. We come to a similar conclusion with larger window sizes than the $0.05N$ width tabulated here.

THIS PAGE INTENTIONALLY LEFT BLANK

V. CHANGE ANALYSIS TECHNIQUES: AN APPLICATION

In this chapter, we illustrate our proposed change detection and localization methods by applying them to the fire detection test data of Gottuk et al. (2002). Our graph-theoretic methods assume data independence, which may be violated in data collected from sensors over a period of time. We show how our methods can be applied to overcome this situation. We also demonstrate a modification that can be employed to reduce the sensitivity of our statistical tests to signals that should not be regarded as requiring detection. Rather than the offline change detection for which our methods are designed, fire detection requires online change analysis. Accordingly, we implement our analysis in a near real-time manner.

A. PROBLEM DESCRIPTION

In this section, we describe the data collected and three challenges they present. We examine a multivariate data set from one of a series of 24 tests as described in Gottuk et al. (2002) and JiJi et al. (2003). These tests were conducted by the Naval Research Laboratory (NRL) onboard a decommissioned U.S. Naval vessel, ex-USS SHADWELL. The tests were part of a larger effort to "develop an early warning fire detection (EWFD) system that is highly immune to nuisance alarms" (Gottuk et al., 2002, p. 1). The goal of this series of tests was to determine the behavior of a real-time integrated alarm system under stimuli resulting from combustion. Each of the 24 tests feature a combustion event involved a different pairing of shipboard location and combustible material. NRL analyzed the data from each test with a probabilistic neural network algorithm. The algorithm was trained with data from more than 40 combustion events conducted previous to the 24 tests (JiJi et al., 2003). Although we prepare the data as described in Section B of this chapter, our methods do not rely on training data.

1. Sensor Locations and Types

Gottuk et al. (2002) positioned 14 clusters of fire detection sensors on the second and third decks (floors) covering 12 areas in the forward section of the decommissioned ship. Figures 18 and 19 show the sensor cluster positions. The deck numbering is from

the top down, so the second deck is above the third deck. EWFD cluster locations are numbered from one to 14. Clusters one through nine were placed on the second deck; clusters 10 through 14 were placed on the third deck. Doors and hatches within the test area remained open, but the test area was cut off from weather and the rest of the ship.

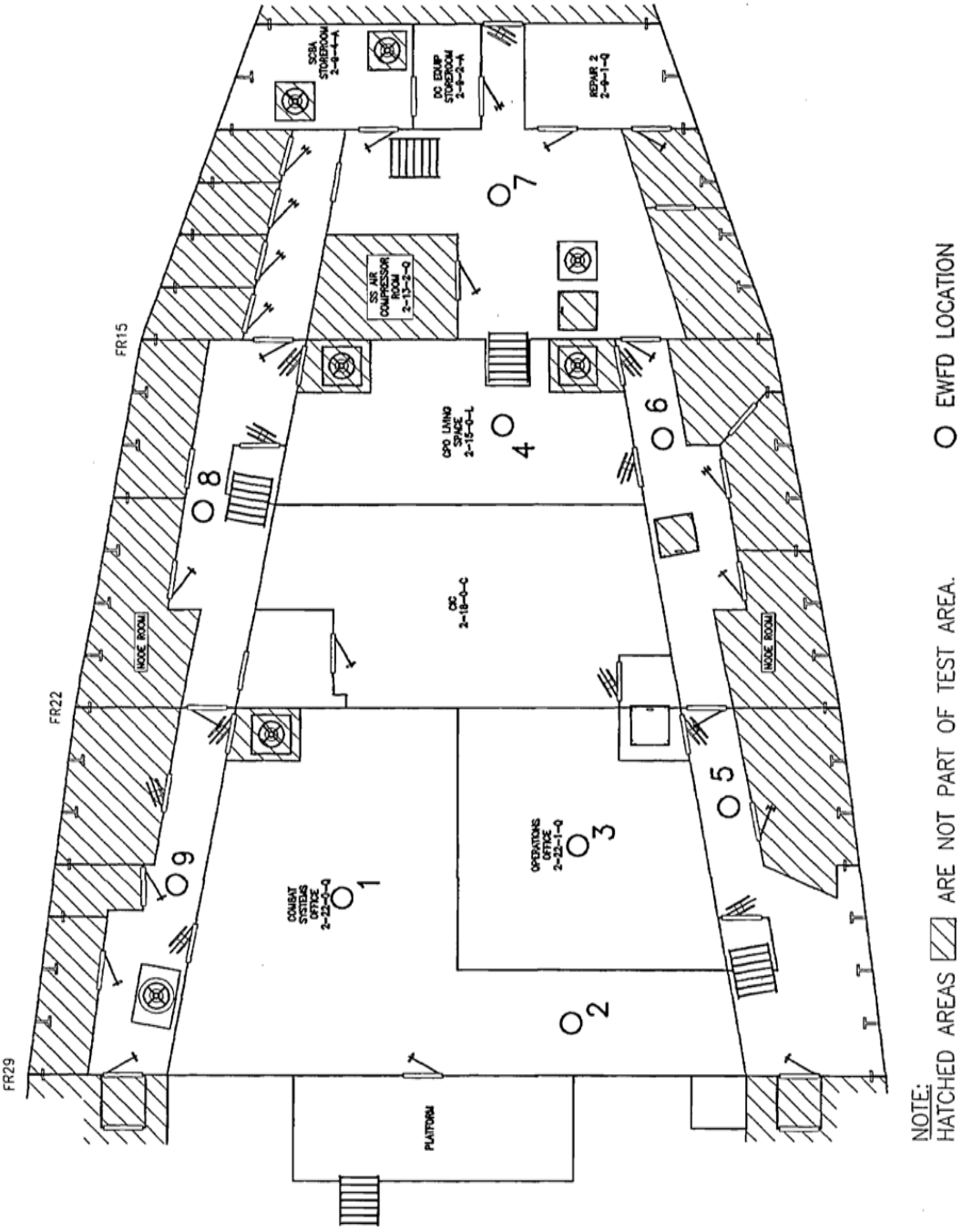

Figure 18. Second deck detection layout (from Gottuk et al., 2002).

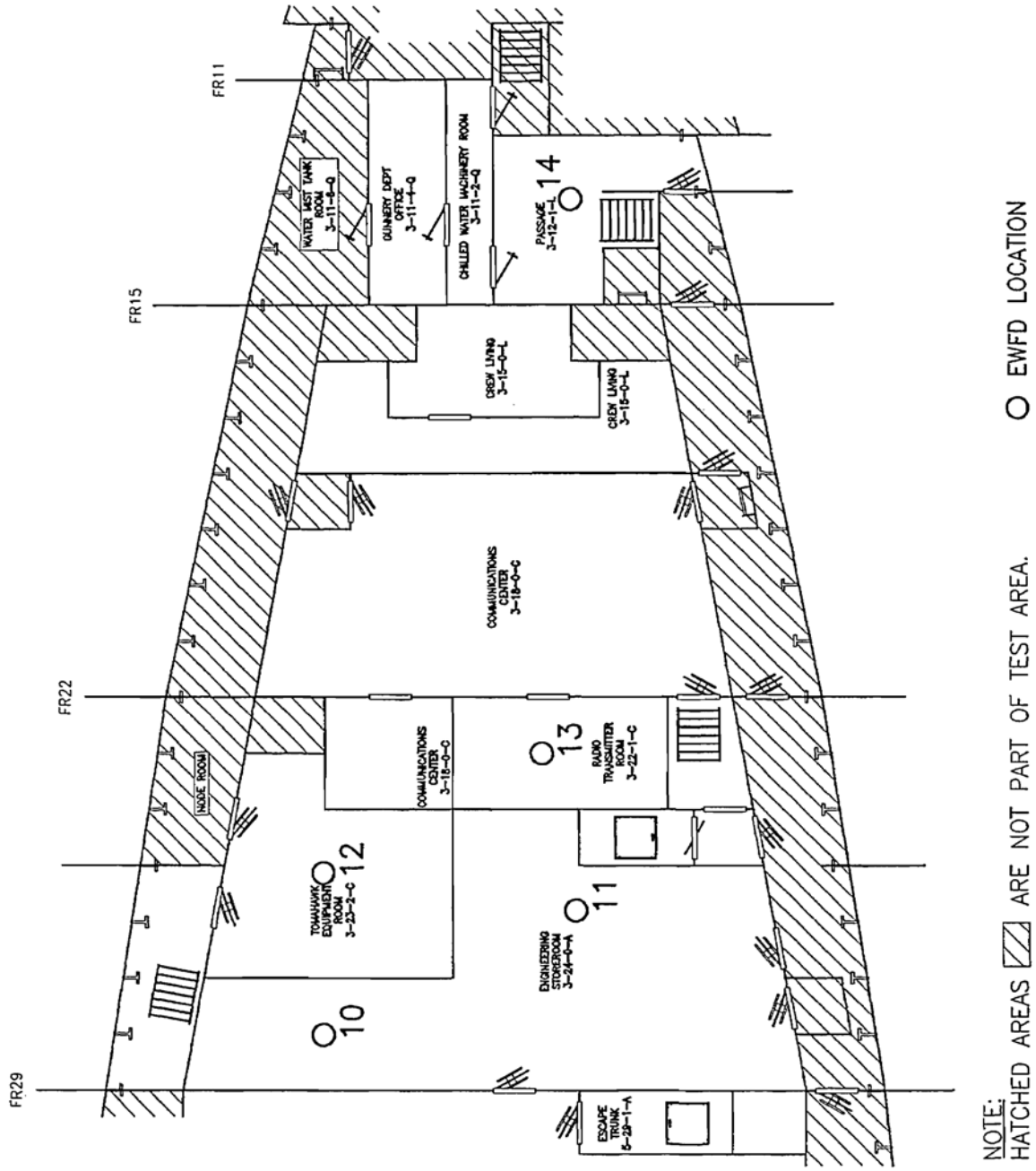

Figure 19. Third deck detection layout (from Gottuk et al., 2002).

In each of the 14 clusters, four different sensors were employed: an ionization smoke detector, a photoelectric smoke detector, a carbon monoxide detector, and a carbon dioxide detector. The sensors remain in continuous operation throughout the test series. We limit consideration to the ionization and photoelectric smoke detectors. During the control period, the carbon dioxide sensors show non-stationarity, and the carbon

monoxide sensors show spiking from possible radio interference (see Appendix D). The ionization and photoelectric sensors were mounted overhead, 2.9 m above the deck.

2. Test Stages and Sensor Response

In each of the 24 alarm system tests, sensor data were recorded at two-second intervals. A notional test timeline is shown in Figure 20. Data recording began at the start of a pre-stimulus period of several minutes prior to the test start time. Source ignition served as a time reference point for a given test. The gap between the start of the test and event initiation ranged from five to 40 minutes. Ten to 75 minutes after event initiation, after alarms would have been expected to trigger, the investigators applied ventilation to clear atmospheric contaminants and establish conditions for the next test. Data collection for a test ended after a few minutes of ventilation.

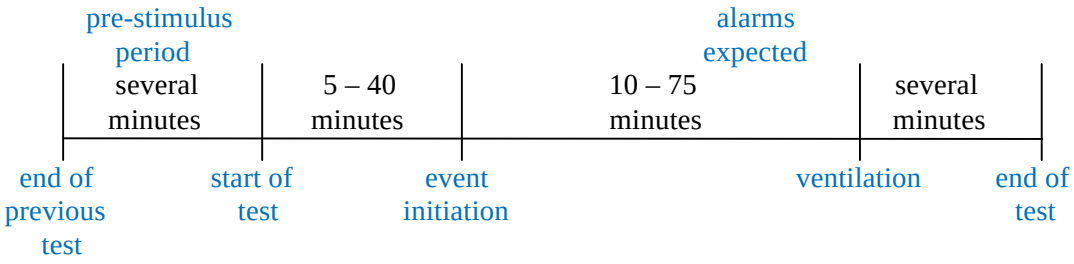

Figure 20. Notional stages for fire tests onboard ex-USS SHADWELL.

The promptness and level of sensor response vary with the concentration of airborne products from a combustion source. A major fire or smoke event quickly produces large changes in combustion sensor measurements. Smoldering events produce responses that are more subtle and slower to develop. Figure 21 illustrates the sensor responses to combustion stimuli in two different test scenarios. Depicted on the left are data from the two sensors at EWFD Location 3, the closest location to a second deck F-76 diesel fuel-spill fire. On the right, we plot data from the two sensors at EWFD Location 10, the closest location to a third deck smoldering computer monitor. We standardize these measurements using the sample means and standard deviations of pre-stimulus data (data collected prior to the start of the test). For the diesel fuel fire event, rapid change is evident in each of the standardized measurements within seconds of the

fuel fire initiation. For the smoldering computer monitor event, the sensor responses are less immediate.

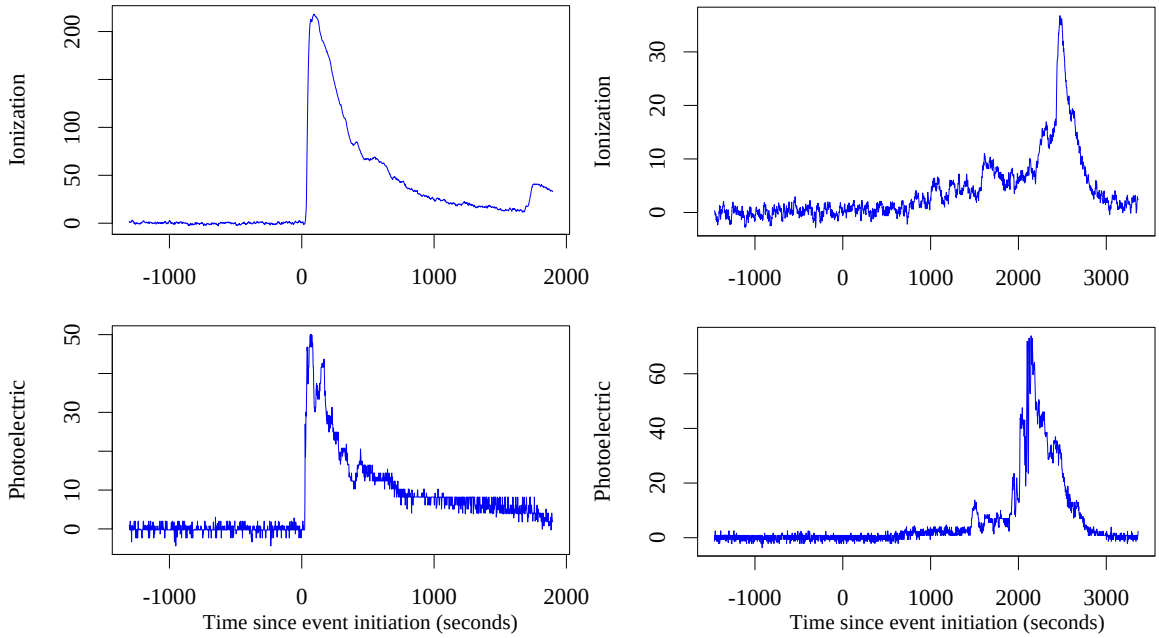

Figure 21. Sensor response to combustion events. To the left, standardized sensor data from Location 3, closest to an F-76 fuel spill fire. To the right, standardized sensor data from Location 10, closest to a smoldering computer monitor. Sensors are ionization (top) and photoelectric (bottom) smoke detectors. The horizontal axes mark the time from event initiation in seconds.

Both ionization and photoelectric sensors use gas diffusion into the unit to produce measurements (JiJi et al., 2003). The ionization sensor contains a radioactive isotope, Americium-241, that produces ionization in the air. Under normal conditions, the ionized air allows current to flow between two electrodes in body of the sensors. The current is continuously compared to the current between electrodes in a sealed reference chamber. Under fire conditions, ions attach to smoke particles entering the sensor, and the current drops below the reference current. The photoelectric sensor contains a light source and an offset photodiode receiver. Under normal conditions, light from the source does not reach the receiver. Under fire conditions, smoke particles entering the sensor interact with the light, deflecting some light to the photodiode receiver. Both sensors

report their respective measures in units of percent obscuration per unit length (NFPA, 2008).

In the more than 1,000 observations during the control periods of both tests of Figure 21, the photoelectric sensors only take values in the set $\{-0.25, -0.24, \ldots, 0.27\}$. Ties may be present when solving for the minimum subgraphs among observations in the absence of a stimulus, however, Figure 21 shows that the photoelectric sensor response in the presence of a combustion event is well above the control-period values.

We demonstrate our graph-theoretic methods in the test case of a smoldering computer monitor, which proved difficult to detect in the analysis reported by Gottuk et al. (2002). The smoldering computer monitor was located in the third deck engineering storeroom (see Figure 19). We consider data only from the five third-deck EWFD Locations (10, 11, 12, 13, and 14), of which Locations 10 and 11 were closest to the combustion source. Considering each location separately, there are five two-dimensional data streams.

3. Challenges in Applying Graph-Theoretic Change Detection Techniques

Although having pre-stimulus "control" data is not strictly necessary for applying our methods, it is useful in the present situation for two reasons. First, it provides a useful baseline for standardization. Second, it allows us to examine if, and how, the sensor measurements are correlated with respect to time. Collected prior to the test start time, these observations are assumed to be representative of the background environment. From them, we observe the characteristics of the sensors without contamination from outside stimuli. We assume that the pre-stimulus data are free from any initial effects, e.g., we expect any instrumentation burn-in to be complete.

In our evaluation of the pre-stimulus observations, we find evidence of dependency in the form of *autocorrelation*. Autocorrelation is the correlation of a series of measurements at one point in time with the same series at a different point in time. An echo is an example of autocorrelation. Time series data, such as the sensor data we consider in the present example, are frequently characterized by dependencies such as

autocorrelation (Box & Jenkins, 1976). It is important to remove autocorrelation because our minimum subgraph algorithms rely on independence in the data. Under the null hypothesis, we assume that any observation has an equal likelihood of being adjacent to any other observation in the subgraph. If autocorrelation is not removed, observations would be more likely to match to their immediate neighbors in the measurement sequence.

We assume that the pre-stimulus observations from each individual sensor form a *second-order stationary* time series. By this, we assert that the mean (μ), variance (σ^2), and autocorrelations are constant for each sensor. We express the autocorrelation (ρ_s) of a second-order stationary time series Y_t, $t = 1, 2, \ldots$, at a particular lag s, $s = 1, 2, \ldots$, in terms of the mean and variance of the time series (Box & Jenkins, 1976):

$$\rho_s = \frac{E\left[\left(Y_t - \mu\right)\left(Y_{t+s} - \mu\right)\right]}{\sigma^2}. \tag{55}$$

The numerator and denominator of (55) are the autocovariance of the time series at lag s and the variance of the time series, respectively.

An autoregressive (AR) model expresses simultaneous autocorrelation of more than one lag value. Such a model, fit to our data, could provide the means to remove autocorrelation, if present. We use the autocorrelation function (ACF) and the partial autocorrelation function (PACF) to detect autocorrelation and estimate parameters of the AR model.

We estimate the autocorrelation for different lag values, and plot the resulting ACF, $\hat{\rho}_s$, as a function of the lag value s. The **`acf`** function in R produces an ACF plot, and displays 95 percent probability bands assuming that the autocorrelation at a particular lag value is equal to zero. An ACF value outside the bands indicates autocorrelation. We do not know the population mean and variance and must estimate them from our data (Box & Jenkins, 1976):

$$\hat{\rho}_s = \frac{\frac{1}{N-s}\sum_{t=1}^{N-s}\left(Y_t - \bar{Y}\right)\left(Y_{t+s} - \bar{Y}\right)}{\frac{1}{N}\sum_{t=1}^{N}\left(Y_t - \bar{Y}\right)^2}. \tag{56}$$

The PACF is a correlation conditioned on the previous observations, and is determined by solving what are known as the Yule-Walker equations (Box & Jenkins, 1976). The **`pacf`** function in R displays 95 percent probability bands assuming that the partial autocorrelation at a particular lag value is equal to zero. An AR model for lag values {1, …, *p*} has the following representation (Box & Jenkins, 1976):

$$Y_t - \mu = \sum_{s=1}^{p} \phi_{ps}\left(Y_{t-s} - \mu\right) + \varepsilon_t; \quad \varepsilon_1, \ldots, \varepsilon_N \text{ are iid } N\left(0, \sigma_\varepsilon^2\right). \tag{57}$$

If the true AR model order is *p*, the PACF is 0 at lag *s*, for $s = p + 1, p + 2, \ldots$.

The change detection methods presented in this dissertation are designed to detect any type of change, without regard to whether the signal is large or small. The goal of fire detection is to detect effects that are sufficiently large to warrant the activation of an alarm. Low-level stimuli, for example due to a person with a lit cigarette entering a room where a smoke detector has been placed, should not generate an alarm. Triggering false alarms can lead to complacency:

> Consider the consequences of a high [smoke detector] false alarm rate on the performance of the subsequent human monitor. A busy human monitor may soon learn to ignore the smoke detector's alarm signal, considering it a false alarm and not worthy of a shift in attention from more pressing duties. The performance of the overall smoke detector-human monitor system would be worse than if the smoke detector were set to emit fewer alarms. (Sorkin & Woods, 1985, p. 52)

Consequently, an adaptation must be made to our graph-theoretic techniques to allow low-level stimuli to not signal a change.

In addition to adequately handling data dependency and reducing sensitivity to low-level change, successful employment in this illustration requires our methods to respond to the immediacy of a fire situation.

B. DATA PREPARATION

In this section, we discuss data handling techniques to overcome the three difficulties presented in Section A of this chapter.

1. Diagnosing and Mitigating Autocorrelation of Sensor Measurements

We demonstrate the estimation and mitigation of autocorrelation with the Location 10 ionization sensor. Figure 22 depicts EWFD Location 10 ionization sensor pre-stimulus data, for which there are 551 observations. Although these measurements fluctuate, they appear to track the immediately prior observation, indicating possible autocorrelation.

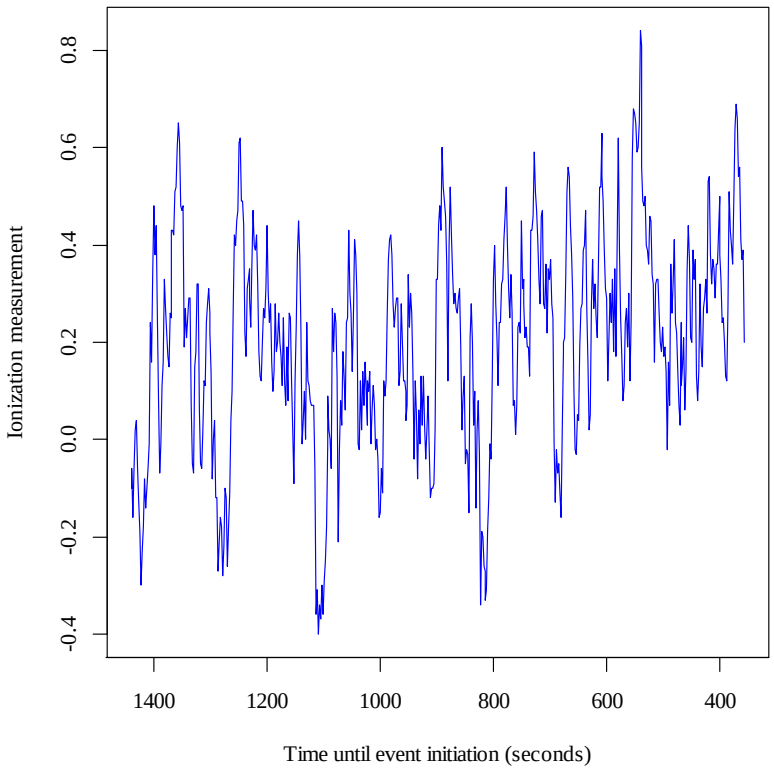

Figure 22. Location 10 ionization sensor pre-stimulus data for the smoldering computer monitor test (551 observations).

We inspect the estimated autocorrelation function (ACF) and partial autocorrelation function (PACF) to determine the type of time-series dependence, if any, that may be present. If dependence is indicated, we use the R function **`ar`** to estimate the

order and the parameters of an AR model. The order is chosen by minimizing the Akaike Information Criterion (AIC). A fuller autoregressive-moving average (ARMA) model (Box & Jenkins, 1976) may be considered if the order of the AR model is large. We calculate the residuals of the estimated AR model, and re-evaluate the estimated ACF and PACF for indications of time-series dependence not removed by the AR model. The process of extracting residuals from a time-series model that has little or no indication of time-series dependence among them is called *whitening*.

We again turn to the ionization sensor data from EWFD Location 10 to illustrate the whitening process. Plots of the ACF and PACF applied to centered pre-stimulus data are shown in Figure 23. Autocorrelation is indicated by one or more lag values of the ACF exceeding the 95 percent probability bands around the value of zero, under the null hypothesis that there is no autocorrelation. The PACF plot gives indication that an AR model of order one may be appropriate to remove the autocorrelation.

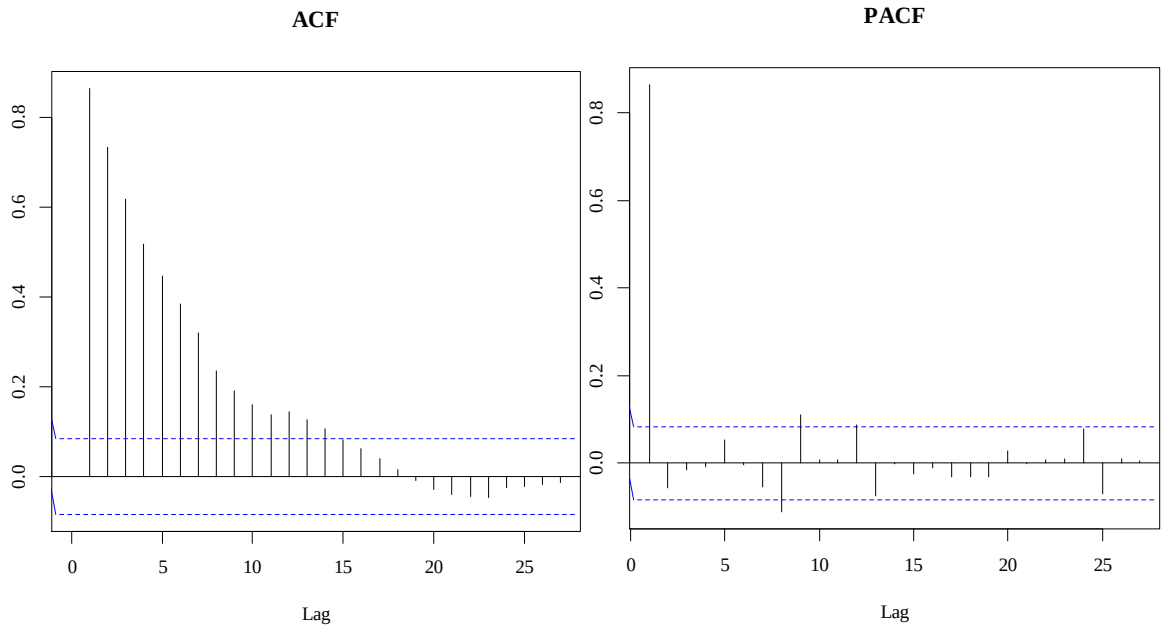

Figure 23. Autocorrelation plots for centered pre-stimulus data from the Location 10 ionization sensor. To the left, the estimated autocorrelation function plot; to the right, the estimated partial autocorrelation function plot. The blue dashed lines indicate the 95 percent probability bands for the absence of correlation.

Having detected autocorrelation, we use the **`ar`** function in R to obtain estimated AR model parameters for the centered Location 10 ionization sensor data:

$$\begin{aligned}\hat{y}_t = {} & 0.9225\left(y_{t-1}-\bar{y}\right)-0.0499\left(y_{t-2}-\bar{y}\right)+0.0057\left(y_{t-3}-\bar{y}\right) \\ & -0.0758\left(y_{t-4}-\bar{y}\right)+0.0657\left(y_{t-5}-\bar{y}\right)+0.0336\left(y_{t-6}-\bar{y}\right) \\ & +0.0568\left(y_{t-7}-\bar{y}\right)-0.2158\left(y_{t-8}-\bar{y}\right)+0.1134\left(y_{t-9}-\bar{y}\right).\end{aligned} \tag{58}$$

We apply this AR model to the centered data, and then test the whitened data for autocorrelation to examine the effectiveness of whitening. As shown in Figure 24, the whitened data show little autocorrelation.

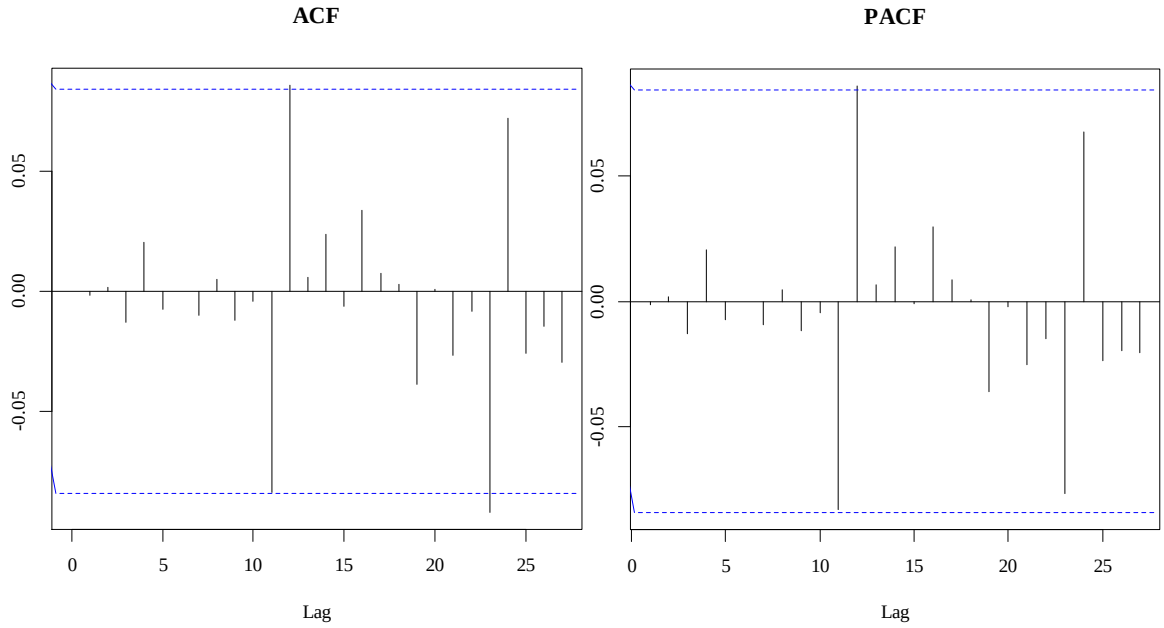

Figure 24. Autocorrelation plots for centered, whitened pre-stimulus data from the Location 10 ionization sensor. To the left, the estimated autocorrelation function plot; to the right, the estimated partial autocorrelation function plot. The blue dashed lines indicate the 95 percent probability bands for the absence of correlation.

Our methods assume that the data are statistically independent, as distances between observations form the basis for obtaining the minimum half factor and the *k*-nearest neighbors (*k*NN) subgraphs, and we appear to have achieved that. We complete

the transformation by dividing the centered, whitened, pre-stimulus data by its standard deviation $(\hat{\sigma})$. A more robust measure, such as the IQR, could be used.

An assumption of normality is not required by our methods. An examination of the normal quantile-quantile plot in Figure 25 shows that the transformed observations from the Location 10 ionization sensor do not severely depart from normality. The AR model we estimate for the Location 10 photoelectric sensor is as follows:

$$\hat{y}_t = 0.0898\left(y_{t-1} - \bar{y}\right). \tag{59}$$

The transformed observations from the Location 10 photoelectric sensor are seen to be nearly-discretely valued.

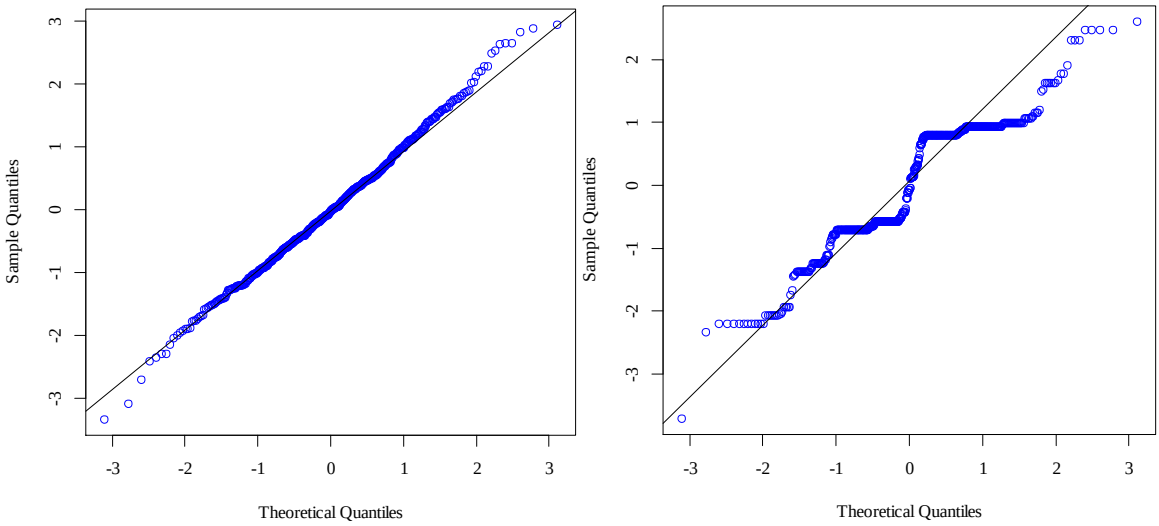

Figure 25. Normal quantile-quantile plots for centered, whitened, standardized pre-stimulus data from the Location 10 ionization sensor (left) and the Location 10 photoelectric sensor (right). Discrete values of the photoelectric sensor are noted in Subsection A.2.

To describe dependence of the transformed data across sensors, we produce a scatterplot of the Location 10 sensors in Figure 26. There is no apparent correlation between the two sensors. This is interesting because it appears that the information provided by the two sensors really is different. What causes perturbations in one sensor is not causing perturbations in the other, and vice versa.

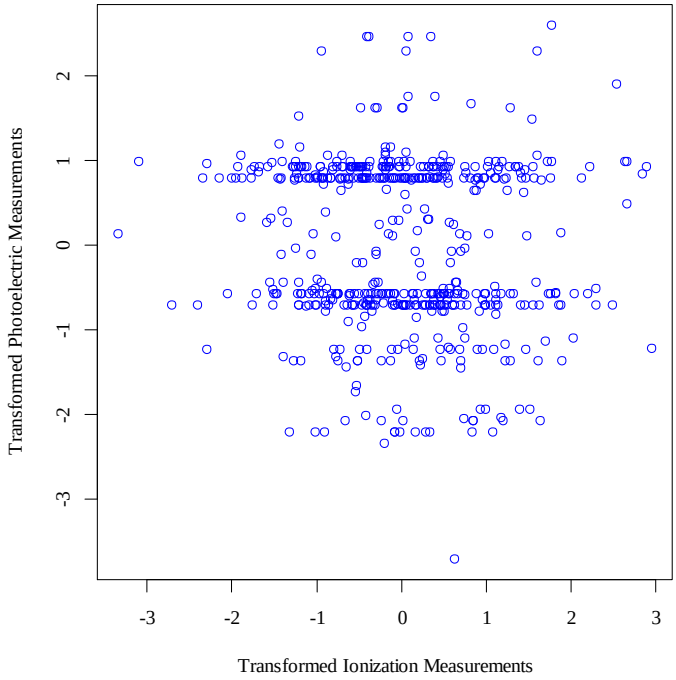

Figure 26. Scatterplot of transformed measurements for Location 10.

We check the transformed data for correlation across the two sensors at each location by estimating the correlation coefficients. The estimates shown in Table 27 indicate relative independence between the two sensors at each location, although the 95-percent confidence interval for the correlation coefficient does not include zero at Location 10. The indicated correlation between the Location 10 ionization sensor and the Location 10 photoelectric sensor in the transformed pre-stimulus observations is weak and does not require mitigation.

Table 27. Estimated value and 95-percent confidence interval for the correlation coefficient between transformed pre-stimulus observations of ionization and photoelectric sensors from each location.

	Location 10	Location 11	Location 12	Location 13	Location 14
Estimate	−0.090	−0.009	0.073	0.001	0.011
95% CI	(−0.173, −0.006)	(−0.093, 0.075)	(−0.011, 0.156)	(−0.083, 0.085)	(−0.073, 0.095)

2. Detection above a Threshold Level

To prevent unwanted alarms due to low-level stimuli, we apply an upper threshold value to the sensor data using the following process:

(1) The pre-stimulus observations are centered and whitened as described in the previous section.

(2) Determine a threshold value such that values below the threshold are considered not to justify an alarm. In practice, this value should reflect expert judgment about the strength of signal considered actionable for fire detection.

(3) Apply a locally weighted scatterplot smoother to the whitened data. We use a LOESS smoother with span equal to 0.25 (Cleveland & Grosse, 1990).

(4) Truncate the smoothed data by taking the minimum of the smoothed values and the threshold value.

(5) Produce a set of residuals by subtracting the truncated, smoothed data from the whitened test observations.

We call these residuals the *thresholded* data and use them in our subsequent analysis. For the sake of illustration, we use the upper quartile plus 1.5 times the interquartile range (IQR) as a threshold value. Our threshold, although not reflective of knowledge of fire detection, is often used for statistical outlier detection in the construction of box plots (Devore, 2011). As with whitening, we apply thresholding separately for each sensor and location. For comparison purposes, we calculate the probability that an observation sampled from the standard normal distribution exceeds the threshold as described above. Standard normal data are expected to have first and third quartiles of −0.6745 and 0.6745, respectively. The expected interquartile range is then 1.349. The probability of a standard normal data exceeding the threshold value of 2.698 is 0.00349 – roughly 3.5 times in a thousand.

Figure 27 provides a graphical illustration of thresholding, using the Location 10 ionization sensor. On the left of Figure 27, we show the whitened pre-stimulus observations in purple, the resulting threshold in black, whitened test observations in blue, smoothed data in orange, and the truncated data in dashed red. On the right of Figure 27 is the thresholded data. Thresholding masks changes that remain under the threshold value, but allows changes of greater magnitude to pass through.

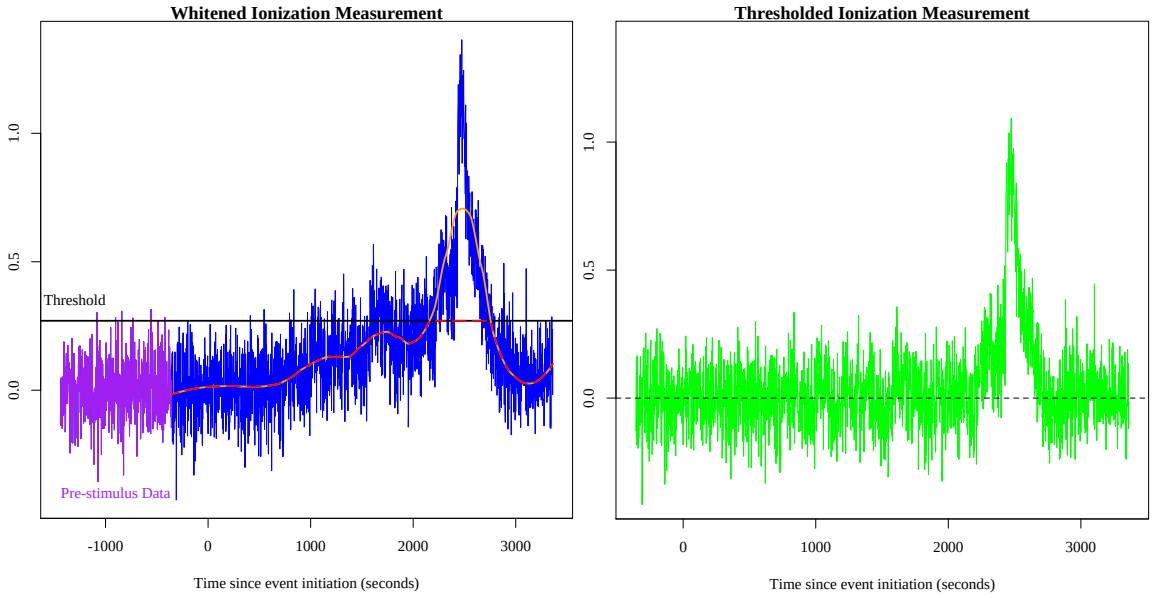

Figure 27. Illustration of thresholding, using the Location 10 ionization sensor. To the left, whitened data is shown in purple for pre-stimulus observations and blue for test observations, plotted against the time since event initiation. The upper outlier threshold determined from pre-stimulus data is the black line. The loess smoother is shown in orange. The truncated loess smoother—the quantity to subtract from the whitened data—is the red dashed line. To the right, the thresholded data—the difference between the whitened data and the truncated loess smoother—are plotted in green against the time since event initiation.

We use a similar process to transform the pre-stimulus data from the other nine sensors. Table 28 displays the mean, AR order, AR parameters, threshold, and standard deviation obtained for each sensor. By transforming each sensor individually, we compensate for differences across sensors of the same type, such as length of service and manufacturing variability.

Table 28. Transformation parameters for centering, whitening, thresholding, and standardizing third deck sensor data from the test of interest (smoldering computer monitor in engineering storeroom). Parameters are based on 551 pre-stimulus observations at each location.

Location and Sensor	Mean	AR Order	AR Parameters			Threshold	SD
10 Ionization	0.201	9	0.9225	−0.0499	0.0057	0.270	0.107
			−0.0758	0.0657	0.0336		
			0.0568	−0.2158	0.1134		
11 Ionization	0.760	2	0.8610	0.0690		0.378	0.139
12 Ionization	0.051	1	0.8665			0.317	0.122
13 Ionization	0.902	8	0.9421	−0.0306	−0.0661	0.327	0.124
			0.0108	−0.0032	−0.0102		
			−0.0170	0.0940			
14 Ionization	0.332	1	0.9498			0.368	0.132
10 Photoelectric	0.018	1	0.0898			0.189	0.060
11 Photoelectric	0.000	1	0.1348			0.151	0.051
12 Photoelectric	0.014	1	0.1107			0.260	0.087
13 Photoelectric	−0.037	5	0.1837	−0.0285	−0.0104	0.225	0.063
			−0.0483	−0.0982			
14 Photoelectric	−0.018	2	0.1885	−0.0673		0.057	0.056

AR: autoregressive model

3. Achieving Detection in Near Real-Time

Ideally, we would analyze the data sequentially. A fire alarm loses its effectiveness if data processing time causes delay. For example, the computation time required to obtain a minimum half factor makes online processing of new sensor observations every two seconds impractical. NN subgraphs are more amenable to real-time updating (Li, Yang, & Han, 2004). In order to provide a common focus for applying our methods, we limit ourselves to batch processing. This is done by analyzing overlapping windows of data for change every minute (30 observations per update).

C. RESULTS

We now apply the change detection and localization methods that we described in Chapters III and IV to the EWFD sensors. Before doing so we discuss several factors that affect realism in the handling of the data.

1. Inclusion of Reference Data

A decision must be made whether to apply our change detection methods relative to a pre-stimulus data set, or to analyze observations in moving windows that are potentially subject to stimulus. We limit our window size (N) due to the computational cost of the minimum half factor and the desire to reduce time latency. We combine a reference set of pre-stimulus observations with labels $1,2,\ldots,n_0$ with real-time observations $1,2,\ldots,n_1$, where $n_0 + n_1 = N$. We include pre-stimulus observations in hopes of detecting more subtle changes, such as a slow build-up of gases, that take longer to develop than the length of our window size. A slow, upward drift may not register as a change when only analyzing observations after the drift began. To compare the effects of including reference data, we conduct three sets of analyses, outlined in Table 29.

Table 29. Durations of pre-stimulus observations and real-time observations for three analysis schemes. Observations are taken every two seconds.

	Pre-stimulus		Real Time		Total
	Observations (n_0)	Duration (min:sec)	Observations (n_1)	Duration (min:sec)	observations (N)
(1)	0	0:00	200	6:40	200
(2)	45	1:30	155	5:10	200
(3)	90	3:00	110	3:40	200

We conduct analysis in windows of 200 observations (1) using no reference set and a moving window of six minutes 40 seconds (N = 200 real-time observations), (2) using a fixed reference set of 90 seconds (n_0 = 45 pre-stimulus observations) and a moving window of five minutes 10 seconds (n_1 = 155 real-time observations), and (3) using a

fixed reference set of 180 seconds (n_0 = 90) and a moving window of three minutes 40 seconds (n_1 = 110).

2. Tests from Minimum Subgraphs

We illustrate change detection and localization using both the minimum half factor and the *k*NN subgraphs for each 200-observation analysis window. For each method, we use permutation tests to obtain *p*-values for testing purposes as described in Chapter III. In order for a change to be detected based on the minimum half factor, the minimum half factor Sum of Pair-Differences (SPD) statistic defined in (22) with $k = 100$, must be less than or equal to the fifth percentile of 10,000 permutations of the index labels for an $\alpha = 0.05$ level of test. In order for a change to be detected based on the *k*NN subgraph with $k = 100$, one of two criteria of a composite *k*NN test must be satisfied: either the *k*NN sum of pair-differences (22) or the *k*NN prior edge count (39) must be less than or equal to the 2.5^{th} percentile of 10,000 permutations of the index labels. We use a critical value of 0.025 for each of these two tests to control the Type I error of the composite test at or below $\alpha = 0.05$.

In addition to the actual change-detection capabilities, a second consideration in choosing a matching method is the computational burden. As noted in Chapter III, obtaining the minimum half factor is best done with proprietary optimization software such as CPLEX. Nearest neighbors is easily implemented in R, and there are several open source *k*NN packages available for R. The empirically observed worst-case computation times for both matching methods grow on the order of $N^{2.5}$, but our simulations have shown that the *k*NN computation times with $k = N/2$ are approximately 40 times faster than minimum half factor computation times for the same sample size.

The data are centered, whitened using the estimated AR model parameters shown in Table 28, and thresholded as described in Section B of this chapter. The transformed data are then used to obtain the minimum subgraphs (minimum half factor, *k*NN) that we use for deriving test statistics and change localization estimates. We compute the statistics for minimum half factor and *k*NN-based tests, apply permutation tests to obtain *p*-values, and signal a change as described in Chapter III.

If a change is detected, the change point is then estimated and the alarm is triggered. The change-point estimate is obtained using the location estimation procedure introduced in Chapter IV Section A, based on the standardized prior edge counts. On our computing platform, detailed in Chapter III Section B, the analyses described in the immediately preceding paragraphs take at most 10 seconds for each new window when using 10,000 permutations, making near-sequential analysis feasible. Our computations were completed on a 64-bit Dell OptiPlex 990 Windows 7 platform with an Intel Core i5-2400 3.1 GHz processor, 8 GB RAM, using R version 3.0.1 with packages cplexAPI (Gelius-Dietrich, 2013) and FNN (Beygelzimer et al., 2013) and using CPLEX version 12.5. We stop the analysis once the algorithm detects change. We conduct a total of thirty tests: at each of the five EWFD locations (10, 11, 12, 13, 14), covering the three reference window sizes (0, 90, 180 seconds), and the two minimum subgraph methods (minimum half factor, *k*NN).

3. Alarm Times and Graphical Results

A summary of the alarm times determined is shown in Table 30. Alarm times are discrete, taking values from 55 seconds to 2,395 seconds in increments of 60 seconds. Ventilation was introduced in the experiment 2,420 seconds after event initiation. Alarm times from the NRL analysis are shown at the bottom of Table 30.

Table 30. Alarm times (and change-point estimates), in seconds after event initiation, by ship location (see Figure 19), duration of pre-stimulus observations, and method of analysis. Minimum half factor refers to forming 100 undirected matches for each of the 200 observations in the analysis window, minimizing the total distance.

Location	Reference Observations (seconds)	Minimum Half Factor Alarm Time and Change Estimate	*k*NN Alarm Time and Change Estimate
10*	none	1,555 (1,465)	1,495 (1,443)
	90	1,555 (1,465)	1,495 (1,443)
	180	1,555 (1,465)	1,495 (1,443)
11*	none	1,015 (927)	1,495 (1,423)
	90	1,015 (927)	1,495 (1,423)
	180	1,315 (1,083)	1,495 (1,423)
12	none	none	none
	90	none	none
	180	none	none
13	none	1,735 (1,609)	1,975 (1,843)
	90	1,735 (1,609)	1,935 (1,843)
	180	1,675 (1,609)	2,155 (2,091)
14	none	none	955 (645)
	90	none	none
	180	none	none

* In the analysis of Gottuk et al. (2002), based on four sensors, Location 10 alarmed at 2,037 seconds and Location 11 alarmed at 2,103 seconds. No other locations alarmed.

Locations 10 and 11, which are closest to the smoldering computer monitor, dominate the earliest alarms. Both graph-theoretic methods produce alarms, although the minimum half factor SPD test does not produce any alarms at Location 14. Location 14 has one alarm from the composite *k*NN test, and it is the first alarm. Overall, including pre-stimulus observations does not positively affect detection in this scenario.

The next two figures expand on the results for Location 10 analyzed with 180 seconds of pre-stimulus data by the minimum half factor SPD test. Figure 28 shows the final window of transformed observation from Location 10. Conducting the minimum half factor SPD test produces an alarm 1,555 seconds after event initiation.

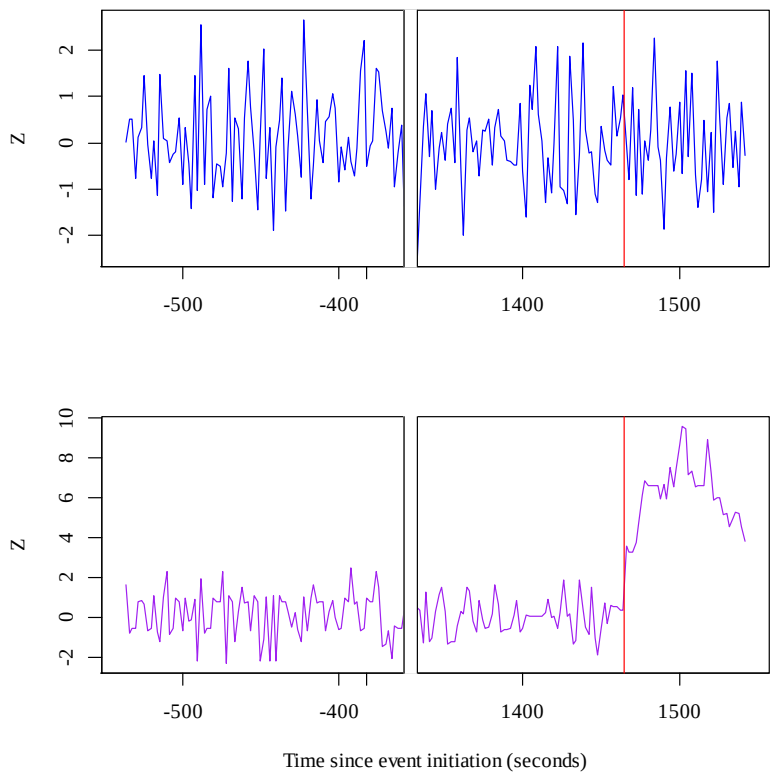

Figure 28. Plot of Location 10 transformed ionization (top) and photoelectric (bottom) sensor observations vs. time since event initiation. The chart includes 180 seconds of pre-stimulus observations. Change is signaled using the minimum half factor sum of pair-differences test. The red vertical line denotes the change point estimated with piecewise regression. The alarm triggers 10 seconds after the raw observations for this window are collected, at 1,555 seconds after event initiation.

Figure 29 shows the piecewise regression on the prior edge count associated with this particular alarm. The optimal break point for the regression is 1,465 seconds after event initiation. This change-point estimate is taken to be the point where the combustion effects were sufficient to justify an alarm. Figures corresponding to the remaining alarms of Table 30 are given in Appendix D.

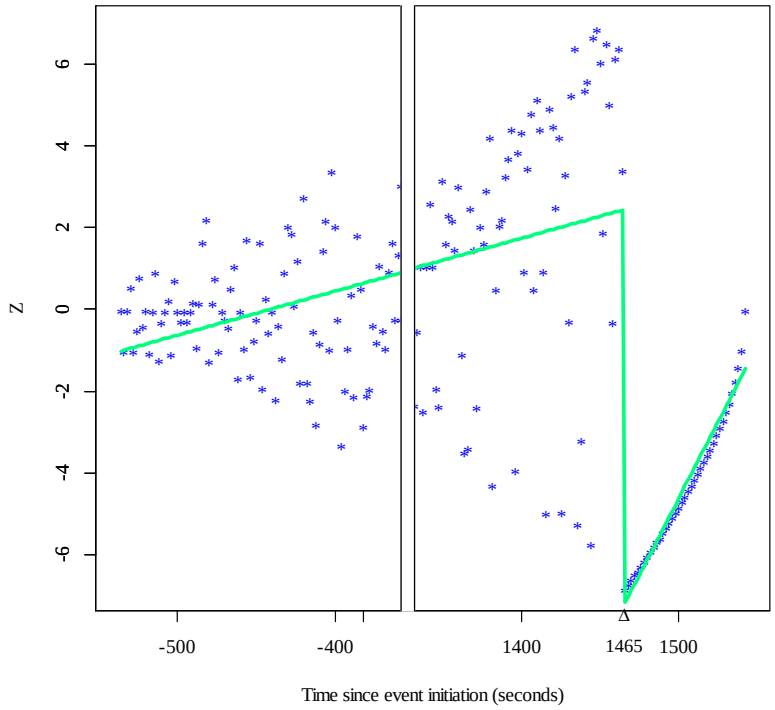

Figure 29. Plot of piecewise regression conducted on the standardized count of minimum half factor prior edges with indices lower than the index point vs. time since event initiation. The estimated change point is 1,465 seconds after event initiation. The chart includes 180 seconds of pre-stimulus observations. Data are from ionization and photoelectric sensors at Location 10.

D. DISCUSSION

The example presented in this chapter demonstrates a real-world application of the graph-theoretic change detection and localization techniques described in Chapters III and IV. Although these methods require data to be statistically independent, respond to change of any nature, and were designed for off-line change detection, we adapt them for this application featuring dependent data, suppression of false alarms, and rapid detection. The presence of a set of control data allows the fitting of a time-series model, based only on current and past observations, that removes much of the temporal dependence. Low-level stimuli that affect the sensors but do not justify the setting of an alarm are handled by filtering with a scatterplot smoother that is truncated at a determined threshold value. And, although we process the data off-line in batches, we do so in a manner that near-real time detectability may be possible. While the NRL analysis

is predicated on training data, our methods have no such requirement. This application also provides assessment of the following:

- Inclusion of Control Data

We find that the decision to either include or exclude an initial run of pre-stimulus observations does not make an appreciable difference in the example we consider. Both minimum half factor and *k*NN alarm times were mostly the same with or without the reference observations, and in two cases, the alarm times were slower when including pre-stimulus observations. It is expected, however, that situations in which change is gradual may benefit from the use of control data.

- Minimum Subgraph Comparison

The minimum half factor method edges the *k*NN method in terms of fastest time to alarm. Of the 15 comparisons between the minimum half factor SPD test and composite *k*NN test shown in Table 30, the minimum half factor signals the first alarm six times, while the composite *k*NN test supplies the first alarm four times.

- Threshold Exploration

In Section B of this chapter, we explain the rationale and process of thresholding for removal of weak signals that do not justify alarms. The purpose of setting the threshold is to appropriately balance the nuisance of false alarms with the risk of missed detection. Figure 30 shows median alarm times obtained with different threshold values (the second horizontal axis value is the threshold value of the illustration). Increasing the threshold has the expected effect of increasing the time to alarm for the minimum half factor SPD test, while the median alarm time of the composite *k*NN test does not change once the lowest threshold is invoked.

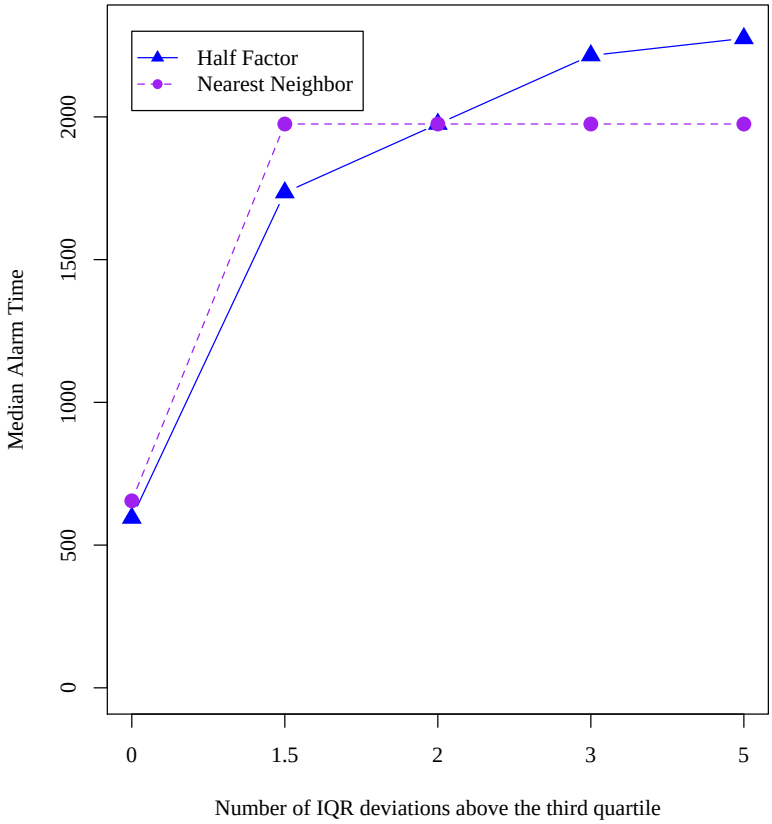

Figure 30. Effects of chosen threshold on the median alarm time for each method (minimum half factor and *k*NN), taking the median across the fifteen tests of each method.

VI. CONCLUSIONS

A. SUMMARY OF MAIN FINDINGS

The need for statistical techniques that detect and localize distributional change in a sequence of multivariate observations continues to grow in modern applications. Graph-theoretic techniques are important members of this class by virtue of their ability to detect change of a general nature without the imposition of strong assumptions about the sampling distribution of the data. The methods proposed in this dissertation compare well with the best parametric procedures when assumptions needed to justify the latter are satisfied. Once change has been detected, we provide a technique for using information from minimum subgraphs to estimate the change point with reasonable accuracy. We also present a real-world example in which apparent limitations to the application of our methods are surmounted.

B. DIRECTIONS FOR FUTURE RESEARCH

There are opportunities for future research in the areas of subgraph formation and usage, change-point localization, dimension reduction, and change-type diagnosis.

1. Subgraph Formation and Usage

Minimum subgraphs and test statistics derived from them that are different from those contained in this dissertation provide fertile ground for exploration. Choosing a different configuration of edge score function and degree of matches may provide increased detection power. Unlike the minimum half factor, which must be completely reformed for each new observation, the *k*NN subgraph easily admits a new observation with less change in the graphical structure (Li, Yang, & Han, 2004). Statistics derived from the *k*NN subgraph can feasibly conduct sequential change detection and localization, which we do not explore in this dissertation.

2. Change-Point Localization

Future research of interest would adapt change-point localization using minimum subgraphs to situations other than a single abrupt change, such as localizing the initiation

of gradual change or allowing for multiple change points. Such an adaptation may also allow for diagnosing non-stationarity (i.e., change initiated at the beginning of the observation sequence) by returning "no change point." This would be the diagnosis for an instrument undergoing continuous drift from the onset of data collection.

3. Dimension Reduction and Localization

The techniques proposed in this dissertation localize the change-point in terms of the observation sequence, but do not localize the variables responsible for the detected distributional change. Can graph-theoretic methods help to localize the change to the changed dimension or dimensions, and/or reduce the number of dimensions to only those of interest? In certain applications such as image analysis or wide-area monitoring, contiguity constraints may allow for reducing the search space. In the manner of Shirabe (2009), a contiguity constraint set restricts change by rejecting the null hypothesis only if the change detected is physically sensible. Such a constraint set defines dimensional adjacencies and requires there to be a path between the dimensions triggering change, as in a network flow problem. Contiguity requirements restrict the dimensional choices, making the problem more manageable.

4. Change-Type Diagnosis

Through use of thresholding applied to fire detection sensor data, we adapted our method to detect only certain types of changes. Other applications may also benefit from knowledge of the type of change detected (location, scale, higher moment). Information from the minimum k-factor and kNN subgraphs may be able to provide diagnostic information regarding the likely nature of the detected change.

APPENDIX A. THE CURSE OF DIMENSIONALITY

There are many meanings that researchers bestow on the phrase the “curse of dimensionality,” but its published origins go back to the early days of dynamic programming. Prior to penning that exact language, RAND Corporation researcher Richard Bellman favored more nautical turns of phrase, which we are obligated to point out as we find ourselves at the *Naval* Postgraduate School. Bellman (1956) dismally notes that “this method founders on the reef of dimensionality” (p. 767). Bellman (1957) catastrophically declares, “…we founder on the shoals of dimensionality” (p. 272). A few years later, no doubt after someone reminded him that he worked at a corporation funded by the U.S. Air Force, Bellman (1961) gamely laments of “…the curse of dimensionality, a malediction that has plagued the scientist from earliest days” (p. 94), although he fails to provide a reference.

THIS PAGE INTENTIONALLY LEFT BLANK

APPENDIX B. ADDITIONAL ENSEMBLE TRACES

We present figures similar to Figure 11 in Chapter III Section C, for the multivariate normal null hypothesis (Figure 31) and mean drift and scale jump (Figure 32) cases. The figures are based on 100 simulations of $N = 200$ observations from a five-dimensional multivariate normal distribution. The thick black line is the critical value. Red traces indicate the ESPM statistic exceeding the critical value based on $\alpha = 0.05$ level of test. Gray traces do not exceed the critical value.

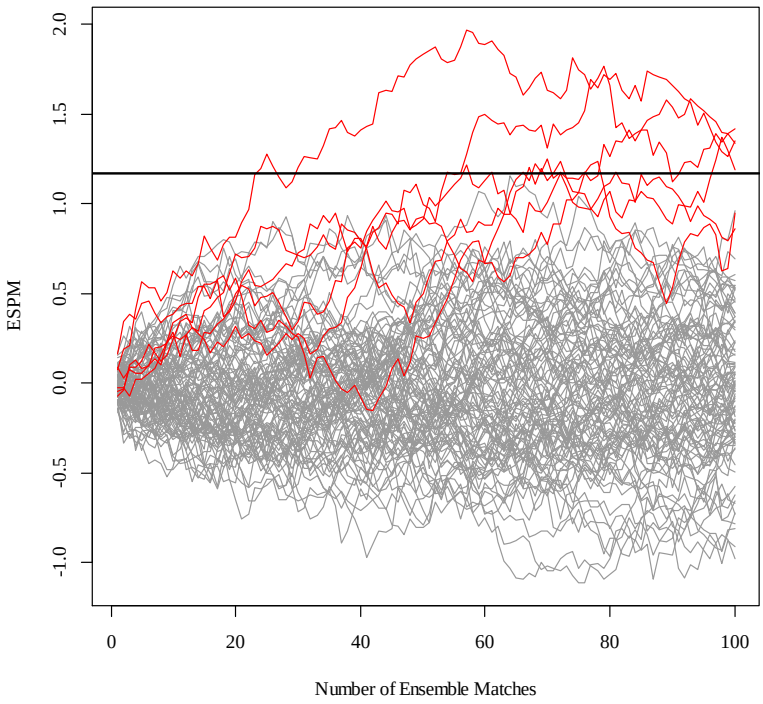

Figure 31. The Ensemble Sum of Pair-Maxima test statistic for observations under the null hypothesis of no change, for matches from 1 to $N/2$.

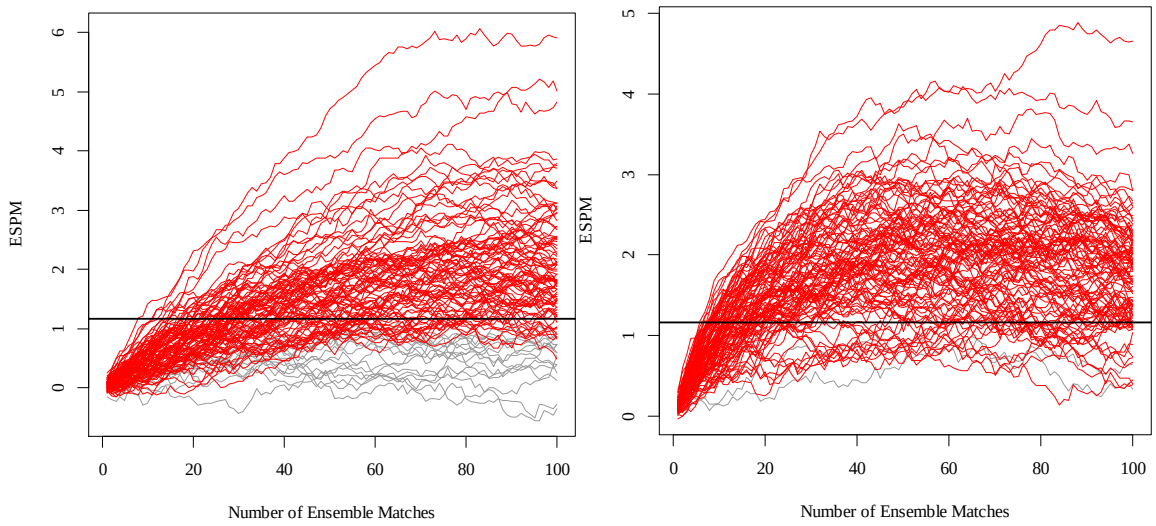

Figure 32. The Ensemble Sum of Pair-Maxima test statistic for observations with a 0.5 magnitude change beginning at observation $\tau = 101$, for matches from 1 to $N/2$. To the left is the case of drift change of magnitude Δ in the first component of the mean vector; to the right is the case of scale jump in all five dimensions.

Considering only the values at $k = 100$, detection power in the mean drift case falls from 0.85 to 0.76. Detection power in the scale jump case falls from 0.98 to 0.81.

APPENDIX C. PRIOR EDGE TABLES

For the standard bivariate normal sample of Chapter II Section A, Table 31 shows the 10 minimum half factor matches of each vertex as well as the prior edge count, its expected value, standard deviation, and standardized value. Table 32 shows the incoming nearest neighbor edge source vertices for each vertex as well as the prior incoming edge count, its expected value, standard deviation, and standardized value.

Table 31. Minimum half factor matches, prior edge count (l_i), expected value, standard deviation, and standardized value for 20 observations drawn from a standard bivariate normal distribution.

Vertex i	Vertices Matched to i										l_i	μ	σ	z_i
1	3	4	5	6	7	11	15	17	18	20	0	0.00	0.00	--
2	3	5	10	11	13	14	15	16	18	19	0	0.53	0.50	−1.05
3	1	2	6	7	8	9	11	15	18	20	2	1.05	0.69	1.38
4	1	5	6	10	12	14	15	16	17	18	1	1.58	0.82	−0.71
5	1	2	4	6	7	10	12	15	17	18	3	2.11	0.91	0.98
6	1	3	4	5	7	11	15	17	18	20	4	2.63	0.98	1.39
7	1	3	5	6	8	9	11	15	18	20	4	3.16	1.04	0.81
8	3	7	9	11	12	13	14	16	19	20	2	3.68	1.08	−1.56
9	3	7	8	11	12	13	14	16	19	20	3	4.21	1.10	−1.10
10	2	4	5	12	13	14	15	16	17	19	3	4.74	1.12	−1.56
11	1	2	3	6	7	8	9	13	19	20	7	5.26	1.12	1.56
12	4	5	8	9	10	13	14	16	17	19	5	5.79	1.10	−0.72
13	2	8	9	10	11	12	14	16	19	20	6	6.32	1.08	−0.29
14	2	4	8	9	10	12	13	16	17	19	7	6.84	1.04	0.15
15	1	2	3	4	5	6	7	10	17	18	8	7.37	0.98	0.64
16	2	4	8	9	10	12	13	14	17	19	8	7.89	0.91	0.12
17	1	4	5	6	10	12	14	15	16	18	9	8.42	0.82	0.71
18	1	2	3	4	5	6	7	15	17	20	9	8.95	0.69	0.08
19	2	8	9	10	11	12	13	14	16	20	9	9.47	0.50	−0.95
20	1	3	6	7	8	9	11	13	18	19	10	10.00	0.00	--

Table 32. Vertex sources for incoming 10-nearest neighbors edges, prior incoming edge count (l_i), expected value, standard deviation, and standardized value for 20 observations drawn from a standard bivariate normal distribution.

Vertex i	Vertices Sending Edges to i																			l_i	μ	σ	z_i
1	3	5	6	7	11	15	17	18	20											0	0	0	0
2	1	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	1	1	0	0
3	1	2	5	6	7	8	9	10	11	14	15	18	19	20						2	1.47	0.61	0.87
4	1	2	3	5	6	10	12	13	14	15	16	17	18	19						3	2.21	0.72	1.1
5	1	4	6	17	18															2	1.05	0.8	1.18
6	1	5	7	15	17	18														2	1.58	0.92	0.46
7	11	20																		0	0.63	0.64	−0.99
8	2	3	7	9	10	11	12	13	14	15	16	18	19	20						3	5.16	0.95	−2.27
9	3	7	8	11	12	13	16	19	20											3	3.79	1.1	−0.72
10	1	2	3	4	5	6	8	9	11	12	13	14	15	16	17	18	19			8	8.05	0.69	−0.08
11	1	3	6	7	8	9	20													6	3.68	1.08	2.15
12	13	14	16																	0	1.74	0.81	−2.15
13	2	4	8	9	10	12	14	16	19	20										6	6.32	1.08	−0.29
14	2	4	5	8	9	10	12	13	16	17	19									8	7.53	1.03	0.46
15	1	2	3	4	5	6	7	8	9	10	11	12	13	14	16	17	18	19	20	14	14	0	0
16	2	4	8	10	12	13	14	17	19											7	7.11	0.91	−0.12
17	1	4	5	6	15	18														5	5.05	0.76	−0.07
18	1	2	3	4	5	6	7	10	15	17										10	8.95	0.69	1.53
19	2	3	4	8	9	10	11	12	13	14	15	16	20							12	12.32	0.46	−0.68
20	7	9	11																	3	3	0	0

For the change case, Table 33 shows the 10 minimum half factor matches of each vertex as well as the prior edge count, its expected value, standard deviation, and standardized value. Table 34 shows the incoming nearest neighbor edge source vertices for each vertex as well as the prior incoming edge count, its expected value, standard deviation, and standardized value.

Table 33. Minimum half factor matches, prior edge count (l_i), expected value, standard deviation, and standardized value for 20 observations drawn from a bivariate normal distribution, with an abrupt change in the mean vector from (0, 0) to (3, 0) starting at observation 11.

Vertex i	Vertices Matched to i										l_i	μ	σ	z_i
1	2	3	4	5	6	7	8	9	10	16	0	0	0	--
2	1	3	4	5	6	7	8	9	10	13	1	0.53	0.50	0.95
3	1	2	4	5	6	7	8	10	18	20	2	1.05	0.69	1.38
4	1	2	3	5	6	7	8	10	11	19	3	1.58	0.82	1.74
5	1	2	3	4	6	7	8	9	10	13	4	2.11	0.91	2.08
6	1	2	3	4	5	7	8	10	19	20	5	2.63	0.98	2.41
7	1	2	3	4	5	6	8	9	10	16	6	3.16	1.04	2.73
8	1	2	3	4	5	6	7	9	10	16	7	3.68	1.08	3.07
9	1	2	5	7	8	10	12	14	15	17	5	4.21	1.10	0.72
10	1	2	3	4	5	6	7	8	9	16	9	4.74	1.12	3.82
11	4	12	13	14	15	16	17	18	19	20	1	5.26	1.12	−3.82
12	9	11	13	14	15	16	17	18	19	20	2	5.79	1.10	−3.43
13	2	5	11	12	14	15	17	18	19	20	4	6.32	1.08	−2.15
14	9	11	12	13	15	16	17	18	19	20	4	6.84	1.04	−2.73
15	9	11	12	13	14	16	17	18	19	20	5	7.37	0.98	−2.41
16	1	7	8	10	11	12	14	15	17	18	8	7.89	0.91	0.12
17	9	11	12	13	14	15	16	18	19	20	7	8.42	0.82	−1.74
18	3	11	12	13	14	15	16	17	19	20	8	8.95	0.69	−1.38
19	4	6	11	12	13	14	15	17	18	20	9	9.47	0.50	−0.95
20	3	6	11	12	13	14	15	17	18	19	10	10	0	--

Table 34. Vertex sources for incoming 10-nearest neighbors edges, prior incoming edge count (l_i), expected value, standard deviation, and standardized value for 20 observations drawn from a bivariate normal distribution, with an abrupt change in the mean vector from (0, 0) to (3, 0) starting at observation 11.

Vertex i	Vertices Sending Edges to i																			l_i	μ	σ	z_i
1	2	3	4	5	6	7	8	9	10	16	19									0	0.00	0.00	--
2	1	3	4	5	6	7	8	9	10	13	16	19	20							1	0.68	0.46	0.68
3	1	2	4	5	6	7	10													2	0.74	0.66	1.91
4	1	2	3	5	6	7	8	10												3	1.26	0.81	2.15
5	1	2	3	4	6	7	8	9	10	16	19									4	2.32	0.9	1.87
6	3	4	7	10																2	1.05	0.8	1.18
7	1	2	3	4	5	6	8	9	10											6	2.84	1.04	3.04
8	1	2	3	4	5	6	7	9	10	11	13	14	15	16	17	18	19	20		7	6.63	0.48	0.76
9	1	2	3	4	5	6	7	8	10	11	12	13	14	15	16	17	18	19	20	8	8.00	0.00	--
10	1	2	3	4	5	6	7	8												8	3.79	1.1	3.81
11	9	12	13	14	15	16	17	18	19	20										1	5.26	1.12	−3.82
12	11	13	14	15	17	18	19	20												1	4.63	1.09	−3.33
13	8	9	11	12	14	15	16	17	18	19	20									4	6.95	1.07	−2.76
14	11	12	13	15	16	17	18	20												3	5.47	1.03	−2.41
15	11	12	13	14	17	18	20													4	5.16	0.95	−1.22
16	1	2	5	8	9	11	12	13	14	15	17	18	19							10	10.26	0.85	−0.31
17	12	14	15																	3	2.53	0.6	0.8
18	9	11	12	13	14	15	16	17	19	20										8	8.95	0.69	−1.38
19	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	20	18	18.00	0.00	--
20	11	12	18																	3	3.00	0.00	--

APPENDIX D. SUPPLEMENTARY APPLICATION FIGURES

This appendix features figures to supplement Chapter V Sections A and B.

A. CHALLENGING DATA

As discussed in Chapter V Section A, the carbon monoxide and carbon dioxide sensor data from this collection were not practically useable. The carbon monoxide data shown on the left in Figure 33 show spiking in the absence of outside stimuli, while the carbon dioxide data shown on the right in Figure 33 show non-stationarity.

The carbon monoxide sensor spiking is attributed to radio frequency interference (RFI). The specific effects of RFI on carbon monoxide sensors are not well documented in literature, but residential carbon monoxide alarms are known to respond to amateur radio transmissions (Low, 2007). We are not relying on post-processing of the data, nor do we consider methods for filtering these RFI spikes. The spiking seems to affect several minutes of data. We could develop a scheme to reject spiking data although, without direct access to these sensors, it is difficult to determine when such a scheme should consider the spike complete. Ultimately, we choose to not use the carbon monoxide data due to the RFI spikes.

The carbon dioxide standardized quantity shows an unexplained increase in value prior to event initiation. Each sensor appears to wander in a different way, increasing or decreasing independently of the other sensors in the absence of any test stimuli. We do not know the reason for this non-stationary behavior. Our methods are not equipped to handle such non-stationarity, so we choose to remove the carbon dioxide sensor from inclusion in further analysis. In our demonstration, we focus on data collected from the ionization and photoelectric sensors.

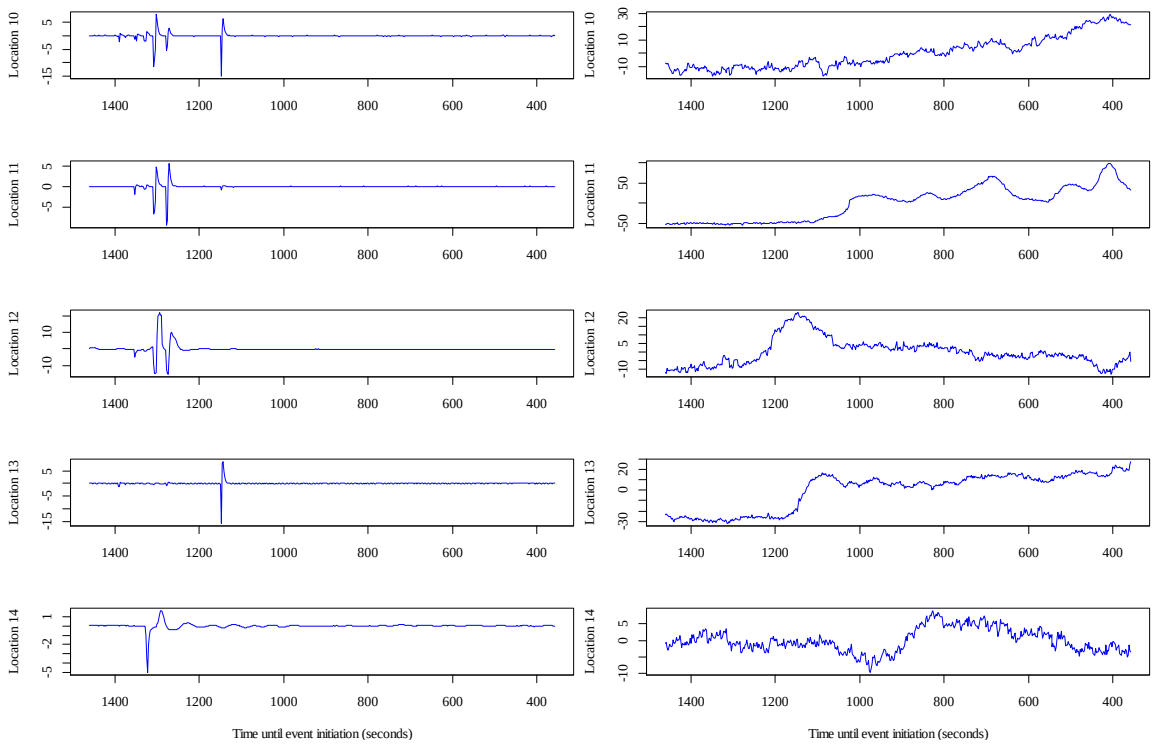

Figure 33. Centered sensor pre-stimulus data for the five third deck EWFD locations. To the left, carbon monoxide measurements, and to the right, carbon dioxide measurements. Each vertical axis is scaled to match the range of the centered sensor response.

B. GRAPHICAL RESULTS

Figures 34 through 43 convey additional results of our trial analogous to Figures 28 and 29 of Chapter V Section C. All figures show the 200-observation window in which change is detected. Figures 34 through 42 show Locations 10, 11, and 13, with minimum half factor results to the left and results from *k*-nearest neighbors to the right. The minimum half factor sum of pair-differences detected no changes at Location 14; we show nearest neighbor results in Figure 43. In all figures, the upper plots show the standardized prior edge count and the fitted piecewise regression line for the window in which a change is detected. The middle plots show the transformed ionization sensor data; the lower plots show the transformed photoelectric sensor data. The red vertical lines denote the change points estimated with piecewise regression for the minimum subgraph. The alarms trigger 10 seconds after the raw observations for these windows are collected.

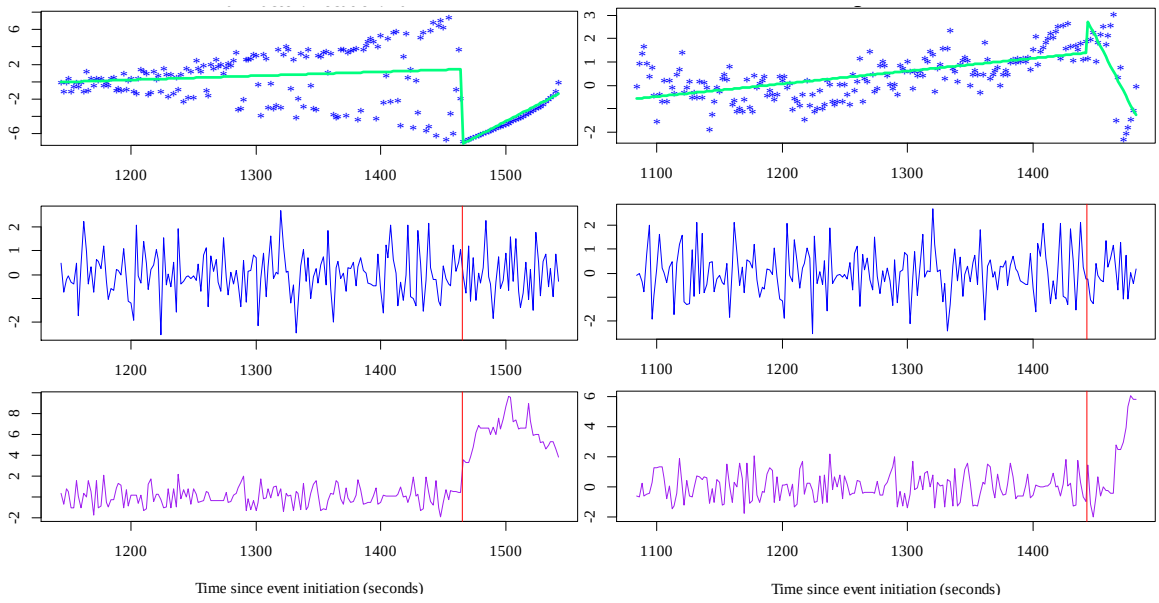

Figure 34. Plots for Location 10 with no pre-stimulus observations. The estimated change points are 1,465 and 1,443 seconds after event initiation; the alarms trigger 1,555 and 1,495 seconds after event initiation.

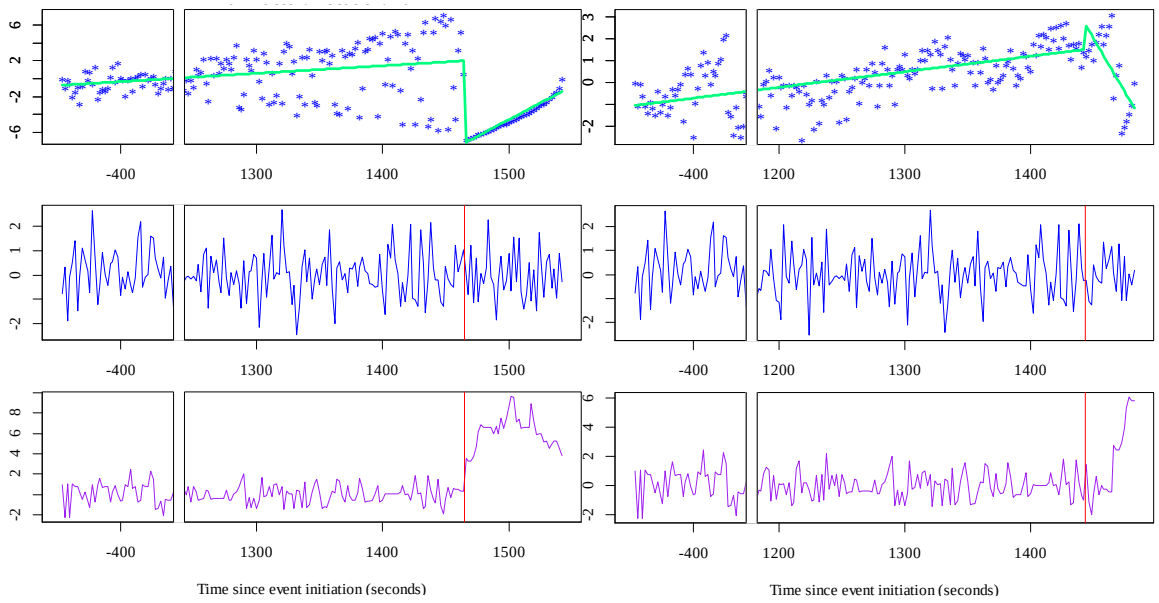

Figure 35. Plots for Location 10 with 90 seconds of pre-stimulus observations. The estimated change points are 1,465 and 1,443 seconds after event initiation; the alarms trigger 1,555 and 1,495 seconds after event initiation.

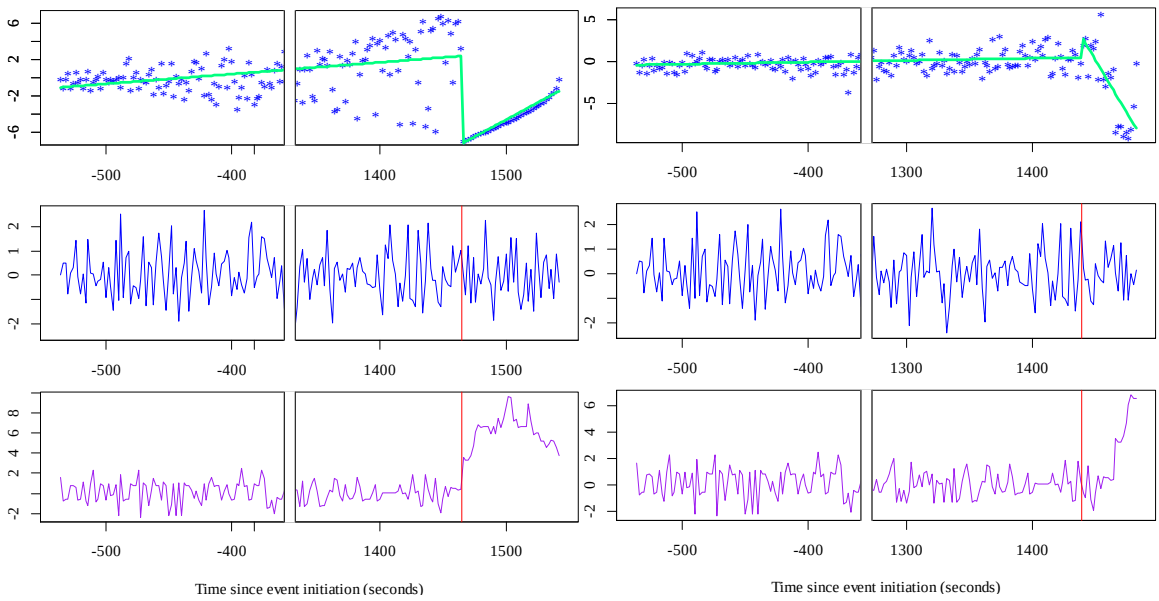

Figure 36. Plots for Location 10 with 180 seconds of pre-stimulus observations. The estimated change points are 1,465 and 1,443 seconds after event initiation; the alarms trigger 1,555 and 1,495 seconds after event initiation.

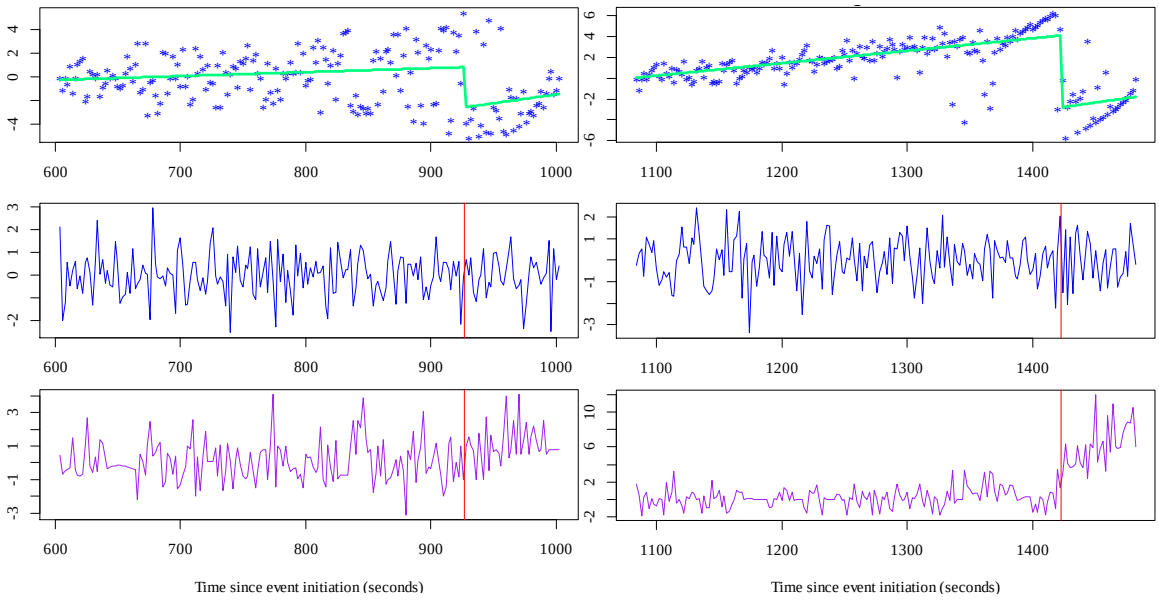

Figure 37. Plots for Location 11 with no pre-stimulus observations. The estimated change points are 927 and 1,423 seconds after event initiation; the alarms trigger 1,015 and 1,495 seconds after event initiation.

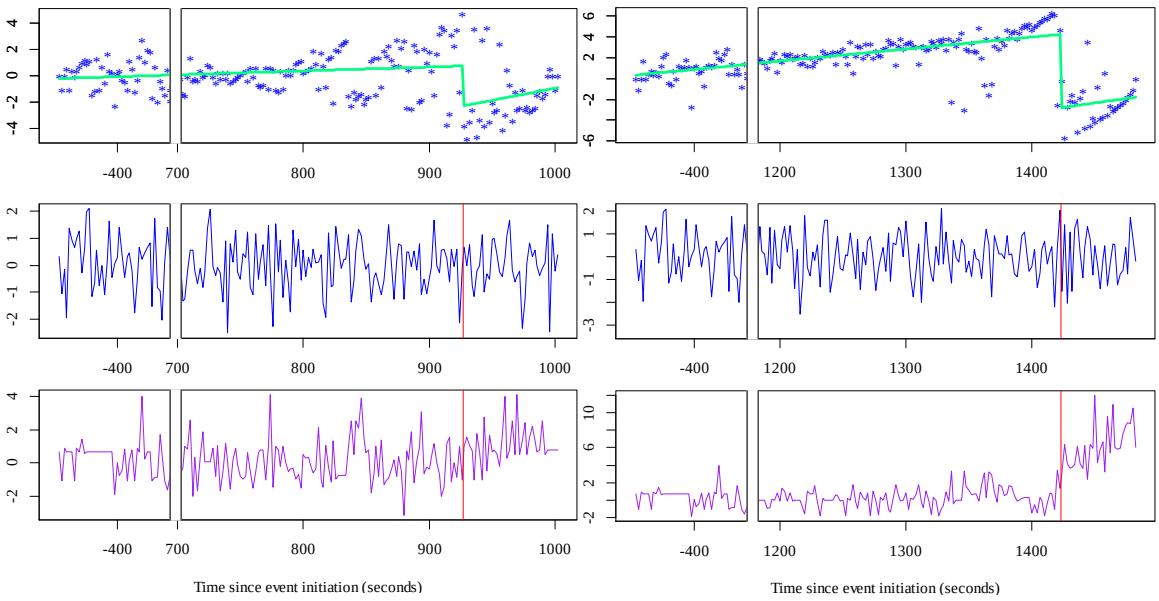

Figure 38. Plots for Location 11 with 90 seconds of pre-stimulus observations. The estimated change points are 927 and 1,423 seconds after event initiation; the alarms trigger 1,015 and 1,495 seconds after event initiation.

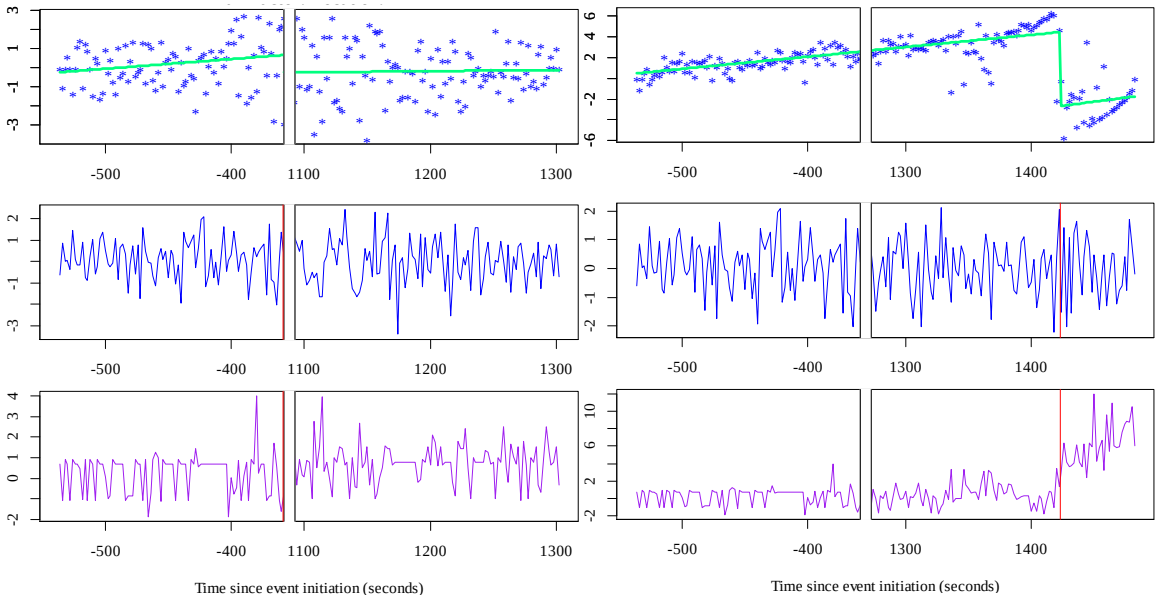

Figure 39. Plots for Location 11 with 180 seconds of pre-stimulus observations. The estimated change points are 1,083 and 1,423 seconds after event initiation; the alarms trigger 1,315 and 1,495 seconds after event initiation.

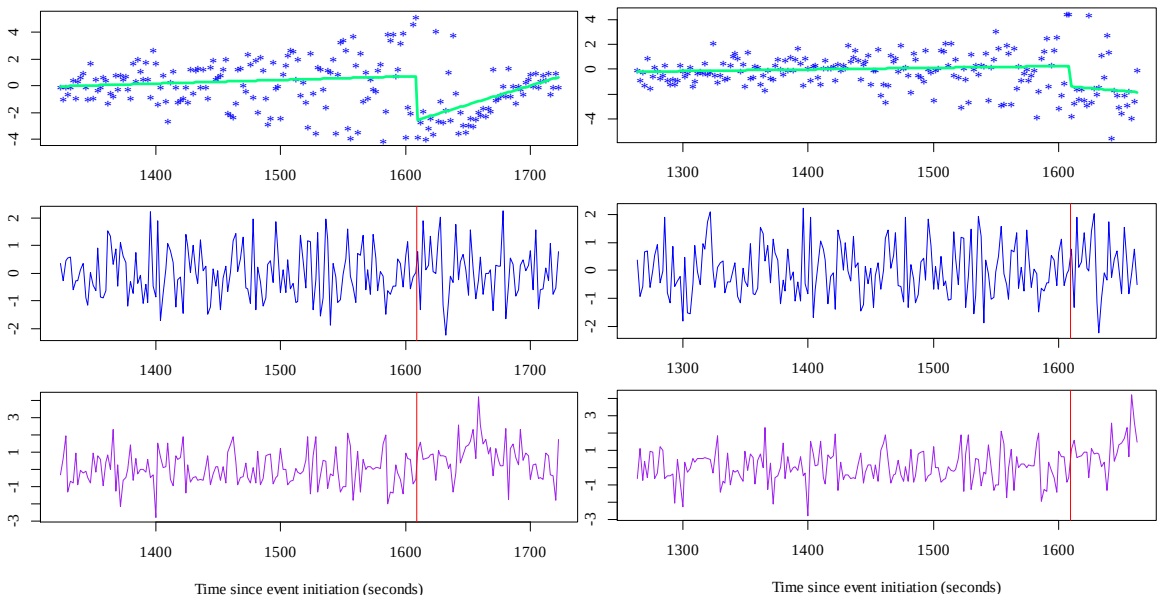

Figure 40. Plots for Location 13 with no pre-stimulus observations. The estimated change points are 1,609 and 1,843 seconds after event initiation; the alarms trigger 1,735 and 1,935 seconds after event initiation.

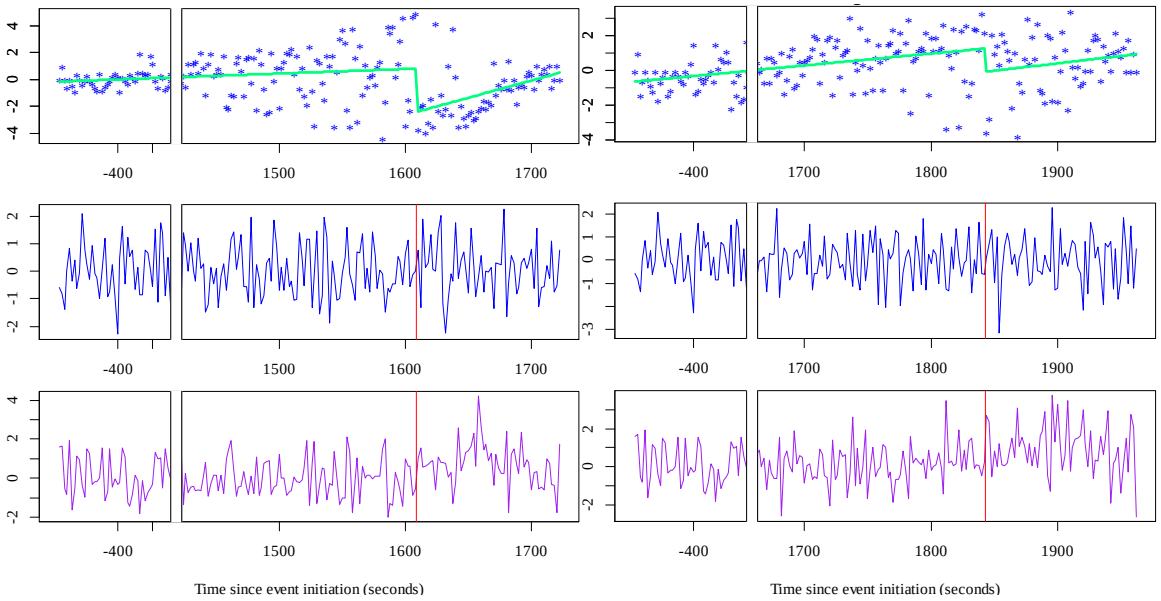

Figure 41. Plots for Location 13 with 90 seconds of pre-stimulus observations. The estimated change points are 1,609 and 1,843 seconds after event initiation; the alarms trigger 1,735 and 1,935 seconds after event initiation.

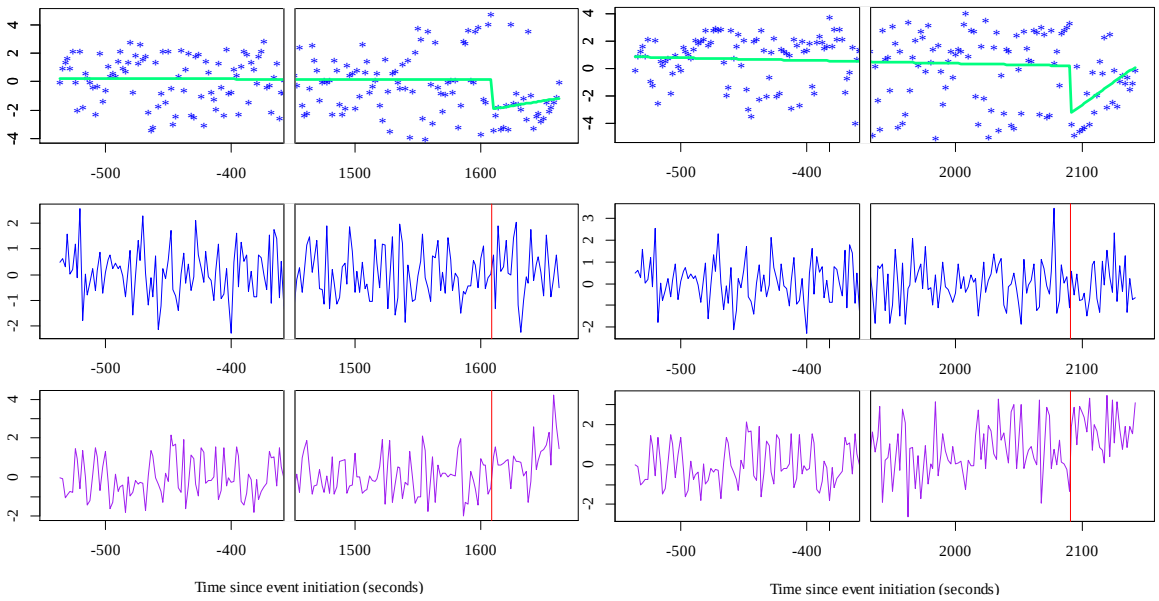

Figure 42. Plots for Location 13 with 180 seconds of pre-stimulus observations. The estimated change points are 1,609 and 2,091 seconds after event initiation; the alarms trigger 1,675 and 2,155 seconds after event initiation.

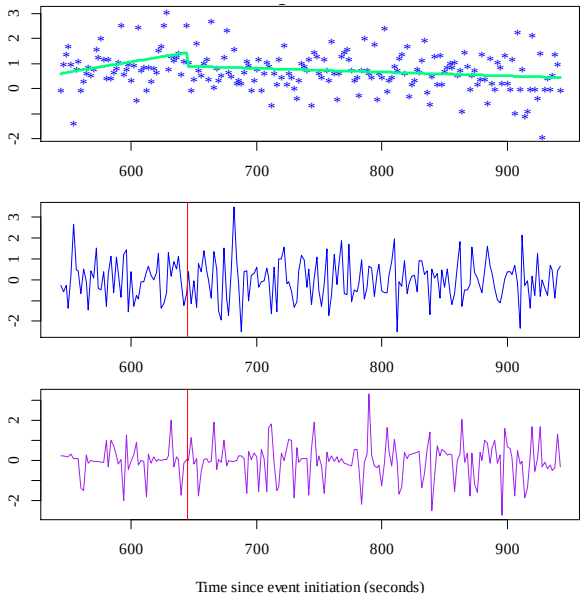

Figure 43. Plots for Location 14 with no pre-stimulus observations. The estimated change point is 645 seconds after event initiation; the alarm triggers 955 seconds after event initiation.

Figures 44 through 48, one for each location, show the transformed data across the entire duration of the test. The graph-theoretic change-point estimates and alarms

based on the graph-theoretic methods are depicted as vertical lines in red and heavy black, respectively. The upper and lower plots show the transformed ionization and photoelectric senor data, respectively. Each vertical axis is scaled to capture the range of the transformed data.

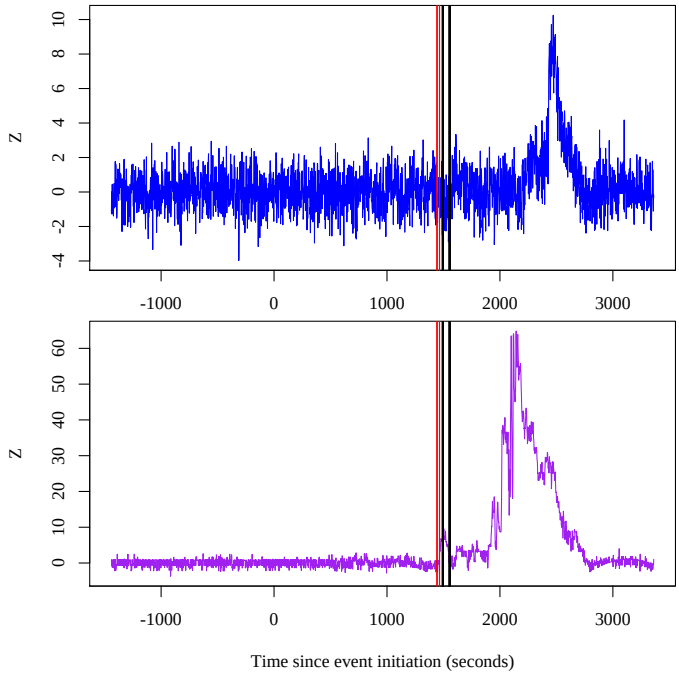

Figure 44. Location 10 transformed measurements, change-point estimates (1,443 and 1,465 seconds after event initiation) and alarms (1,495 and 1,555 seconds after event initiation).

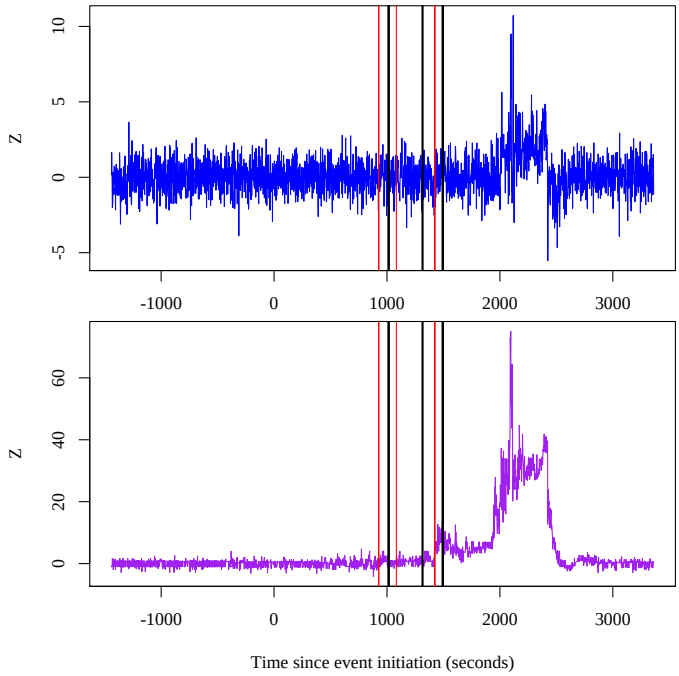

Figure 45. Location 11 transformed measurements, change-point estimates (927, 1,083, and 1,423 seconds after event initiation) and alarms (1,015, 1,315, and 1,495 seconds after event initiation).

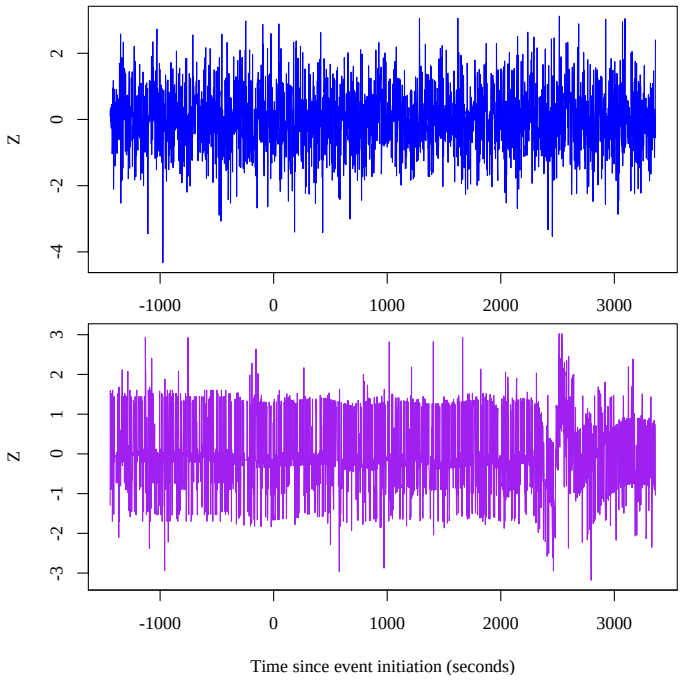

Figure 46. Location 12 transformed measurements. No alarms are triggered.

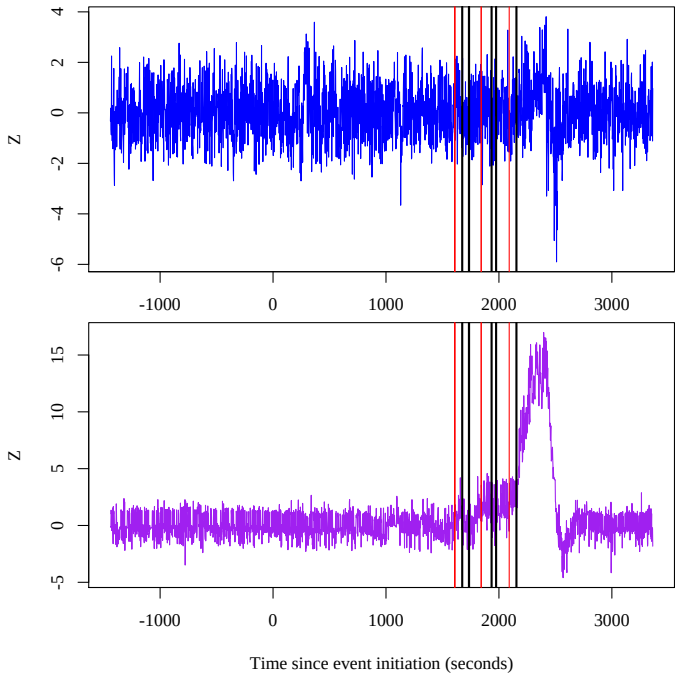

Figure 47. Location 13 transformed measurements, change-point estimates (1,609, 1,843, and 2,091 seconds after event initiation) and alarms (1,675, 1,735, 19,35, 1,975, and 2,155 seconds after event initiation).

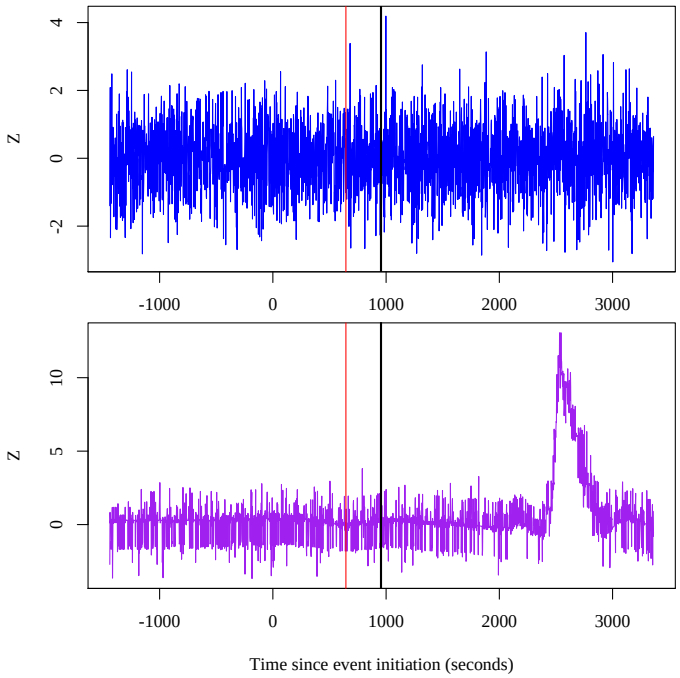

Figure 48. Location 14 transformed measurements, change-point estimate (645 seconds after event initiation) and alarm (955 seconds after event initiation).

LIST OF REFERENCES

Altman, N. S. (1992). An introduction to kernel and nearest-neighbor nonparametric regression. *The American Statistician, 46*(3), 175–185.

Anderson, I. (1972). Sufficient conditions for matchings. *Proceedings of the Edinburgh Mathematical Society (Series 2), 18*(02), 129–136.

Arya, S., Mount, D. M., Netanyahu, N. S., Silverman, R., & Wu, A. Y. (1998). An optimal algorithm for approximate nearest neighbor searching fixed dimensions. *Journal of the ACM (JACM), 45*(6), 891–923.

Bellman, R. E. (1956). Dynamic programming and Lagrange multipliers. *Proceedings of the National Academy of Sciences of the United States of America, 42*(10), 767.

Bellman, R. E. (1957). On a Dynamic Programming Approach to the Caterer Problem-I. *Management Science, 3*(3), 270–278.

Bellman, R. E. (1961). *Adaptive control processes: A guided tour* (Vol. 4). Princeton, NJ: Princeton University Press, 94–95.

Beyer, K., Goldstein, J., Ramakrishnan, R., & Shaft, U. (1999). When is "nearest neighbor" meaningful? In *Database Theory—ICDT'99*. Berlin: Springer, 217–235.

Beygelzimer, A., Kakadet, S., Langford, J., Arya, S., Mount, D., & Li, S. (2013). FNN: Fast Nearest Neighbor search algorithms and applications. R package version 1.1.

Bhattacharya, P. K. (1994). Some aspects of change-point analysis. *Lecture Notes-Monograph Series, 23*, 28–56.

Bhattacharyya, M. N., & Layton, A. P. (1979). Effectiveness of seat belt legislation on the Queensland road toll—an Australian case study in intervention analysis. *Journal of the American Statistical Association, 74*(367), 596–603.

Bickel, P. J., & Krieger, A. M. (1992). Extensions of Chebychev's inequality with applications. *Probability and Mathematical Statistics, 13*, 293–310.

Biswas, M., Mukhopadhyay, M., & Ghosh, A. K. (2014). A distribution-free two-sample run test applicable to high-dimensional data. *Biometrika, 101*(4), 913–926.

Bodgan, C. E. (2014). *Applications of Graph-Theoretic Tests to Online Change Detection* (No. USNA-TSPR-424). Naval Academy Annapolis MD.

Box, G. E., & Cox, D. R. (1964). An analysis of transformations. *Journal of the Royal Statistical Society. Series B (Methodological)*, 211–252.

Box, G. E., & Jenkins, G. M. (1976). *Time series analysis: Forecasting and control*. San Francisco: Holden-Day Inc., 9, 23–33, 53–66.

Brodsky, E., & Darkhovsky, B. S. (1993). *Nonparametric methods in change point problems* (No. 243). Springer Science & Business Media, 11.

Brodsky, E., & Darkhovsky, B. S. (2000). *Nonparametric statistical diagnosis: problems and methods* (Vol. 509). Dordrecht, The Netherlands: Springer Science & Business Media, 83–126.

Chen, H., & Friedman, J. H. (2015). *A new graph-based two-sample test for multivariate and object data*. Retrieved from http://arxiv.org/pdf/1307.6294v5.pdf

Chen, H., & Zhang, N. (2015). Graph-based change-point detection. *The Annals of Statistics, 43*(1), 139–176.

Chen, J., & Gupta, A. K. (2001). On change point detection and estimation. *Communications in Statistics-Simulation and Computation, 30*(3), 665–697.

Chen, J., & Gupta, A. K. (2011). *Parametric statistical change point analysis: With applications to genetics, medicine, and finance.* New York: Springer Science & Business Media, 89–94.

Cleveland, W. S., & Grosse, E. (1991). Computational methods for local regression. *Statistics and Computing, 1*(1), 47–62.

Crawley, M. J. (2007). Regression. *The R book, Second Edition*. West Sussex, England: John Wiley & Sons, 425–430.

Crosier, R. B. (1988). Multivariate generalizations of cumulative sum quality-control schemes. *Technometrics, 30*(3), 291–303.

Darkhovsky, B. S. (1994). Nonparametric methods in change-point problems: A general approach and some concrete algorithms. *Lecture Notes-Monograph Series*, *23*, 99–107.

Devore, J. (2011). *Probability and statistics for engineering and the sciences*. Belmont, CA: Cengage Learning, 37.

Dries, A., & Rückert, U. (2009). Adaptive concept drift detection. *Statistical Analysis and Data Mining, 2*(5-6), 311–327.

Faraway, J. J. (2005). *Linear models with R*. Boca Raton, FL: CRC Press, 110.

Friedman, J. H., & Rafsky, L. C. (1979). Multivariate generalizations of the Wald-Wolfowitz and Smirnov two-sample tests. *The Annals of Statistics*, 697–717.

Gelius-Dietrich, G. (2013). cplexAPI: R Interface to C API of IBM ILOG CPLEX. R package version 1.2.9.

Good, P. I. (2005). *Permutation, parametric and bootstrap tests of hypotheses* (Vol. 3). New York: Springer, 34–36.

Gottuk, D. T., Wright, M. T., Wong, J. T., Pham, H. V., Rose-Pehrsson, S. L., Hart, S. … Street, T. T. (2002). *Prototype early warning fire detection system: test series 4 results* (No. NRL/MR/6180--02-8602). Washington, DC: Naval Research Lab.

Hall, P., & Tajvidi, N. (2002). Permutation tests for equality of distributions in high-dimensional settings. *Biometrika, 89*(2), 359–374.

Henze, N. (1988). A multivariate two-sample test based on the number of nearest neighbor type coincidences. *The Annals of Statistics*, 772–783.

Holland, M. D., & Hawkins, D. M. (2014). A control chart based on a nonparametric multivariate change-point model. *Journal of Quality Technology, 46*(1), 63–77.

Hotelling, H. (1947). Multivariate quality control illustrated by the air testing of sample bomb sights. *Techniques of Statistical Analysis*, 111–184.

IBM (2013). ILOG CPLEX Optimization Studio. http://www-01.ibm.com/software/commerce/optimization/cplex-optimizer/index.html

Ion, R.A. (2001). *Nonparametric statistical process control*. Ph.D. Thesis, University of Amsterdam, The Netherlands, 61–78.

James, B., James, K. L., & Siegmund, D. (1992). Asymptotic approximations for likelihood ratio tests and confidence regions for a change-point in the mean of a multivariate normal distribution. *Statistica Sinica, 2*(1), 69–90.

JiJi, R. D., Hammond, M. H., Williams, F. W., & Rose-Pehrsson, S. L. (2003). Multivariate statistical process control for continuous monitoring of networked early warning fire detection (EWFD) systems. *Sensors and Actuators B: Chemical, 93*(1), 107–116.

Khodadadi, A., & Asgharian, M. (2008). Change-point problem and regression: an annotated bibliography. *COBRA Preprint Series, Paper, 44*.

Kolmogorov, V. (2009). Blossom V: a new implementation of a minimum cost perfect matching algorithm. *Mathematical Programming Computation, 1*(1), 43–67.

Kotsiantis, S. B. (2007). Supervised Machine Learning: A Review of Classification Techniques. *Informatica, 31*, 249–268.

Koul, H. L., & Qian, L. (2002). Asymptotics of maximum likelihood estimator in a two-phase linear regression model. *Journal of Statistical Planning and Inference, 108*(1), 99–119.

Lee, T. S. (2010). Change-point problems: bibliography and review. *Journal of Statistical Theory and Practice, 4*(4), 643–662.

Li, J., Zhang, X., & Jeske, D. R. (2013). Nonparametric multivariate CUSUM control charts for location and scale changes. *Journal of Nonparametric Statistics, 25*(1), 1–20.

Li, Y., Yang, J., & Han, J. (2004, June). Continuous k-nearest neighbor search for moving objects. In *Scientific and Statistical Database Management, 2004. Proceedings. 16th International Conference on*. IEEE, 123–126.

Low, R. (2007). Re: [RFI] CO detector [Online forum comment]. Retrieved from Contesting website: http://lists.contesting.com/_rfi/2007-02/msg00043.html

Lowry, C. A., Woodall, W. H., Champ, C. W., & Rigdon, S. E. (1992). A multivariate exponentially weighted moving average control chart. *Technometrics, 34*(1), 46–53.

Lung-Yut-Fong, A., Lévy-Leduc, C., & Cappé, O. (2012). *Homogeneity and change-point detection tests for multivariate data using rank statistics*. Retrieved from http://arxiv.org/pdf/1107.1971v3.pdf

Matteson, D. S., & James, N. A. (2014). A nonparametric approach for multiple change point analysis of multivariate data. *Journal of the American Statistical Association, 109*(505), 334–345.

National Fire Protection Association (NFPA) Task Group on Smoke Detector Technology. (2008). *Minimum Performance Requirements for Smoke Alarm Detection Technology*. Quincy, MA: National Fire Protection Association Technical Committee on Single- and Multiple-Station Alarms and Household Fire Alarm Systems. Retrieved from http://www.nfpa.org/safety-information/for-consumers/fire-and-safety-equipment/smoke-alarms/ionization-vs-photoelectric

Preuss, P., Puchstein, R., & Dette, H. (2014). Detection of multiple structural breaks in multivariate time series. *Journal of the American Statistical Association*, (accepted). Retrieved from http://dx.doi.org/10.1080/01621459.2014.920613

Qui, P., & Hawkins, D. (2003). A nonparametric multivariate cumulative sum procedure for detecting shifts in all directions. *Journal of the Royal Statistical Society: Series D (The Statistician), 52*(2), 151–164.

R Core Team (2013). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. http://www.R-project.org/

Rosenbaum, P. R. (2005). An exact distribution-free test comparing two multivariate distributions based on adjacency. *Journal of the Royal Statistical Society: Series B (Statistical Methodology), 67*(4), 515–530.

Ruth, D. M. (2009). *Applications of Assignment Algorithms to Nonparametric Tests for Homogeneity* (Ph.D. dissertation, Naval Postgraduate School, Monterey, CA).

Ruth, D. M., & Koyak, R. A. (2011). Nonparametric tests for homogeneity based on non-bipartite matching. *Journal of the American Statistical Association, 106*(496), 1615–1625.

Schilling, M. F. (1986). Multivariate two-sample tests based on nearest neighbors. *Journal of the American Statistical Association, 81*(395), 799–806.

Seneta, E. W. (n.d.). Carlo Emilio Bonferroni (version 6). In *StatProb: The Encyclopedia Sponsored by Statistics and Probability Societies*. Retrieved March 19, 2015 from http://statprob.com/encyclopedia/CarloEmilioBONFERRONI.html

Shirabe, T. (2009). Districting modeling with exact contiguity constraints. *Environment and planning. B, Planning & design, 36*(6), 1053.

Sorkin, R. D., & Woods, D. D. (1985). Systems with human monitors: A signal detection analysis. *Human-Computer Interaction, 1*(1), 49–75.

Tran, D. H., Gaber, M. M., & Sattler, K. U. (2014). Change detection in streaming data in the era of big data: models and issues. *ACM SIGKDD Explorations Newsletter, 16*(1), 30–38.

Vaidya, P. M. (1989). An $O(n \log n)$ algorithm for the all-nearest-neighbors problem. *Discrete & Computational Geometry, 4*(1), 101–115.

Wackerly, D., Mendenhall, W., & Scheaffer, R. (2008). *Mathematical statistics with applications*. Belmont, CA: Cengage Learning, 126–127, 208, 401.

Wald, A., & Wolfowitz, J. (1940). On a test whether two samples are from the same population. *The Annals of Mathematical Statistics, 11*(2), 147–162.

West, D. B. (2001). *Introduction to graph theory* (Vol. 2). Upper Saddle River: Prentice Hall, 2–9, 53, 67–97, 107–145.

Woodall, W. H., & Ncube, M. M. (1985). Multivariate CUSUM quality-control procedures. *Technometrics, 27*(3), 285–292.

Zhou, C., Zou, C., Zhang, Y., & Wang, Z. (2009). Nonparametric control chart based on change-point model. *Statistical Papers, 50*(1), 13–28

Zhou, M., Zi, X., Geng, W., & Li, Z. (2014). A distribution-free multivariate change-point model for statistical process control. *Communications in Statistics-Simulation and Computation*, (accepted). Retrieved from http://dx.doi.org/10.1080/03610918.2013.835411

INITIAL DISTRIBUTION LIST

1. Defense Technical Information Center
   Ft. Belvoir, Virginia

2. Dudley Knox Library
   Naval Postgraduate School
   Monterey, California