View metadata, citation and similar papers at core.ac.uk
brought to you by CORE
provided by Clemson University: TigerPrints

Clemson University
TigerPrints

All Dissertations | Dissertations

5-2012

Clustering and Classification of Multivariate Stochastic Time Series in the Time and Frequency Domains

Ryan Schkoda
Clemson University, ryan.schkoda@alumni.clemson.edu

Follow this and additional works at: https://tigerprints.clemson.edu/all_dissertations

Part of the Mechanical Engineering Commons

Recommended Citation
Schkoda, Ryan, "Clustering and Classification of Multivariate Stochastic Time Series in the Time and Frequency Domains" (2012). *All Dissertations*. 907.
https://tigerprints.clemson.edu/all_dissertations/907

This Dissertation is brought to you for free and open access by the Dissertations at TigerPrints. It has been accepted for inclusion in All Dissertations by an authorized administrator of TigerPrints. For more information, please contact kokeefe@clemson.edu.

Clustering and Classification of Multivariate Stochastic Time Series in the Time and Frequency Domains

---

A Dissertation
Presented to
the Graduate School of
Clemson University

In Partial Fulfillment
of the Requirements for the Degree
Doctor of Philosophy
Mechanical Engineering

---

by
Ryan F. Schkoda
May 2012

Accepted by:
Dr. John Wagner, Committee Chair
Dr. Robert Lund
Dr. Ardalan Vahidi
Dr. Darren Dawson

Abstract

The dissertation primarily investigates the characterization and discrimination of stochastic time series with an application to pattern recognition and fault detection. These techniques supplement traditional methodologies that make overly restrictive assumptions about the nature of a signal by accommodating stochastic behavior. The assumption that the signal under investigation is either deterministic or a deterministic signal polluted with white noise excludes an entire class of signals – stochastic time series. The research is concerned with this class of signals almost exclusively. The investigation considers signals in both the time and the frequency domains and makes use of both model-based and model-free techniques.

A comparison of two multivariate statistical discrimination techniques, one based on a traditional covariance statistic and one based on a more recently proposed periodogram based statistic, is carried out through simulation study. This investigation validates the utility of the periodogram based statistic over the covariance based statistic. The periodogram based statistic proves more useful in identifying statistical dissimilarities in multidimensional time series than the more traditional statistic.

Attention is then focused on using the periodogram based statistic as a distance measure for clustering and classifying time series, which is motivated by the periodogram method's increased discrimination capability. The test statistic is used in both clustering and classification algorithms, and the performance is evaluated though a simulation study. This measure proves capable of grouping like series together while simultaneously separating dissimilar series from one another.

Finally, the techniques are adapted to the time-domain where they are used to cluster multidimensional, non-stationary, climatological data. The non-stationary model accounts for seasonal means, seasonal standard deviations, and stochastic components. The statistical approach results in the development of a level-α test for assessing signal equality. This improves upon typical dendrogram techniques by defining a level under which the distance should be considered zero. Climatological time series from the west coast, Gulf of Mexico, and east coast are analyzed using the aforementioned techniques.

To complement the time series analysis work, some effort (Appendix A) is focused on improving the bachelor of science in the department of mechanical engineering via the undergraduate laboratories. This is accomplished by identifying desired outcomes and implementing specific improvements in the undergraduate laboratory courses over a period of four years. The effects of these improvements are quantified with survey results. Overall, the improvements are very well received and result in significant increases in student satisfaction.

Dedication

This work is dedicated to my mother, Cindy Schkoda.

Acknowledgments

I must acknowledge the work of Dr. Robert Lund and Nan Su from the Department of Mathematical Sciences at Clemson University for their assistance in performing the research presented in Chapter 3. Their research efforts were supported by NSF Grant DMS 0905570.

I would like to recognize the mutual support of many good friends and colleges over the years. A short and incomplete list of these individuals includes Andrew R. Streett, Dr. Steven M. Bower, Joseph Olles, Gregory Penoyer and Clodoaldo Silva. Thank you.

Table of Contents

List of Tables

List of Figures

Chapter 1

Introduction

Getting the *most* out of a device or system is a lofty but noble goal and increasing its efficiency has long been an engineering objective. To this end, one would like to maximize the number of working hours for a device before it eventually fails or needs repair. This particular task is the subject of system health monitoring.

The occurrence of fault means the device or system is not operating properly. This improper operation mean inefficiency and the potential for further damage to the larger system. Fault diagnosis seeks to identify whether or not a fault has occurred. And assuming that a slowly progressing fault has occurred, prognostics seeks to determine how much longer the device can stay in service before system shutdown and part replacement is necessary. A general description for system health monitoring, a statistical approach and the necessary tools are presented herein.

1.1 System Health Monitoring

Effective fault detection and diagnosis allows for efficient operation/use of devices and systems. Efficient use translates into cheaper operating costs and extended useful life. Modern day diagnostic capacity has been bolstered by the current state of data storage devices, data transport devices and data processing capabilities. As a result, the field of fault detection and diagnosis is expanding into areas that were once inaccessible.

1.2 Context

From a global perspective, fault detection exists on three levels and is achieved by employing techniques belonging to one of two broad categories [1, 2]. Level 1 is detection. The most basic goal of any fault detection scheme is to identify if a fault is in fact present. This is accomplished by any number of techniques but is essentially a change detection. Assuming the original state or behavior of a system is optimal and any deviation from that is decremental is key for justifying change detection as a means of fault detection. The second level is isolation. Successfully isolating a fault means determining what signal, sensor, or parameter estimate has deviated from nominal. Finally, the last level is identification. Identification is determining the size and nature of a particular fault and doing so allows operators to take necessary actions to clear the fault.

Achieving any one of these three levels of fault detection and isolation requires a strategy or methodology of analyzing the system under investigation. The literature recognizes two broad classes of techniques, model-based and model-free. The model-based techniques employ some type of mathematical model (linear, nonlinear, deterministic or stochastic) of the system being investigated and use this a basis of comparison during the diagnostic test. This model will be created based on either physical principals or some understanding about the dynamics of the system. Again, once a model has been produced it represents healthy and expected operation of the system.

The diagnostic determination will be based on the analysis of residuals. Residuals are the difference between system output and model output when subjected to the same input. This concept is illustrated in Figure 1.1.

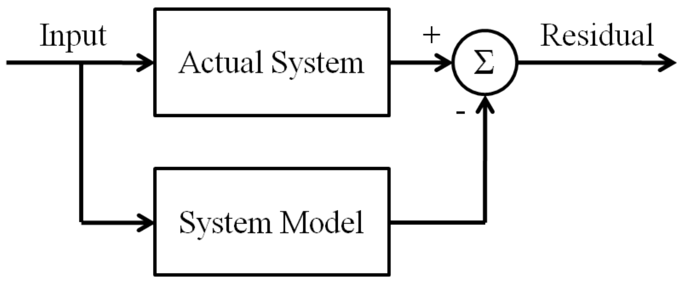

Figure 1.1: Illustration of how residuals are generated. The system model is subjected to the same inputs as the actual system. The difference between the two outputs make up the residuals.

Model-free techniques, as the name suggests, do not make use of a explicit model as part of the fault detection and diagnosis task. Instead these techniques process the signal coming from the system under investigation and compare it to either past or reference operation data (see Figure 1.2). Certain characteristic features of the signal, or simply *features*, are extracted from the raw signal and used as the basis of comparison to the reference data. The particular features being used will depend heavily on the system being analyzed but can include magnitude, variance, frequency content, rate of change etc.

It should be noted that there is some overlap between these two broad categories. For instance, depending on your particular definition of “model” a technique may fall into either the model-based or the model-free category. For example, consider a fully trained neural network that produces the expected output based on a general system input. The output of the neural network is then compared to the actual system output to make decisions regarding system health.

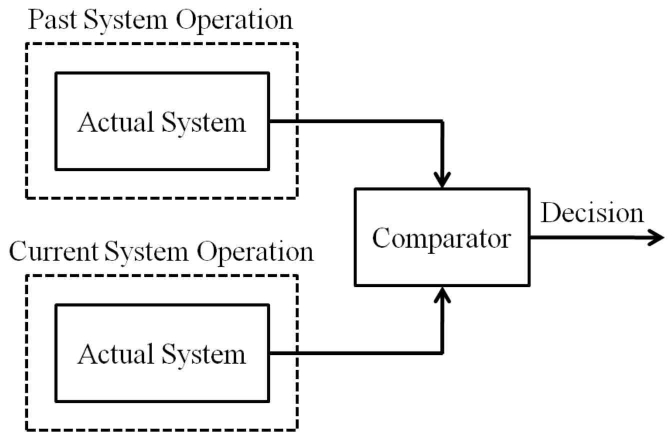

Figure 1.2: Illustration of the model free diagnostic approach. Current system output is compared to past (or reference) output to make a decision about system health.

At first, this may appear to be a model based technique since the neural network is producing outputs (based on measured inputs) that are then being used in a residual analysis. However, one could argue that because the neural network's training procedure requires large amounts of past operating data, this is an example of a data-driven or model-free approach.

The dissertation considers both model-free and model-based techniques. Signals are evaluated in the frequency domain by analyzing the statistical behavior or spectral estimators and signals are evaluated in the time domain by analyzing prediction errors. The frequency domain analysis does not make use of a particular model while the time domain analysis makes use of a non-stationary model.

1.3 Time Series Analysis

In the field of mechanical engineering, it is extremely common, almost taken for granted, that dynamic systems exhibit time history dependency. This dependency is related to the concept of causality and accounts for the fact that the current state of a system depends on past inputs to the systems and the past state of the system as opposed to future inputs or future states of the system. This time dependency is not necessary but is, again, extremely common. Most any dynamic system whose independent variable is time possesses this causal, time dependence.

In other fields, such as statistics, this time dependence may not be as common. Certain ares of statistics (such as engineering statistics) are concerned with Independent Identically Distributed (IID) random variables or at the very least random variables that are uncorrelated. Under these conditions, realizations of the random variables are independent of (uncorrelated with) past or future realizations. A classic example of an IID random variable is the roll of a fair die. The outcome of a roll is in no way dependent on the outcome of the previous roll nor is it dependent on rolls that are yet to come.

Time series analysis is the breach of mathematical statistics concerned with stochastic processes whose independent variable is time. A time series model seeks to describe the dynamic behavior of a stochastic process. The breadth of this field quite large and there are time series models for many different stochastic processes including Autoregressive (AR), Moving Average (MA), Autoregressive Moving Average (ARMA), Autoregressive Integrative Moving Average (ARIMA) and models that account for heteroskedasticity such as the Autoregressive Conditional Heteroskedasticity (ARCH) as well as a host of nonlinear time series models.

This area of statistical analysis allows for the modeling of stochastic process that exhibit a time dependency. This research considers the AR, MA and ARMA family of models as well as a narrow class of nonlinear time series models. And as will be shown in the following section, these models are closely related to engineering systems.

1.4 Deterministic Versus Stochastic Processes

There is an important distinction between a stochastic process and a deterministic process. Many engineering disciplines consider deterministic systems and their associated inputs and responses. Whether those functions are continuous (sine, cosine, exponential), piecewise-continuous (ramp function) or discontinuous (step function) does not matter. Regardless of which deterministic function is used as an input to a dynamic system, applying the same input will result in the same (determined) output. However, if the input to a dynamic system is stochastic the output will be stochastic as well and is generally not repeatable.

White noise is perhaps the most basic stochastic process. However, white noise may be used drive an ARMA model to produce an extremely wide range of stochastic series. A series $\{X_t\}$ is said to be an ARMA(p, q) process if $\{X_t\}$ is stationary and for every t,

$$X_t - \phi_1 X_{t-1} - \cdots - \phi_p X_{t-p} = Z_t + \theta_1 Z_{t-1} + \cdots + \theta_q Z_{t-q} \tag{1.1}$$

where $\{Z_t\} \sim \mathrm{WN}(0, \sigma^2)$ [3]. Now $\{X_t\}$ is a time series whose covariance will depend on the values of the parameters ϕ and θ.

Interestingly, these ARMA models are discrete analogs of continuous time LTI systems like those represented with transfer functions. The takeaway is that the terms *deterministic system* and *stochastic system* are somewhat misleading. More appropriate terminology would be deterministic and stochastic *signal*, where signal is referring to the output of the dynamic system. The nature of the system alone is not enough to completely characterize the system response.

The following example is meant to highlight the similarities between a continuous-time differential equation model and a discrete-time difference equation. Also, to highlight what effect the nature of the input can have on the output.

Consider the following system model,

$$5\ddot{x} + 3\dot{x} + 7x = 7f(t) \tag{1.2}$$

with corresponding transfer function

$$TF(s) = \frac{X}{F}(s) = \frac{7}{5s^2 + 3s + 7} \tag{1.3}$$

representing a second order, under damped system. Figure 1.3 shows the response of such a system when the input is a pure sinusoid. Notice the regular, continuous nature of the response.

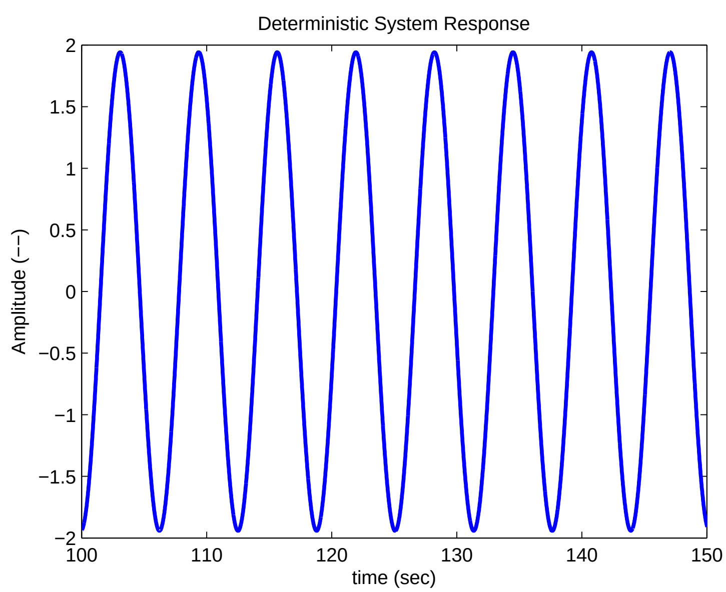

Figure 1.3: Continuous dynamic system response to a sinusoidal input.

Equation 1.4 shows the result of discretizing the system assuming a zero order hold and a step size of $T = 0.1$ to the system in eq. (1.2),

$$x(k-2) - 1.928x(k-1) + 0.9418x(k) = 0.006854f(k-1) + 0.006718f(k). \quad (1.4)$$

Equation 1.4 shows that the current value of the dynamic variable depends on past values of the dynamic variable as well as both current and past values of the input function. Figure 1.4 shows the output of the system described in eq. (1.4) for a sampled sinusoidal input. Notice that the discrete response demonstrates the same regular structure as the continuous system shown in Figure 1.3.

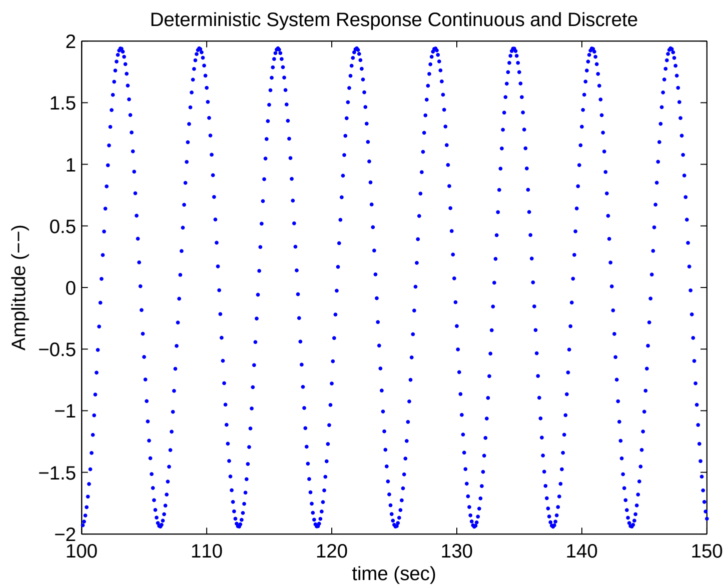

Figure 1.4: Discrete dynamic system response to a sinusoidal input.

At this point please recall the definition of the ARMA process from eq. (1.1). Notice the similarity in form between the ARMA model and eq. (1.4), they are identical except for the driving or forcing function. Consider the impact that the forcing function has on the output. Figure 1.5 shows the response to the same system used to produce the response in Figure 1.4 but with white noise input, Z_t, instead of the sinusoid. Notice here that the output is not completely random (uncorrelated noise)

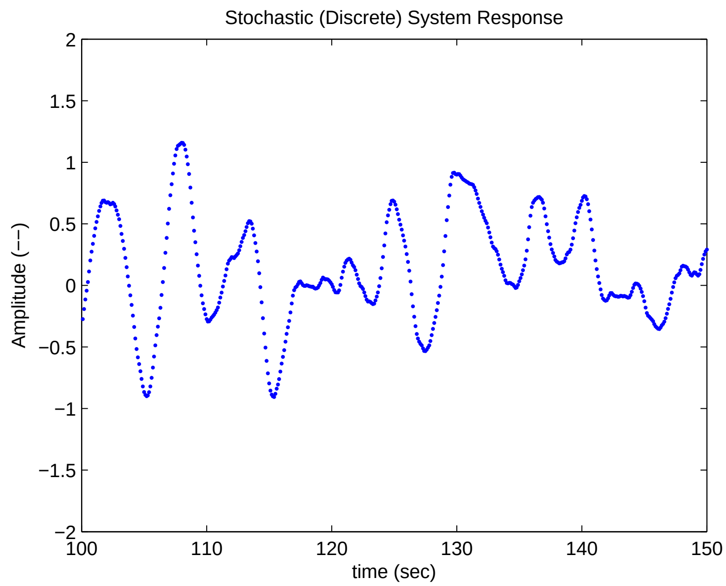

Figure 1.5: Discrete dynamic system response to white noise input.

but it certainly does not have a regular or predictable pattern. The same system produced two outputs with very different character due to changes in the input. This output is a stochastic time series and must be analyzed differently from a function. Accordingly, the diagnostic techniques associated with either type of signal are different.

Figure 1.6 shows how the nature of the system output is dependent on the input. The system definition is neither deterministic nor stochastic. The system is the same. It is the nature of the input that determines the nature of the output. The proper analysis techniques may be applied only after the character of the signal has been determined.

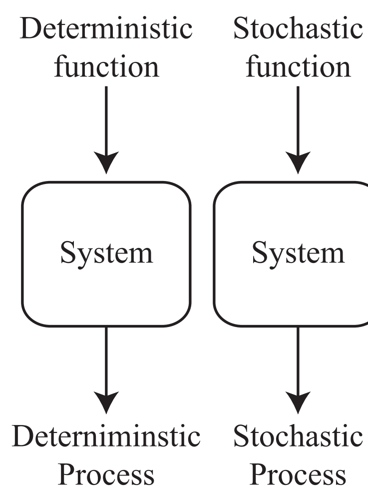

Figure 1.6: The nature of an LTI system response depends on the nature of the input. If the input is deterministic the output is deterministic. If the input is stochastic the output is stochastic.

Chapter 2

Literature Review

The research being presented draws on theoretical and applied areas of both mathematics and engineering. What is presented is meant to put the subsequent chapters in context and provide a rational for the proposed research. The review is not exhaustive but does construct a meaningful outline of what efforts have been made thus far.

The review is organized, more or less, in the order I studied the different areas during the research projects. The initial motivation for the research was an application to the fault detection and diagnosis of commercial scale wind turbines. This led to an investigation of the individual components and then to a more global review of fault detection and diagnosis. Pattern recognition was identified as a means of identifying and diagnosing faults and prompted a review of clustering and classification literature. Time series analysis was then identified as the mathematical framework that would be used in the research and a review of the time series literature was necessary. Finally, the background topics section identifies concepts that do not fit neatly into any previously mentioned area.

2.1 Wind Turbines

The research was initially motivated with an application to wind turbine diagnostics. Wind turbines have become large, expensive devices that operate on a slim financial margin and any degradation of performance can threaten their financial success [4–6].

Much literature has been written about wind turbine fault detection and operation; from books [7–10] to diagnostic review articles [11–13] to journal articles [14–21] as well as in magazine articles [22]. Still more has been written with more of an emphasis on fault detection and diagnosis but with an application toward wind turbines [23–26].

Much of the wind turbine diagnostic research previously mentioned is model based. These diagnostic techniques rely on models of varying specificity in their diagnostic reasoning. A sampling of articles that address modeling and simulation of the turbine and its components include [27–40] and [41–43] address the issue of turbine-grid and farm-grid interaction.

Case studies have been carried out to gain both experience with wind turbine operation as well as to verify the economic feasibility of such instillations. Operation data was made available from the first two years of the Tennessee Valley Authority's (TVA) Buffalo Mountain Wind Power Project [44]. This activity was part of the U.S. Department of Energy/Electric Power Research Institute (EPRI) Utility Wind Turbine Verification Program [45] whose purpose was to gain experience in operating large scale wind turbines and to verify economic feasibility.

In addition to feasibility or siting studies, some researchers have used case studies to evaluate the effects of condition monitoring on life cycle costs [46] and to gain experience with unconventional fault detection techniques [47].

2.2 Electric Machines

The review of electric machines was peripheral to that fault detection and diagnosis of wind turbines but was necessary to understand this important application area. The weakest link in the wind turbine chain is the gearbox followed closely by the generator. The electric generator is an electric machine that has been in existence for a long time and whose behavior has been studies extensively [48–50]. Separate from wind turbines, electric machines are used in many, many applications and there has certainly been a need to develop effective fault detection methodologies for them. Fault detection methodologies utilizing ARMA modeling [51, 52] and statistical time-frequency modeling [53] have also been developed and applied to electric machine diagnosis.

The electric machines on commercial scale wind turbines tend to be induction generators. Various review articles have been written on fault detection for this class of motor [54–56] as well as articles dealing with induction machines used specifically in commercial scale wind turbines [57–62].

2.3 Turbo Machinery

Electric machines fall into a broader class of rotating devices known as turbo machinery. There is significant overlap between diagnosing electric machines and other turbo machinery such as gas turbines [63] and gearboxes. Roemer in [64] presented an overview of selected prognostic technologies with application to engine health management and in [65] presented prognostics and health management software for gas turbine engine bearings. Gearboxes have been the subject of investigation in [66, 67] and in [68] Watson investigated the utility of very high frequency monitoring in the health management of engine gearboxes and generators. Tool wear has also been the subject of diagnostics and statistical modeling [69, 70].

Diagnostic literature will sometimes include research regarding performance predictions such as the paper prepared by Sekhon [71] which presented a comparison of trending strategies for gas turbine performance or remaining useful life (prognostic) research [72].

2.4 Fault Detection and Diagnosis

The area of fault detection and diagnosis is a mature field employing various engineering and mathematical techniques to accomplish its goals (see chapter 1). This diversity stems from the wide range of systems being analyzed. Developing an effective fault detection methodology requires intimate knowledge of the system under investigation and tailoring the technique to fit that application. An example of such an application may be found in [73] where Upadhyaya presents a case study of an application of stochastic modeling of nuclear power plant dynamics.

There are both classic texts [74–76] as well as more modern texts [1, 2, 77–79] outlining some of the more global aspects and application of fault detection. These texts provide an expansive foundation, covering model-based and model-free techniques, feature selection, and discussions about how advancing technology has caused the field to evolve.

One such technique is the Fourier transform, which has proven to be indispensable in many fault detection applications. The transform grants access to the spectral properties to both deterministic and stochastic signals which may, in turn, be used to assess system health. When the observed signal is deterministic the spectrum may be computed and when the observed signal is stochastic the spectrum may be estimated by the periodogram. This transform and its utility have been studied and documented extensively in the literature [80, 81].

As discussed earlier, the nature of a signal (stochastic or deterministic) is important in selecting the proper diagnostic approach. It is possible, however, that a signal exhibits deterministic and stochastic character simultaneously. If the two components of the signal can be identified and separated form one another then they may be analyzed separately. Fisher first developed tests to identify deterministic components of stochastic signals [82].

2.5 Pattern Recognition

A background in pattern recognition [83] is necessary to properly apply the concepts of clustering and discrimination. Pattern Recognition (PR) may be used to aid in making decisions based on observations. This particular aspect of pattern recognition is very much aligned with fault detection and diagnosis. By exposing a PR algorithm to healthy and faulty sample data, the algorithm can assist an operator in analyzing incoming data. Clustering and classification has been used to analyze time series in the past [84–86] and will be extended in this research.

2.6 Time Series Methods

The research focuses principally on the analysis of multivariate time series. Three primary texts in this field are [3] (with a more introductory text found in [87]) along with [88] and [89]. Brockwell and Davis [87] serves as a text book for an introductory time series analysis course, providing basic explanations of stochastic processes and presents various exemplary time series.

Time series analysis is the study of a signals' second order properties. These concepts are closely related to the Fourier transform, by way of the spectrum and its estimator, the periodogram. Tukey [90] provides as lengthy but insightful discussion on the connection between analysis of variance and spectrum analysis which helps to make interconnections clear.

These quantities are theoretical and, for observed series, must be estimated. Estimating the mean, auto- and cross-covariance poses no problem. Unfortunately, the intuitive spectral estimator turns out to be inconsistent. In [91], Welch developed a technique that mitigates the stochastic nature and allows one to better estimates the spectra. Caiado in [92] developed techniques for comparing time series of different length and in [93] investigated metrics for time series classification. These metrics were similar to the ones investigated as part of this research expect that they were for the univariate case. The proposed research is applied to multivariate stochastic processes.

2.7 Multivariate Analysis

The literature is well developed in terms of testing whether of the means of two multivariate populations are the same. These techniques were developed in the 1930s [94] and '40s and found an application in multivariate quality control [95, 96], model fitting [97], discrimination [98, 99] and clustering of multivariate time series [100]. The proposed research seeks to further the body of literature concerning multivariate statistical analysis based on a signals second order properties. This type of analysis would strengthen the ability to compare multivariate stochastic signals based on observed realizations.

The research conducted by Bassily et al. in [101] and [102] concerned discrimination of multivariate time series. This research offered the opportunity to be furthered by subsequent research. The proposed research addresses some of the open issues associated with that research.

Principal Component Analysis (PCA) [103] was investigated as a possible way to characterize multivariate series. In the literature, PCA has been used for fault detection [104] as well as multivariate quality control [105] and for more general signal and model analysis [106]. Traditional PCA, however, tries to identify the static relationships among variables by answering the question, "Is there any correlation among samples collected *at the same time*?" This approach is marred when there is correlation amongst samples as its development assumes independent or at least uncorrelated observations.

A variation of PCA investigated during the course of this research is dynamic principal component analysis or dynamic PCA. The technique was pioneered by Ku in [107] and is aimed at accounting for the time history dependence of stochastic signals (characteristic of a time series) using PCA.

Dynamic PCA starts with the Hankel matrix (a quantity familiar to system identification literature [108]) and applies a type of principal component analysis to estimate ARMA model parameters. This technique has two primary drawbacks. First, the order of the model is determined by the number of lags included in the augmentation. One must know either the appropriate order of the model or perform the analysis for numerous lags and determine what number of lags works best. Also, the technique does not work very well when the system input is pure white noise. The technique works much better when the input is some type of colored noise.

2.8 Background Topics

The field of fault detection and diagnosis is broad and requires a healthy background in various areas. Gut and Ross provide a necessary foundation in probability in [109, 110] while Ross offers an excellent introduction to stochastic processes in [111]. Anderson provides an approachable text in multivariate statistics [112] while Golub in [113] provides a reference for linear algebra and matrix computations. The extremely influential Claude E. Shannon wrote [114] and pioneered work in the area of information theory. This work is closely related the concepts of information sufficiency [115], discriminations criterion [116] and quantification of entropy [117].

Two other useful references are the The Handbook of Data Mining [118] and the NIST/SEMATECH e-Handbook of Statistical Methods [119]. The NIST manual is particularly helpful as it includes step-by-step procedures for performing hundreds of statistical and process analyses.

Chapter 3

A Comparison of Multivariate Signal Discrimination Measures

The research problem studied was the comparison of several techniques to discriminate two multivariate stationary signals. The compared methods include Gaussian likelihood ratio variance/covariance matrix tests – perhaps best viewed as principal component analyses (PCA) without dimension reduction aspects – and spectral-based tests gauging equality of the autocovariance function (over all lags) of the two signals. We show how one can make inappropriate conclusions with PCA tests, even when dimension augmentation techniques are used to incorporate non-zero lag autocovariances into the analysis. The various discrimination methods are first discussed. A simulation study is then presented that illuminates the various properties of the methods. An analysis of experimentally collected gearbox data is also presented.

3.1 Introduction

Given two d-dimensional series $\{\mathbf{X}_t\}$ and $\{\mathbf{Y}_t\}$ that are preprocessed to a zero-mean stationary setting, this paper considers how to assess whether (or not) the two signals have the same time series dynamics. This is useful in discrimination and classification pursuits. For example, if a test signal $\{\mathbf{Y}_t\}$ is deemed to have different dynamics than a reference signal $\{\mathbf{X}_t\}$ that is known to be "healthy", the test signal could be deemed unhealthy. Signal discrimination problems are fundamental (see [114, 115]) and are well-developed when discriminating series via means or first moments; here, Hotelling T^2 or Q statistics are frequently relied upon as in [95] and [94]. In 1986, [98] considered discrimination of two univariate constant-mean series based on their sample autocovariances. Speech signals, for example, are typically of constant mean, regardless of what words are being spoken. Here, word-to-word changes are best identified through autocovariances shifts and monitoring of the mean is insufficient to identify dynamic changes. [100] seeks to discriminate an earthquake from a covert underground nuclear test; again, the crux lies with constant-mean data.

The classical way of discriminating $\{\mathbf{X}_t\}$ and $\{\mathbf{Y}_t\}$ through second order characteristics is via a Gaussian likelihood ratio. Such a test compares the sample variance matrix of the two series. Elaborating, conclusions are based on how different the two sample variance matrices

$$N^{-1}\sum_{t=1}^{N}\mathbf{X}_t\mathbf{X}_t', \qquad N^{-1}\sum_{t=1}^{N}\mathbf{Y}_t\mathbf{Y}_t'$$

are from each other. Section 3.2 shows how to do this. Here, N is the sample length of the two series, which are assumed equal for convenience. When the dimension d is large, this comparison is typically made after a dimension reduction transformation, usually some type of principal component analysis (PCA), is done. Without dimension reduction aspects, covariance comparisons are not truly PCA techniques; however, they share the commonality in that conclusions are made only from sample variances.

Basing signal equality conclusions exclusively on sample variances can produce erroneous conclusions when the two series are not multivariate white noise. A more comprehensive test would compare the sample autocovariances

$$\hat{\mathbf{\Gamma}}_{\mathbf{X}}(h) = N^{-1} \sum_{t=1}^{N-h} \mathbf{X}_{t+h}\mathbf{X}_t'$$

and

$$\hat{\mathbf{\Gamma}}_{\mathbf{Y}}(h) = N^{-1} \sum_{t=1}^{N-h} \mathbf{Y}_{t+h}\mathbf{Y}_t'$$

over all suitable lags $h \geq 0$. Such tests for multivariate series were discussed in [100], [102], [99], and the references within.

PCA methods have been extended to handle cases where correlation at non-zero series lags is present. This is typically done through a dimension augmentation scheme. For example, if $\mathbf{\Gamma}_X(1)$ and/or $\mathbf{\Gamma}_Y(1)$ are believed to be non-zero, one could compare the sample covariance matrices of the $2d$-dimensional vectors $\{\mathbf{X}_t^*\}$ and $\{\mathbf{Y}_t^*\}$, where $\mathbf{X}_t^* = (X_{2t-1,1}, \ldots, X_{2t-1,d}, X_{2t,1}, \ldots, X_{2t,d})'$ and $\mathbf{Y}_t^* = (Y_{2t-1,1}, \ldots, Y_{2t-1,d}, Y_{2t,1}, \ldots, Y_{2t,d})'$. If the sample variance of $\{\mathbf{X}_t^*\}$ and $\{\mathbf{Y}_t^*\}$ agree, then one concludes that $\mathbf{\Gamma}_{\mathbf{X}}(0) = \mathbf{\Gamma}_{\mathbf{Y}}(0)$ and $\mathbf{\Gamma}_{\mathbf{X}}(1) = \mathbf{\Gamma}_{\mathbf{Y}}(1)$. Higher order comparisons are constructed via analogous reasoning. Of course, such dimension augmentation tactics shorten the observed series length; also, there is no clear maximum lag to augment by when autocovariances at all lags are non-zero, the typical case in practice.

Bassily [102] and Lund [99] attack the problem with different techniques.

Specifically, two multivariate covariance functions are equal if and only if their spectral densities are equal at all frequencies (the spectrum is assumed to have no point masses). From this, signal equality tests that compare the periodograms of both series were devised (Section 3.2 elaborates). This paper revisits these methods and shows how one can fool variance-based tests for signal equality, even when the dimension is augmented to account for non-zero autocovariances at higher lags. The pros and cons of the various methods are demonstrated by simulating multivariate stationary signals with various properties and then applying the tests. An application to a series of gearbox vibrations is included.

The rest of this paper proceeds as follows. Section 3.2 presents the signal processing background needed for the methods. Section 3.3 then shows how the techniques compare on simulated series with various autocovariance properties. Section 3.4 analyzes several gearbox vibration data series. Section 3.5 reviews the content of the paper and makes some closing remarks.

3.2 Background

We work with two zero-mean d-dimensional covariance stationary signals $\{\mathbf{X}_t\}$ and $\{\mathbf{Y}_t\}$ observed at times $t = 1, \ldots, N$. The covariance matrices at lag $h \geq 0$ are

$$\mathbf{\Gamma}_{\mathbf{X}}(h) = E[\mathbf{X}_{t+h}\mathbf{X}_t'], \quad \mathbf{\Gamma}_{\mathbf{Y}}(h) = E[\mathbf{Y}_{t+h}\mathbf{Y}_t'].$$

3.2.1 Testing Equality of Variances

The classical test for signal equality of zero-mean stationary series merely compares the sample variance matrices of the two observed series. The null hypothesis is that $\mathbf{\Gamma}_{\mathbf{X}}(0) = \mathbf{\Gamma}_{\mathbf{Y}}(0)$. A Gaussian likelihood ratio statistic for testing this hypothesis is

$$\lambda = \left(2^d \frac{\det \left[\hat{\mathbf{\Gamma}}_{\mathbf{X}}(0) \hat{\mathbf{\Gamma}}_{\mathbf{Y}}(0) \right]^{1/2}}{\det \left[\hat{\mathbf{\Gamma}}_{\mathbf{X}}(0) + \hat{\mathbf{\Gamma}}_{\mathbf{Y}}(0) \right]} \right)^N , \tag{3.1}$$

where det indicates matrix determinant. This statistic is derived in [112], pg. 404. Values of λ are in $[0, 1]$ and the null hypothesis is rejected when λ is too small to be explained by random chance. Authors have used this test when the series are non-Gaussian white noise without drastic performance degradations. Here, the usual central limit caveat applies: the test works well for large N provided marginal distributions of the series are not heavy-tailed. Applying the test when the data are autocorrelated (i.e, not white noise) is more problematic. This aspect will be demonstrated in the next Section.

In great generality, $-2\ln(\lambda)$ has an asymptotic (as $N \to \infty$) chi-squared distribution (see [120], [121]). The degrees of freedom is equal to the number of parameters that are saved when the two signals have the same covariance matrix. Since covariance matrices of a d-dimensional signal are $d \times d$ symmetric matrices, $d(d+1)/2$ free parameters are saved; that is, $d(d+1)/2$ is the appropriate degrees of freedom. Phrased another way, λ asymptotically behaves as $e^{-L/2}$, where L is a chi-squared random variate with $d(d+1)/2$ degrees of freedom. From this, it follows that λ has the asymptotic density

$$f_\lambda(x) = \frac{[-\ln(x)]^{d(d+1)/4-1}}{\Gamma\left(\frac{d(d+1)}{4}\right)}, \quad 0 \le x \le 1.$$

Here, $\Gamma(\alpha)$ represents the usual Gamma function at argument $\alpha > 0$ (the use of Γ as both a covariance and a function should cause no confusion). This density can be used to extract percentiles; however, exact formulas cannot be given since the

Table 3.1: Ninety-fifth percentiles for λ

Dimension	Ninety-Fifth Percentile
1	0.1478
2	0.02025
3	0.001839
4	1.035e-4
5	3.580e-6

antiderivative of $\ln(x)^\beta$ for $\beta > 0$ has no explicit formula. Table 3.1 lists how small λ must be to warrant rejection of equal variances with 95% statistical confidence for several values of d. A plot of the asymptotic density of λ for $d = 2$ is shown in Figure 3.1.

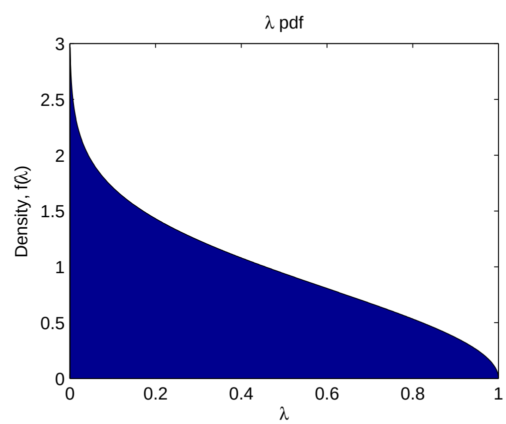

Figure 3.1: Probability density of λ when $d = 2$

3.2.2 Testing Equality of the Autocovariance Functions

A spectral approach to testing equality of multivariate autocovariance functions was developed in [102]. Since $\mathbf{\Gamma_X}(h) = \mathbf{\Gamma_Y}(h)$ for all lags $h \geq 0$ if and only if $\mathbf{f_X}(\omega) = \mathbf{f_Y}(\omega)$ for all frequencies $\omega \in [0, 2\pi)$ (with respect to the Lebesgue measure), where

$$\mathbf{f_X}(\omega) = \frac{1}{2\pi} \sum_{h=-\infty}^{\infty} \mathbf{\Gamma_X}(h) e^{-i\omega h}$$

and

$$\mathbf{f_Y}(\omega) = \frac{1}{2\pi} \sum_{h=-\infty}^{\infty} \mathbf{\Gamma_Y}(h) e^{-i\omega h}$$

are the theoretical spectral densities of $\{\mathbf{X}_t\}$ and $\{\mathbf{Y}_t\}$ at frequency ω, respectively.

Bassily [102] estimates the spectral densities of the two series and statistically compares their ratios. Specifically, the discrete Fourier transforms (DFTs) of the series are first computed via

$$\mathbf{J_X}(\omega_j) = N^{-1/2} \sum_{t=1}^{N} \mathbf{X}_t e^{-it\omega_j}$$

and

$$\mathbf{J_Y}(\omega_j) = N^{-1/2} \sum_{t=1}^{N} \mathbf{Y}_t e^{-it\omega_j}$$

at all Fourier frequencies $\omega_j = 2\pi j/N$, for $j = 0, \ldots, N-1$ (see [80] and [81] for Fourier transform basics). The raw (unsmoothed) spectral densities are estimated via

$$\hat{\mathbf{f}}_\mathbf{X}(\omega_j) = \frac{\mathbf{J_X}(\omega_j)\mathbf{J}^*_\mathbf{X}(\omega_j)}{2\pi}, \qquad \hat{\mathbf{f}}_\mathbf{Y}(\omega_j) = \frac{\mathbf{J_Y}(\omega_j)\mathbf{J}^*_\mathbf{Y}(\omega_j)}{2\pi}.$$

Here, the asterisk denotes complex conjugation. The raw spectral estimates are then smoothed in a uniform manner over $2M+1$ Fourier frequencies closest to the Fourier frequency being considered:

$$\hat{\mathbf{f}}^s_{\mathbf{X}}(\omega_j) = \frac{\sum_{k=-M}^{M} \hat{\mathbf{f}}_{\mathbf{X}}(\omega_{j+k})}{2M+1}, \qquad \hat{\mathbf{f}}^s_{\mathbf{Y}}(\omega_j) = \frac{\sum_{k=-M}^{M} \hat{\mathbf{f}}_{\mathbf{Y}}(\omega_{j+k})}{2M+1}.$$

Here, M is a positive integer, representing a smoothing bandwidth, that satisfies $2M+1 \geq d$ (this is needed for technical reasons rooted in the finiteness of variances). The choice of M does not usually influence practical conclusions about signal equality. In smoothing the raw spectral estimates, frequencies outside of $[0, 2\pi)$ are rounded modulo 2π to mimic the periodic nature of the DFT; for example, $\hat{\mathbf{f}}_{\mathbf{X}}(\omega_{j+N}) = \hat{\mathbf{f}}_{\mathbf{X}}(\omega_j)$.

Bassily [102] bases signal equality conclusions on the statistic

$$\bar{\Delta} = \frac{1}{\frac{N}{2}-1} \sum_{j=1}^{\frac{N}{2}-1} |\Delta(\omega_j)|. \tag{3.2}$$

Here, the $\Delta(\lambda_j)$'s are the log determinant of the ratios of the smoothed spectral density estimates:

$$\Delta(\omega_j) = \log\left(\det\left(\hat{\mathbf{f}}^s_{\mathbf{X}}(\omega_j)\right)\right) - \log\left(\det\left(\hat{\mathbf{f}}^s_{\mathbf{Y}}(\omega_j)\right)\right). \tag{3.3}$$

Under the null hypothesis of equal autocovariance functions, $\Delta(\omega_j)$ should be statistically close to zero for every non-zero Fourier frequency ω_j. Bassily [102] shows that $\Delta(\omega_j)$ has an asymptotic distribution that does not depend on j for $j = 1, 2, \ldots, N/2-1$ or the common spectral density of $\{\mathbf{X}_t\}$ and $\{\mathbf{Y}_t\}$. From this structure, a test for signal equality based on $\overline{\Delta}$ is easily constructed based on the central limit theorem (the $\Delta(\omega_j)$'s for varying j are approximately independent). Such a test rejects equality of autocovariance functions when

$$\bar{\Delta} > \mu_M + z_\alpha \frac{\sigma_M}{\sqrt{\frac{N}{2}-1}}. \tag{3.4}$$

Here, z_α denotes a quantile that cuts off an upper tail area of α in the standard normal distribution ($z_\alpha = 1.645$ when $\alpha = 0.05$) and μ_M and σ_M are the theoretical mean and variance of $|\Delta(\omega_j)|$. Note that this is a one sided test.

The constants μ_M and σ_M^2 depend on both M and d and are difficult to derive. [99] derives explicit expressions when $d = 1$, but the computations for the multidimensional case are intense. However, simulations with Gaussian white noise readily provide good estimates of them. These estimates are given in tables in [102].

The detection power of the $\bar{\Delta}$ statistic can be increased if the signals are known to be band-limited. Specifically, if the spectrums of $\{\mathbf{X}_t\}$ and $\{\mathbf{Y}_t\}$ are known to be limited to the interval $[\omega_L, \omega_U]$, then eq. (3.2) is modified to

$$\bar{\Delta} = C^{-1} \sum_{j:\omega_j \in [\omega_L, \omega_U]} |\Delta(\omega_j)|,$$

where C is the number of distinct Fourier frequencies in the interval $[\omega_L, \omega_U]$. The rejection region is the same as in eq. (3.4), except that $N/2-1$ is replaced by C. One should take C large enough to induce asymptotic normality of averages (a typical rule of thumb takes $C \geq 30$). Detection power increases because many frequencies where no differences occur are excluded in the analysis, accentuating the importance of differences in the considered frequency increments.

3.3 Method Comparison

This section studies the properties of the λ and $\bar{\Delta}$ statistics through specifically designed simulations to illustrate various points. In all cases, the issues are apparent in dimension $d = 2$ and at 95% statistical confidence. The smoothing parameter $M = 5$ and series length $N = 1024$ are also common to all cases. In all cases, one hundred thousand simulations were conducted.

First, the λ and $\bar{\Delta}$ statistics were computed for each simulated realization of $\{\mathbf{X}_t\}$ and $\{\mathbf{Y}_t\}$, each realization containing zero-mean Gaussian white noise. In this case, the covariance matrix of $\{\mathbf{X}_t\}$ and $\{\mathbf{Y}_t\}$ was taken as the two-dimensional identity matrix. Hence, this case, which we refer to as Case I, is a scenario where the two signals have the same dynamics. Table 3.2 shows empirically aggregated proportions of runs where the λ and $\bar{\Delta}$ reject the null hypothesis of signal equality at level 5%. As both proportions are close to 5%, both methods have worked well in this case.

Our second case is one where $\{\mathbf{X}_t\}$ and $\{\mathbf{Y}_t\}$ do not have the same variance (lag-zero covariance matrix). Here, $\{\mathbf{X}_t\}$ and $\{\mathbf{Y}_t\}$ are zero-mean Gaussian white noise with the covariance matrices

$$\mathbf{\Gamma}_{\mathbf{X}}(0) = \begin{bmatrix} 1.0 & 0.0 \\ 0.0 & 1.0 \end{bmatrix}, \qquad \mathbf{\Gamma}_{\mathbf{Y}}(0) = \begin{bmatrix} 1.1 & 0.1 \\ 0.1 & 1.0 \end{bmatrix},$$

respectively. Table 3.2 displays the proportions of times the λ and $\bar{\Delta}$ tests reject signal equality at confidence 95%. In this case, the likelihood ratio statistic λ has worked best, drastically so, as seen by its larger empirical rejection proportion. This is not unexpected: while both methods should ideally reject signal equality, the two signals differ only in their variances; covariances at all higher lags are zero. While the λ statistic focuses solely on variance differences, the $\bar{\Delta}$ statistics must consider all covariance lags. This essentially degrades the detection power of the $\bar{\Delta}$ test in this case.

Case III considers a situation where $\{\mathbf{X}_t\}$ and $\{\mathbf{Y}_t\}$ have the same variances, but where there is non-zero autocorrelation at non-zero lags; that is, the series under consideration are not multivariate white noise. We do this by examining solutions to the vector autoregressive moving-average (VARMA) model of autoregressive order 2 and moving-average order 1. Specifically, both $\{\mathbf{X}_t\}$ and $\{\mathbf{Y}_t\}$ obey the VARMA difference equation

$$\mathbf{X}_t = \mathbf{\Phi}_1\mathbf{X}_{t-1} + \mathbf{\Phi}_2\mathbf{X}_{t-2} + \mathbf{Z}_t + \mathbf{\Theta}_1\mathbf{Z}_{t-1}.$$

Here, the autoregressive matrix coefficients were chosen as

$$\mathbf{\Phi}_1 = \begin{bmatrix} 0.40 & 0.05 \\ 0.05 & 0.30 \end{bmatrix}, \quad \mathbf{\Phi}_2 = \begin{bmatrix} -0.48 & 0.10 \\ 0.10 & -0.06 \end{bmatrix},$$

and the moving-average coefficient matrix was selected as

$$\mathbf{\Theta}_1 = \begin{bmatrix} 0.30 & 0.10 \\ 0.10 & 0.50 \end{bmatrix}.$$

Also, $\{\mathbf{Z}_t\}$ is chosen as white noise with an identity covariance matrix. The Case III performance characteristics reverse from Case II with the λ statistic erroneously rejecting signal equality about 19% of the time. Most statisticians view this false alarm rate as unacceptable in a 95% test. The $\bar{\Delta}$ statistic, however, rejects signal equality at approximately the intended 5% rate.

Case IV represents an exacerbated version of Case III. Here, the two series are taken as vector autoregressions of order one. Specifically, both series follow the VAR(1) dynamics

$$\mathbf{X}_t = \mathbf{\Phi}\mathbf{X}_{t-1} + \mathbf{Z}_t,$$

where $\{\mathbf{Z}_t\}$ is taken as Gaussian white noise with the identity covariance matrix and

Table 3.2: Method detection and false alarm probabilities

	λ	$\bar{\Delta}$
Case I	5.14%	5.39%
Case II	57.69%	7.05%
Case III	19.04%	5.54%
Case IV	73.19%	5.62%
Case V	8.12%	100.00%
Case VI	100.00%	100.00%
Case VII	7.68%	15.61%

$$\mathbf{\Phi} = \begin{bmatrix} 0.90 & 0.10 \\ -0.10 & 0.90 \end{bmatrix}.$$

The dynamics of this model lie near the boundary of the multivariate causality region of a VAR(1) model, as is seen by the near unit diagonal entries in $\mathbf{\Phi}$. In this case, the λ statistic erroneously rejects signal equality at a whopping 73% rate. The false alarm (Type I error) of the $\bar{\Delta}$ test is also getting a bit larger than the specified 5%, but not drastically so. Taken together, the last two cases show that likelihood ratio tests to detect variance changes perform suboptimally unless the signals are known to be white noise. At this point, one can also question the detection power of the $\bar{\Delta}$ statistic as it performed poorly in the one case where the signals were truly different (Case II). The next three cases will perhaps remedy this concern.

Case V moves to a situation designed to fool the λ statistic. Specifically, our $\{\mathbf{X}_t\}$ is taken as the first-order moving-average satisfying

$$\mathbf{X}_t = \mathbf{Z}_t + \mathbf{\Theta}\mathbf{Z}_{t-1}.$$

and $\{\mathbf{Y}_t\}$ is taken as white noise

$$\mathbf{Y}_t = \mathbf{\Xi}_t.$$

The caveat here is that we select the parameters $\mathbf{\Theta}$, $\text{Var}(\mathbf{Z}_t) = \mathbf{\Sigma_Z}$, and $\text{Var}(\mathbf{\Xi}_t) = \mathbf{\Sigma_\Xi}$ so that $\mathbf{\Gamma_X}(0) = \mathbf{\Gamma_Y}(0)$. To do this, we take

$$\mathbf{\Theta} = \begin{bmatrix} 0.70 & 0.30 \\ 0.30 & 0.50 \end{bmatrix}, \quad \mathbf{\Sigma_Z} = \begin{bmatrix} 1.00 & 0.00 \\ 0.00 & 1.00 \end{bmatrix},$$

and

$$\mathbf{\Sigma_\Xi} = \begin{bmatrix} 1.58 & 0.36 \\ 0.36 & 1.34 \end{bmatrix},$$

In this case, the two series have different dynamics, but have the same lag-zero variance matrix. The empirical probabilities in Table 3.2 reflect this property: the λ statistic opts for equivalent signal dynamics only slightly more than the 5% nominal false alarm rate; however, the $\bar{\Delta}$ statistic makes the correct conclusion of signal inequality in all of the one hundred thousand runs.

Summarizing to this point, the λ test degrades under correlation but is more powerful at detecting variance changes when only variance changes are truly present.

One can reduce equality of autocovariance problems to variance comparisons through dimension augmentation techniques. For example, suppose that the signal's autocovariances are only non-zero at lags $0, 1, \ldots, \kappa$ and set

$$\mathbf{X}_n^* = (\mathbf{X}'_{(n-1)(\kappa+1)+1}, \ldots, \mathbf{X}'_{n(\kappa+1)})'.$$

Then $\{\mathbf{X}_t\}$ and $\{\mathbf{Y}_t\}$ have the same autocovariances at lags $h = 0, \ldots, \kappa$ if $\{\mathbf{X}_n^*\}$ and $\{\mathbf{Y}_n^*\}$ have equal variances. For example, in Case V, $\mathbf{X}_t^* = (X_{2t-1,1}, X_{2t-1,2}, X_{2t,1}, X_{2t,2})'$

and $\mathbf{Y}_t^* = (Y_{2t-1,1}, Y_{2t-1,2}, Y_{2t,1}, Y_{2t,2})'$. Of course, such tactics may not represent an efficient way of proceeding when κ is large as series sample sizes are reduced.

Case VI shows empirical probabilities of signal equality rejection when 4-dimensional vectors are made to analyze the signals generated in Case IV. We will not rerun the $\bar{\Delta}$ analyses, preferring to emphasize that the $\bar{\Delta}$ method naturally handles autocorrelation and that there is no need to do any sort of dimension augmentation. The rejection probability of the λ statistic in Case V increases to 100% when the dimension is augmented to four dimensions. Since moving averages are completely characterized by their lag-zero and lag-one autocovariances, dimension augmentation works very well here.

Selection of the dimension to augment by is problematic. If one selects the augmentation dimension too small, higher order covariances will not be considered (which is suboptimal if these autocovariances are non-zero). On the other hand, if the selected dimension is too large, then the sample size becomes significantly smaller and discrimination power is lost.

Our last case is intended to show that there are no easy ways of selecting augmentation dimensions. We do this by constructing two series where the signals have different dynamics, but where both the lag-zero and lag-one autocovariances agree. That is, we want $\{\mathbf{X}_t\}$ and $\{\mathbf{Y}_t\}$ to have different dynamics, but $\mathbf{\Gamma}_{\mathbf{X}}(0) = \mathbf{\Gamma}_{\mathbf{Y}}(0)$ and $\mathbf{\Gamma}_{\mathbf{X}}(1) = \mathbf{\Gamma}_{\mathbf{Y}}(1)$. Case VII shows signal equality rejection probabilities in such a case. This was done by mixing two univariate signals with equal lag-zero and lag-one autocovariances. Specifically, suppose that $\{X_{t,1}^*\}$, and $\{X_{t,2}^*\}$, the components of $\{\mathbf{X}_t^*\}$, both follow the same AR(1) dynamics

$$X_{t,1}^* = \phi X_{t-1,1}^* + Z_{t,1}, \qquad X_{t,2}^* = \phi X_{t-1,2}^* + Z_{t,2},$$

where $\{Z_{t,1}\}$ and $\{Z_{t,2}\}$ are independent zero-mean unit-variance Gaussian white noise series. Hence, the two components of $\{\mathbf{X}_t^*\}$ are independent AR(1) series having the same univariate covariances at all lags. Now suppose that both components of $\{\mathbf{Y}_t^*\}$ obey the same MA(1) dynamics:

$$Y_{t,1}^* = \eta_{t,1} + \theta\eta_{t-1,1}, \qquad Y_{t,2}^* = \eta_{t,2} + \theta\eta_{t-1,2},$$

where $\{\eta_{t,1}\}$ and $\{\eta_{t,2}\}$ are independent zero-mean variance σ_η^2 Gaussian white noise series. A simple computation shows that $\{X_{t,1}^*\}$ and $\{Y_{t,1}^*\}$ have the same lag-zero and lag-one autocovariances when

$$\phi = \frac{\theta}{1+\theta^2}, \qquad \sigma_\eta^2 = \frac{1+\theta^2}{1+\theta^2+\theta^4}.$$

To mix the two components (so that $\{X_{t,1}\}$ and $\{X_{t,2}\}$ are not independent), set

$$\mathbf{X}_t = \mathbf{L}\begin{pmatrix} X_{t,1}^* \\ X_{t,2}^* \end{pmatrix}, \qquad \mathbf{Y}_t = \mathbf{L}\begin{pmatrix} Y_{t,1}^* \\ Y_{t,2}^* \end{pmatrix},$$

where

$$\mathbf{L} = \begin{bmatrix} 1/2 & 1/3 \\ -1/3 & 1/2 \end{bmatrix}.$$

Then $\{\mathbf{X}_t\}$ and $\{\mathbf{Y}_t\}$ have different signal dynamics, yet, by construction, $\mathbf{\Gamma}_{\mathbf{X}}(0) = \mathbf{\Gamma}_{\mathbf{Y}}(0)$ and $\mathbf{\Gamma}_{\mathbf{X}}(1) = \mathbf{\Gamma}_{\mathbf{Y}}(1)$.

The Case VII probabilites use $\phi = 1/4$. The values $\theta = 2-\sqrt{3}$, and $\sigma_\eta^2 = 0.9952$ were then chosen to satisfy the above constraints. The Table 3.2 rejection proportions show that while the $\bar{\Delta}$ statistic does not detect signal inequality well, the λ statistic is almost completely fooled. Because of this, we do not consider comparing signals whose autocovariances match to a higher number of lags as the pattern is clear: the λ statistic will have more difficulty correctly discriminating such signals.

Overall, the $\bar{\Delta}$ tests seems to perform well without the need for dimension augmentation. Performance of the classical λ test can degrade should autocorrelations in the series be present (i.e., this test performs well for white noise discrimination only). We suggest that the $\bar{\Delta}$ statistic be considered should conclusions on signal equality have importance.

3.4 Gearbox Analysis

To demonstrate discrimination capabilities on actual data, the λ and $\bar{\Delta}$ statistics will be computed for three experimentally collected gearbox vibration signals of dimension $d = 2$. Our goal here lies with fault diagnosis. In fault diagnosis schemes, a known healthy signal is compared to a test signal, which may be healthy or unhealthy. An unhealthy signal is indicative of faults. Such an approach has been used to diagnose faults in wind turbine gearboxes (see [12, 26]), gas turbines (see [63, 71]), electric motors (see [52, 54, 55, 66]), and general rotating components as in [25]. See [1, 2, 51, 69, 74–76, 79, 92, 93] for other fault detection research.

The vibration data used here comes from The Prognostics and Health Management Society (PHM Society) as part of their *2009 PHM Challenge Competition Data Set*. Similar data sets are found at NASA's Prognostics Center of Excellence's prognostic-data-repository (`http://ti.arc.nasa.gov/tech/dash/pcoe/prognostic-data-repository/`). The data were collected from a generic, three-axis gearbox with accelerometers mounted on the input side and output side (see Figure 3.2). The input pinion had 32 teeth, the input-side idler gear 96 teeth, the output-side idler gear 48 teeth, and the output gear 80 teeth, resulting in the 5 to 1 reduction ratio.

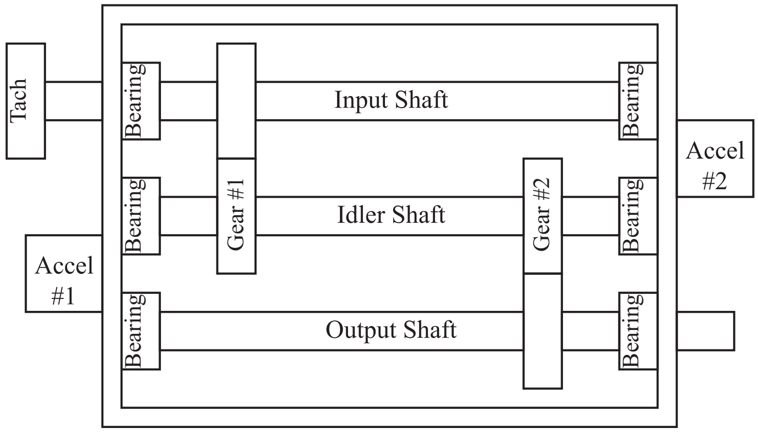

Figure 3.2: System diagram of the generic industrial gearbox used in the 2009 PHM Society competition showing the location of the accelerometers and the physical relation of the components.

The vibration data set, as a whole, contains over 560 two dimensional series. These series correspond to gearbox runs at 30, 35, 40, 45, and 50 Hz under high and low loadings, all repeated twice. This frequency sequence was run again for numerous fault cases, including chipped teeth, broken teeth, eccentric gears, bent shafts, imbalanced shafts, and inner and outer bearing defects. This battery of tests was repeated for helical and spur gears. The series were collected at 66.6kHz and are of length $N = 266000$.

Our investigation focuses on three series. Series A and B were collected from the gearbox when no faults were present (healthy data). Series C was collected after various faults were introduced (faulty data). The faults present in series C include an eccentric gear, a gear with a broken tooth, and a bearing with a fouled ball.

Figure 3.3 plot segments of the component series. Notice that the data appear to have a constant mean (roughly) and were sampled at a very high frequency. In fact, the entire data length corresponds to only 3.99 seconds of runtime. In truth, non-stationarity is likely present in these series. Plausibly, there are many deterministic sinusoids embedded in the series, a prominent one residing at 30Hz. We will combat local variance change aspects by making sliding subsegments of length 1024.

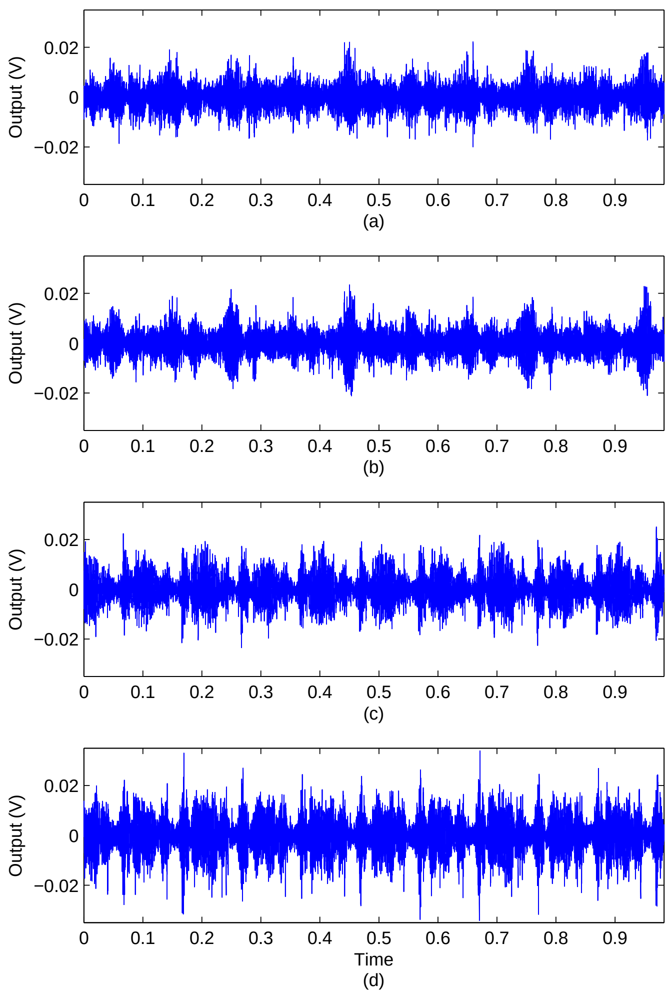

Figure 3.3: Sample of data to be analyzed. (a) Gear 1, Component 1. (b) Gear 1, Component 2. (c) Gear 2, Component 1. (d) Gear 2, Component 2.

Smoothed periodograms of the components of the healthy series A and faulty series C are plotted in Figure 3.4. The smoothing uniformly weights eleven adjacent periodogram ordinates. The periodograms of the healthy and faulty data appear pretty similar. Observe that all significant spectral content is located below 12,000Hz and that the more significant spectral content is found below 1,000Hz. This is to be expected. The input shaft for the data being analyzed is rotating at 30Hz and with a gear reduction ratio of 5:1, the output shaft will be rotating at 6Hz. The spectral contributions from the rotating shafts and gears as well as the tooth interactions are expected to be at lower frequencies, particularly below 1,000Hz. Because of this, we band-limit all $\bar{\Delta}$ statistics to $[0, 1,000]$Hz. That is not to say the higher frequencies are totally negligible. A broken tooth, for example, creates a short-duration disturbance once-per-gear revolution. This once per cycle, short-duration disturbance may be similar to a impulse-train type disturbance and may affect the system accordingly. Impact Technologies identified such behavior and exploited it in their ImpactEnergy™ detection algorithms [64, 65, 68, 72, 77, 78].

To compare signals, each series will be segmented into non-overlapping segments of length 1024, resulting in roughly 250 subsegments. Each subsegment will be compared to the corresponding subsegment in the other series and referred to as a trial. Each trial calculates a λ and a band-limited $\bar{\Delta}$. Once all 250 comparisons are made, the percentage of trials that exceed the 95^{th} percentile for each corresponding statistic will be reported.

Table 3.3 summarizes the outcomes. For the case where the comparison is between two like signals, the λ statistic declares them different 96.4% of the time (all conclusions are made at level 5%) while the $\bar{\Delta}$ statistic declares them different only 50.0% of the time. In truth, there are likely some subtle differences between the two healthy case runs. However, as there is significant non-zero correlation at many lags in this data, one believes the $\bar{\Delta}$ results to be more realistic.

When comparing signal A to signal C, the λ statistic declares them different

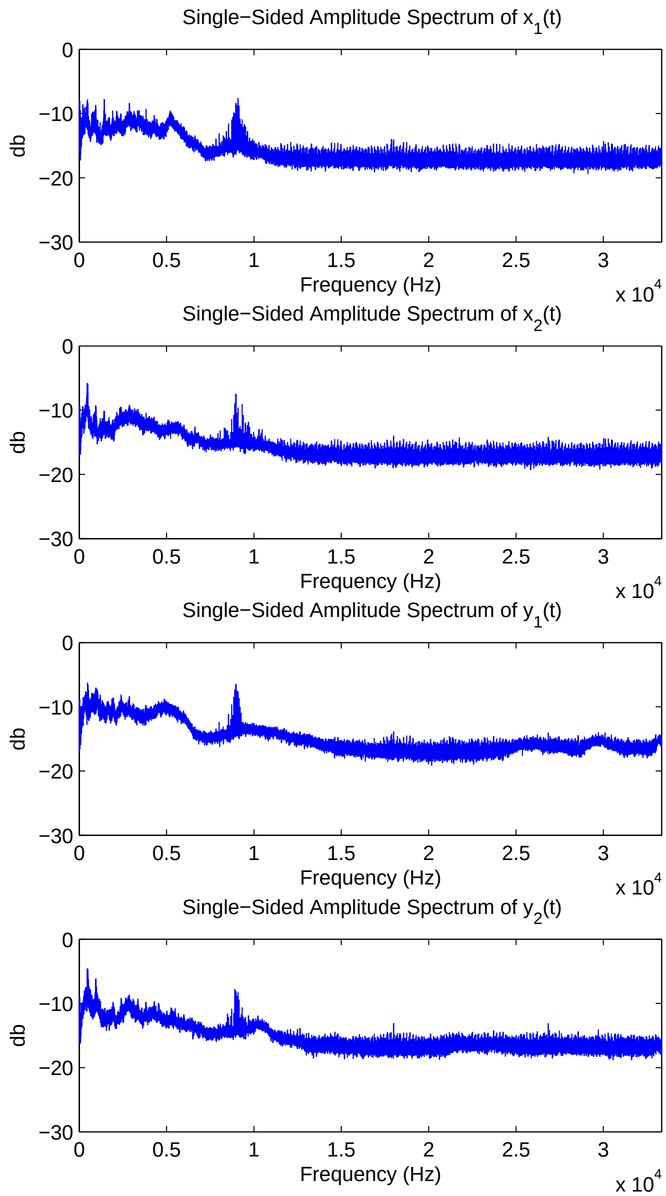

Figure 3.4: Periodograms of healthy and faulty signals. Healthy signal, input accelerometer (top left). Healthy signal, output accelerometer (bottom left). Faulty signal, input accelerometer (top right). Faulty signal, output accelerometer (bottom right). Notice the change in the periodograms from healthy signal to faulty signal.

Table 3.3: Detection powers

Test/Comparison	λ	$\bar{\Delta}$
Healthy (A)-Healthy (B)	96.4%	50.0%
Healthy (A)-Faulty (C)	100.0%	87.2%
Healthy (B)-Faulty (C)	96.4%	88.4%

100.0% of the time while the $\bar{\Delta}$ statistic declares them different 87.2% of the time. When comparing signal B to signal C, the λ statistic declares them different 96.4% of the time while the $\bar{\Delta}$ statistic declares them different 88.4% of the time. Overall, it appears that both statistics capably identified that the signals were born of two different processes.

3.5 Summary

This section compared two multivariate signal discrimination techniques under various scenarios. The likelihood ratio statistic λ rejects signal equality in a reliable manner only when the series are white noise. However, when the series are in truth white noise, the λ statistic has a larger discrimination power than the $\bar{\Delta}$ statistic. In cases where some autocovariances at lags one or more are non-zero, the $\bar{\Delta}$ statistic is more reliable. In fact, a simple VAR(1) case was constructed where the false alarm rate of the λ statistic was approximately 15 times higher than advertised. Applications to an experimental set of gearbox vibrations showed similar structure. Overall, it is wise to base signal equality conclusions on the $\bar{\Delta}$ statistic when the signals are not multivariate white noises.

Chapter 4

Multivariate Time Series Clustering in the Frequency Domain

This chapter extends results from Chapter 3 by exploring the utility of a test statistic in typical clustering and classification algorithms. The discrimination capability of the periodogram-based hypothesis test is adapted for use in an agglomerative hierarchical clustering algorithm and the Nearest Neighbor Rule (NNR) classification algorithm. The results demonstrate the measure's ability to effectively group and classify signals based on their dynamic character.

4.1 Introduction

Various physical processes may be modeled as the output of a dynamic system that has been excited by random or stochastic input. The dynamic system filters the stochastic input to produce a stochastic output. This stochastic output may be considered a time series and some common forms have been surveyed in [122]. The character of the dynamic system is contained in the output. The goal of the research in this chapter is to use the spectral content in the output to cluster and classify the signals. Estimated spectra are well studied [81] and have been used as the basis of comparison in the literature [98, 123], but are extend in this paper. In his 1961 paper [124] Jenkins provides "a simplified account of the motivation behind the spectral analysis of time-series" which provides a thorough introduction to the spectral analysis of stochastic time series. The research presented in this chapter is fundamentally different from much of the published work because it analyzes multidimensional time series. At every instance in time there are d observations. Additionally, this work systematically considers the uncertainly in the signal beyond just measurement uncertainty.

Clustering or classification/discrimination algorithms start by analyzing the similarity or dissimilarity between objects. This measure of similarity is usually refereed to as a distance. The concept of distance may be rather literal, for instance, when attempting to group spatially oriented objects based on their proximity to one another. However, the concept of distance may become more abstract when the features being used for comparison are separated by something other than Euclidean distance like color or shape.

Established clustering algorithms (agglomerative hierarchical clustering) and classification algorithms (NNR) each require some type of similarity/dissimilarity measure to operate. There are numerous well defined distance metrics in the literature but those metrics concerning statistical time series are few. Coates [98] and Shumway [125] published more classical research on discriminating between stationary time series via estimated spectra while De Souza [126], Maharaj [127], Piccolo [128] and Tong [129] focused on parametric and Swanepoel [130] on nonparametric methods for a similar purpose. Authors have even used common statistical techniques to analyses more deterministic signals such as wake shedding patters in wind tunned testing. In 1992 Shaw [131] clustered the spectra of oscillating was shedding patterns. Liao in 2005 [132] published a survey paper that outlines much of the past and (then) present work in the area of time series clustering and classification.

With as popular as some of these techniques have become in the data mining and signal processing literature, they are not without their critics. In 2003, Keogh [133] went on the record as being critical of published advancements in the area of time series analysis and data mining. He claimed that many of the reported advancements are not significant when compared to the variance of results associated with analyzing real world data or changing minor implementation details.

There have also been critiques of recently published work. In his 2005 work, Keogh [134] claims that "clustering of time-series subsequences is meaningless." This comment is directed towards those attempting to cluster ordered subsequences of longer parent series. The claims are somewhat off putting because they are broad but the authors provides specific context for which he feels his claims are valid.

4.2 The Models

Various different time series models will be used as part of a simulation study and are described here. The first model, $\{\mathbf{X}_t\}_1$ is multivariate Gaussian white noise with an identity covariance matrix,

$$\{\mathbf{X}_t\}_1 = \mathbf{Z}_t, \quad \mathbf{Z}_t \sim \mathrm{WN}(0, \mathbf{\Lambda}) \tag{4.1}$$

where

$$\mathbf{\Lambda} = \begin{bmatrix} 1.0 & 0.0 \\ 0.0 & 1.0 \end{bmatrix}.$$

This is perhaps the simplest of all stochastic models.

The second model, $\{\mathbf{X}_t\}_2$, is the same as the second model except for a slightly different covariance matrix in the noise sequence.

$$\{\mathbf{X}_t\}_2 = \mathbf{Z}_t, \quad \mathbf{Z}_t \sim \text{WN}(\mathbf{0}, \mathbf{\Lambda}) \tag{4.2}$$

where

$$\mathbf{\Lambda} = \begin{bmatrix} 1.2 & 0.0 \\ 0.0 & 1.0 \end{bmatrix}.$$

The third model under consideration, $\{\mathbf{X}_t\}_3$, is that of a vector auto-regressive moving-average model with autoregressive order 2 and moving-average order 1, VARMA(2,1) with the autoregressive matrix coefficients

$$\mathbf{\Phi}_1 = \begin{bmatrix} 0.40 & 0.05 \\ 0.05 & 0.30 \end{bmatrix}, \quad \mathbf{\Phi}_2 = \begin{bmatrix} -0.48 & 0.10 \\ 0.10 & -0.06 \end{bmatrix},$$

and the moving-average coefficient matrix

$$\mathbf{\Theta}_1 = \begin{bmatrix} 0.30 & 0.10 \\ 0.10 & 0.50 \end{bmatrix}.$$

A sequence generated by this model satisfies the difference equation

$$\mathbf{X}_t = \mathbf{\Phi}_1\mathbf{X}_{t-1} + \mathbf{\Phi}_2\mathbf{X}_{t-2} + \mathbf{Z}_t + \mathbf{\Theta}_1\mathbf{Z}_{t-1} \tag{4.3}$$

for a corresponding realization of $\mathbf{Z}_t$.

The fourth model, $\{\mathbf{X}_t\}_4$, is chosen to be a vector autoregressive of order one, VAR(1), according to

$$\mathbf{X}_t = \mathbf{\Phi}\mathbf{X}_{t-1} + \mathbf{Z}_t, \tag{4.4}$$

with

$$\mathbf{\Phi} = \begin{bmatrix} 0.90 & 0.10 \\ -0.10 & 0.90 \end{bmatrix}$$

and where $\{\mathbf{Z}_t\}$ is taken as Gaussian white noise with the identity covariance matrix.

Model five, $\{\mathbf{X}_t\}_5$, is a first-order moving-average satisfying

$$\mathbf{X}_t = \mathbf{Z}_t + \mathbf{\Theta}\mathbf{Z}_{t-1}$$

with

$$\mathbf{\Theta} = \begin{bmatrix} 0.70 & 0.30 \\ 0.30 & 0.50 \end{bmatrix},$$

and $\{\mathbf{Z}_t\}$ is taken as Gaussian white noise with the identity covariance matrix.

Model six, $\{\mathbf{X}_t\}_6$, is white noise

$$\{\mathbf{X}_t\}_6 = \mathbf{Z}_t, \quad \mathbf{Z}_t \sim \mathrm{WN}(\mathbf{0}, \mathbf{\Lambda}) \tag{4.5}$$

where

$$\mathbf{\Lambda} = \begin{bmatrix} 1.58 & 0.36 \\ 0.36 & 1.34 \end{bmatrix}.$$

This model is white noise like models one and two but the covariance matrix is not similar to that of the covariance matrix for either model one or model two. Recall models one and two have very similar covariance matrices.

Models seven, denoted $\{\mathbf{X}_t\}_7$, and eight, denoted $\{\mathbf{X}_t\}_8$, are designed to be correlated counterparts to models one and two. The two models are vary similar in their covariance structure but are distinctly different in their underlying dynamics. Model seven is based off an AR(1) model while model eight is based off an MA(1) model. However, both the lag-zero and lag-one autocovariances agree. That is, $\{\mathbf{X}_t\}_7$ and $\{\mathbf{X}_t\}_8$ to have different dynamics, but $\mathbf{\Gamma}_{\mathbf{X}7}(0) = \mathbf{\Gamma}_{\mathbf{X}8}(0)$ and $\mathbf{\Gamma}_{\mathbf{X}7}(1) = \mathbf{\Gamma}_{\mathbf{X}8}(1)$. This can be partially observed in Figure 4.1. The autocorrelation function for the first component of a realization of $\{\mathbf{X}_t\}_7$ and $\{\mathbf{X}_t\}_8$ are shown on the same graph. Notice that the ordinates at lag-0 and lag-1 lie on top of one another but differ beyond that until they both converge to zero. For a detailed discussion of how these signals are generated please refer to Chapter 3.

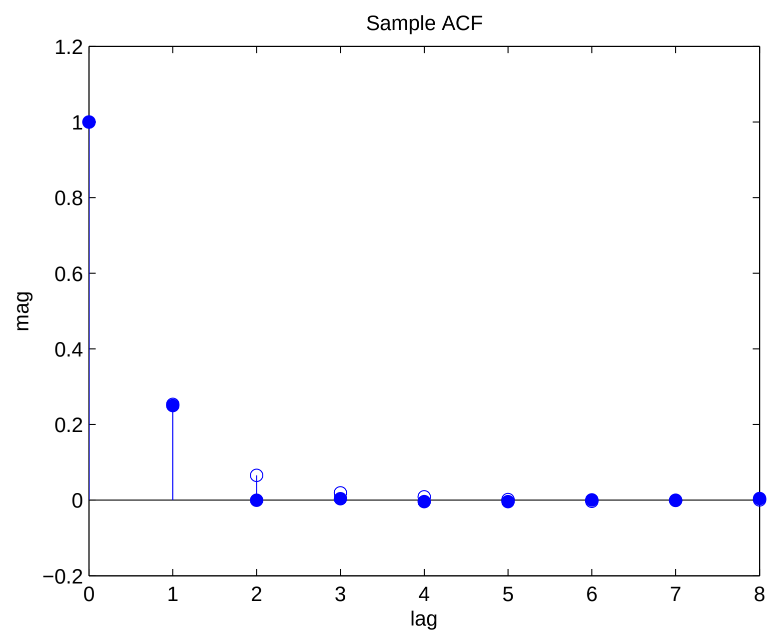

Figure 4.1: ACF for the first component of a realization of models seven and eight. Notice how the covariance is the same for lag-0 and lag-1 but different thereafter. Additional components are excluded to save space.

While there are eight distinct models present, one can make the argument

Table 4.1: Group definitions

<table>
<tr><th colspan="2">Case 1</th><th colspan="2">Case 2</th><th colspan="2">Case 3</th></tr>
<tr><td>G1</td><td>M1</td><td rowspan="2">G1</td><td>M1</td><td rowspan="2">G1</td><td>M1</td></tr>
<tr><td>G2</td><td>M2</td><td>M2</td><td>M2</td></tr>
<tr><td>G3</td><td>M3</td><td>G2</td><td>M3</td><td>G2</td><td>M3</td></tr>
<tr><td>G4</td><td>M4</td><td>G3</td><td>M4</td><td>G3</td><td>M4</td></tr>
<tr><td>G5</td><td>M5</td><td>G4</td><td>M5</td><td>G4</td><td>M5</td></tr>
<tr><td>G6</td><td>M6</td><td>G5</td><td>M6</td><td>G5</td><td>M6</td></tr>
<tr><td>G7</td><td>M7</td><td>G6</td><td>M7</td><td rowspan="2">G6</td><td>M7</td></tr>
<tr><td>G8</td><td>M8</td><td>G7</td><td>M8</td><td>M8</td></tr>
</table>

that models 1 & 2 (zero mean Gaussian white noise with only slightly differing lag-0 covariance matrices) and models 7 & 8 (a VAR(1) and VMA(1) that share the same lag-0 and lag-1 covariance matrices) are dynamically very similar. This dynamic similarity may make it reasonable to consider models 1 & 2 to be grouped together and to group models 7 & 8 together. Whether or not we choose to group models 1 & 2 and models 7 & 8 means that the original eight distinct models may be reasonably segmented into either 8, 7 or 6 distinct groups (see Table 4.1). This will have an impact on both the clustering and classifications.

4.3 Hypothesis Testing

Given two d-dimensional series $\{\mathbf{X}_t\}$ and $\{\mathbf{Y}_t\}$ that have been preprocessed so that they are stationary this paper investigates the ability to cluster these series based on their spectral properties. Consider the hypothesis test

$$
\begin{aligned}
H_o : &= \boldsymbol{\Gamma}_{\mathbf{X}}(0) = \boldsymbol{\Gamma}_{\mathbf{Y}}(0), \quad \forall t && (4.6)\\
H_1 : &= \boldsymbol{\Gamma}_{\mathbf{X}}(0) \neq \boldsymbol{\Gamma}_{\mathbf{Y}}(0), \quad \forall t && (4.7)\\
TS : &= \mathrm{dist}\left(\{\mathbf{X}_t\}, \{\mathbf{Y}_t\}\right) && (4.8)\\
RR : &= \bar{\Delta} > \mu_M + z_\alpha \frac{\sigma_M}{\sqrt{\frac{N}{2} - 1}} && (4.9)
\end{aligned}
$$

where the definition of the rejection region comes from eq. (3.4), $\bar{\Delta}$ is defined in eq. (3.2) and Δ is defined in eq. (3.3).

In this situation, the statistic $\bar{\Delta}$ is interpreted as a distance,

$$\mathrm{dist}\left(\{\mathbf{X}_t\}, \{\mathbf{Y}_t\}\right) = \bar{\Delta}. \tag{4.10}$$

This test was shown, in Chapter 3, to have more favorable discriminations characteristics than a more typical covariance based test.

Take a moment to interpret the distance and the practical significance of it being greater than $\mu_M + z_\alpha \sigma_M / \sqrt{N/2 - 1}$. The distance between two series is itself a random variable (labeled DIST in Figure 4.2). If the two series $\{\mathbf{X}_t\}$ and $\{\mathbf{Y}_t\}$ are identical, that is $\{\mathbf{X}_t\} = \{\mathbf{Y}_t\}$, $\forall t$, then the distance between them will be zero. If the difference between the two series can be accounted for by the random variation of $\mathbf{Z}_t$, then the distance between the two series should fall *below* $\mu_M + z_\alpha \sigma_M / \sqrt{N/2 - 1}$, $(1 - \alpha)\%$ of the time. This fact allows the operator to better interpret the resulting dendrogram from the clustering algorithm. In order for two groups to be considered separate at the α level, the distance between then should be at least $\mu_M + z_\alpha \sigma_M / \sqrt{N/2 - 1}$. Otherwise, the distance between the two groups is likely due to statistical variation *only* and should not be attributed to differing dynamics. Figure 4.2 is meant to help illustrate this point.

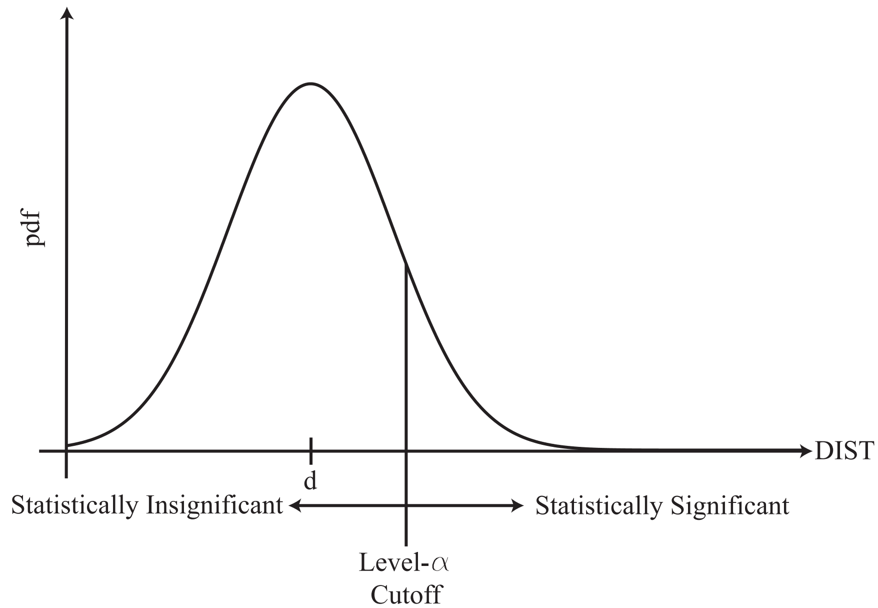

Figure 4.2: The pdf of the random variable DIST. Realizations of DIST greater than the level-α cutoff suggest a statistically significant distance.

This analysis is not meant to replace the interpretation of the structure of the dendrogram. The proposed analysis is meant to provide a "statistical floor" for the dendrogram, under which distances are to be considered either zero or at least not statistically significant. For series that are statistically indistinguishable, the distance will fall within the acceptable region $(1-\alpha)\%$ of the time.

4.4 Simulation Study: Clustering

This section describes the structure and results of the clustering simulation study. All series under investigation have dimensionality two, $d = 2$. In addition the smoothing parameter $M = 5$ (see Chapter 3) and series length $N = 1024$ is kept constant throughout.

Four test series were simulated according to each of the eight models. The resulting thirty two models were then analyzed using an agglomerative hierarchical clustering algorithm. Using eq. (4.10) as the distance measure between elements makes this algorithm different from the a typical agglomerative hierarchical clustering algorithm.

Figure 4.3 is a dendrogram resulting from the hierarchical clustering algorithm. The dendrogram shows considerable separation when sufficiently high up the tree.

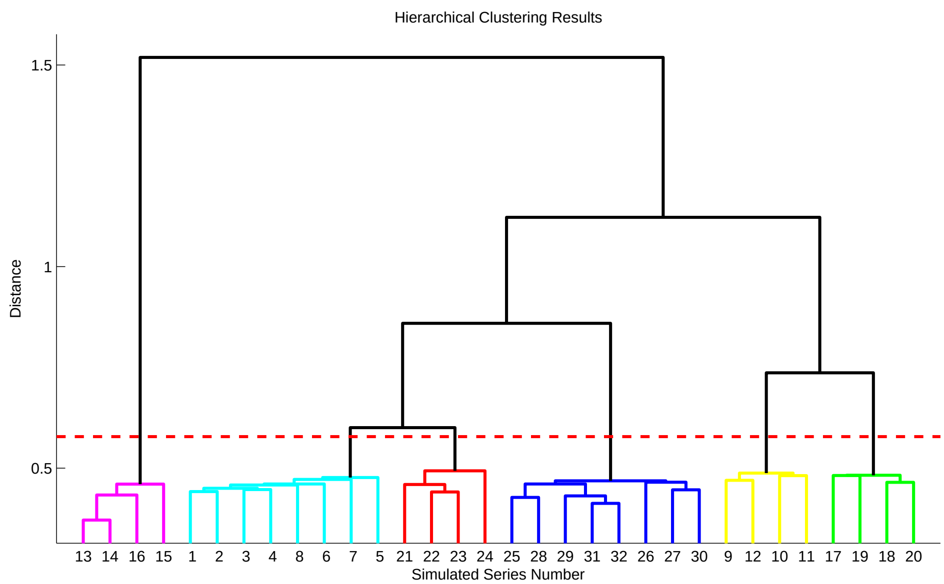

Figure 4.3: Hierarchical Dendrogram showing the natural grouping of the stochastic time series.

The results of the clustering algorithm were quantified using a known scoring metric [132]. Let G be the set of k ground truth clusters and C be the set of clusters resulting from the clustering algorithm. The similarity measure presented in [132] is

$$Sim(G, C) = \frac{1}{k} \sum_{i=1}^{k} \max_{1 \le j \le k} Sim_2(G_i, C_j) \tag{4.11}$$

Table 4.2: Hierarchical clustering results

Assumed Number of Groups	% Error
8 (Case 1)	18.33
7 (Case 2)	10.39
6 (Case 3)	0

where

$$Sim_2(G_i, C_j) = \frac{2|G_i \cap C_j|}{|G_i| + |C_j|}$$

and $|\cdot|$ denotes the cardinality of the set. Note: The subscript "2" has been added for clarification.

This type of analysis is possible because the series are simulated and ground truth is known exactly. The results of the clustering algorithm for all three cases is shown in Table. 4.2. When considering Case 1 (eight groups are assumed) there is 18.33% error. This amount of error is not unexpected because of the known similarity among some of the models. For Case 2 (seven groups are assumed) there is 10.39% error. This case is one where models 1 and 2 are considered to be part of the same group. This merging is justified because of the similarity in the structure of models 1 and 2. For Case 3 (six groups are assumed) there is 0% error. These results highlight the technique's ability to cluster stochastic signals, but also makes it clear that the technique has difficulty distinguishing between series with very similar dynamics.

4.5 Simulation Study: Classification

The nearest neighbor (NNR) algorithm is a supervised classification scheme requiring training data as well as a test data. This algorithm works by computing pairwise distances between each member of the test set and each member of the training set. This is an exhaustive search and has the potential to take a long time depending on the size of both the test set and training set. For the 1-NNR algorithm, each element of the test set is classified as being in the same group as the element closest to it according to the defined distance. There are two basic variations of the NNR algorithm, the 3-NNR and the 5-NNR. The 3- and 5-NNR algorithms uses a voting scheme to assign a label to each element of the teat set based on the 3 and 5 nearest neighbors, respectively. While possible, a tie is quite unlikely and was not encountered during these simulations. Each implementations of the NNR algorithm uses an exhaustive search and are not exceptionally efficient. However, the algorithm performed rather well when classifying the test set.

A test set of five series from each of the eight models (forty series total) and a training set of twenty five series from each of the eight models (two hundred total) were simulated. Each element of the test series was then classified using the 1-, 3- and 5-NNR.

Evaluation of the nearest neighbor classification algorithm is performed by direct comparison of the ground truth and results vector. Let H be the ground truth vector and Q be the classification results. The similarity measure, or percentage of correct classification is

$$Sim(H, Q) = \frac{1}{n} \sum_{i=1}^{n} P_i \tag{4.12}$$

where n is the number of elements in both H and Q and

$$P = (H == Q) \tag{4.13}$$

where "==" denotes the logical comparison between H and G with a Boolean output vector.

Table 4.3 summarizes the results of the NNR clustering algorithm. When all

Table 4.3: Nearest neighbor classification results

Assumed Number of Groups	1-NNR	3-NNR	5-NNR
8	20	22.5	17.5
7	5	5	7.5
6	0	0	0

eight models were assigned a separate groups (Case 1) the percent for the 1-NNR was 20%, for the 3-NNR was 22.5% and for the 5-NNR was 17 .5%. When models 1 and 2 were considered to be part of the same group (Case 2) resulting in only 7 groups total, the percent for the 1-NNR was 5%, for the 3-NNR was 5% and for the 5-NNR was 7.5%. When models 7 and 8 were also considered to be part of the same group (Case 3) the percent error for the 1-NNR was 0%, for the 3-NNR was 0% and for the 5-NNR was 0%.

4.6 Summary

This research explored the ability of a test statistic, based on spectral density estimators, to serve as the similarity/dissimilarity measure in traditional clustering and classification algorithms. The technique was shown to be useful in both an agglomerative hierarchical clustering scheme as well as in an NNR classification scheme.

The clustering scheme was able to group signals with the same or similar dynamics but had difficulty separating those with subtle differences. The same was true for the classification scheme. When signals with similar dynamics were considered members of the same group, the classification scheme performed well. However, classification was poor when the algorithm was expected to distinguish between signals with only subtle differences. This behavior (observed in both schemes) is not unexpected and ultimately due to the short signal length. The basis for comparison is an estimated parameter, or in this case, and estimated function. As the signal length increases, more information about each signal becomes available and allows for more convincing comparisons between signals. With only a limited signal length it is not possible to determine conclusively (statistically significantly) if two signals are truly alike or not alike based on their spectral properties.

Chapter 5

Multivariate Time Series Clustering in the Time Domain

This chapter develops a modeling and analysis routine for multivariate, climatological time series. Data will be analyzed that represent temperature and wind speed variations over a long time period (~ 5 years). The experimental measurements come from offshore buoys located around the United States. After the data are modeled, the prediction errors are analyzed and used as a basis of comparison for grouping signals displaying similar weather patterns. Results show convincing groupings and provide a level-α test for interpreting those groupings.

The techniques are first demonstrated with a simulation study and then implemented on the actual data.

5.1 Introduction

This chapter expands on Chapter 4 by considering non-stationary time series in the time domain. The research successfully clusters climatological time series by comparing prediction errors and using the difference to calculate pairwise distances between each series. These distance measures are used in a hierarchical clustering algorithms to groups the time series. Signals with similar dynamics naturally group together. The approach is first implemented on simulated data as a simulation study and then applied to actual data.

Many of the same references cited in Chapter 4 are relevant to the following work since both are concerned with clustering and classification of time series, but there are some differences. The data in this chapter is being analyzed in the time domain as opposed to the frequency domain, and the raw data is non-stationary. Discrimination of non-stationary series has been considered in [135] but was done so in the frequency domain. Assumptions about the structure of the non-stationarity can be made because it is climatological data. Maharaj [136] investigated the classification of time series using a p-value which was used to make binary decisions regarding signal equality as part of a larger simulation. This approach requires many samples to be effective. Additionally, the effects of signal length were not adequately investigated.

This research includes the development of a level-α interpretation of the computed distances, which are random variables. The level-α interpretation suggests whether or not the separation between two groups is statistically significant.

5.2 Wind Speed Variation and Temperature Modeling

Modeling of wind speed and temperature variation is difficult due to the obvious non-stationarity in both the mean and variance. To address this issue, consider the d-dimensional, non-stationary time series model

$$\{\mathbf{X}_t\} = \boldsymbol{\mu}_{\mathbf{X},t} + \mathbf{S}_{\mathbf{X},t}\mathbf{U}_{\mathbf{X},t}, \quad \{\mathbf{Y}_t\} = \boldsymbol{\mu}_{\mathbf{Y},t} + \mathbf{S}_{\mathbf{Y},t}\mathbf{U}_{\mathbf{Y},t} \tag{5.1}$$

where $\mathbf{X}_t$ and $\mathbf{Y}_t$ are $d \times 1$ vectors, $\boldsymbol{\mu}_{\mathbf{X},t}$ is the $d \times 1$ seasonal mean vector, $\mathbf{S}_{\mathbf{X},t}$ is the $d \times d$, diagonal, seasonal covariance matrix and the subscript $_\mathbf{X}$ and $_\mathbf{Y}$ associate the parameters with their respective series. The term $\mathbf{U}_{\mathbf{X},t}$ is the time series component,

$$\mathbf{U}_{\mathbf{X},t} = \mathbf{\Phi}_\mathbf{X}\mathbf{U}_{\mathbf{X},(t-1)} + \mathbf{Z}_{\mathbf{X},t} \tag{5.2}$$

where $\mathbf{Z}_{\mathbf{X},t}$ is a $d \times 1$ Gaussian random vector and $\mathbf{\Phi}_\mathbf{X}$ is the first order, vector, autoregressive coefficient. The Y component is defined similarly. This model was inspired by the periodic autoregressive models of [137] but was ultimately chosen after visually inspecting the data in the time domain (see Figure 5.1).

The experimental data used in this analysis is made available by The National Oceanic and Atmospheric Administration (NOAA) National Data Buoy Center (NDBC) `http://www.ndbc.noaa.gov/`. This is a United States government entity that publishes the data, in part, for scientific investigation.

5.3 Processing of the Data

The data must be processed rather extensively. These steps include both preprocessing to "clean up" the raw data as well as primary processing to implement the techniques developed in this chapter.

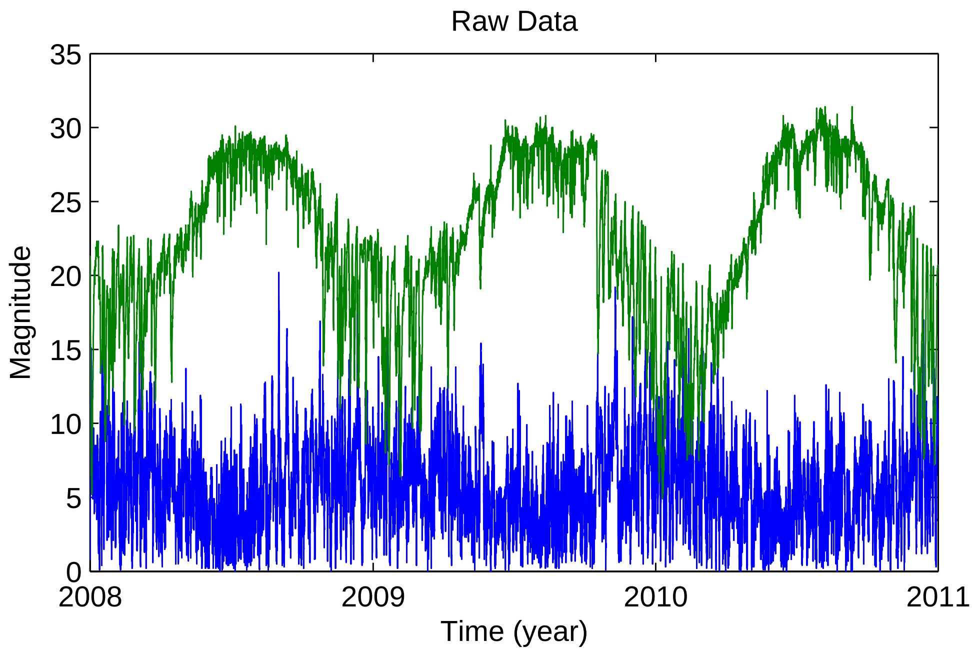

Figure 5.1: Two channels of raw data that are typical for this analysis. The data is hourly averages of temperature (upper) and wind speed (lower) measurements. Only three years are shown for clarity.

5.3.1 Error Removal

Bad or missing data is a reality when analyzing real (non-simulated) experimental data. Data acquisition failure in the form of sensor failure, transmission failure, storage failure, and file corruption may lead to bad or missing data. Preprocessing of the data includes removing any apparent outliers or obviously failed acquisition attempts. Statistical outliers are data points that lie anywhere from 3 to 8 standard deviations outside of the local means. Care must be taken when determining the thresholds. If the raw data is not normally distributed or non-stationary, then using a 3 standard deviation rule of thumb may excise valid data points. Due to the non-stationarity, and non-normality of the data at hand, 8 standard deviations was chosen as the threshold.

There is another situation where no thresholds are necessary and that occurs when the data acquisition system acknowledges that it did not collect a valid data point. Some acquisition systems will "write in" their maximum or minimum values (usually a series of 9s) when an acquisition attempt fails. For instance, data points might display temperature values of 999.99 or wave height values of 99.99. These are clearly incorrect and are analogous to the acquisition system denoting the IEEE's NaN value. If there are errors in any of the channels being analyzed at a particular time, then all data points for that instance of time are removed.

Once bad or missing data is removed, it is ignored rather than interpolated in the subsequent analysis. Incorporating interpolated data into statistical calculations has consequence. By ignoring missing data, we reduce the number of data points used to estimate parameters but avoid basing calculations on samples that are directly related to one another.

5.3.2 Daily Averaging

The analysis focuses on long term behavior of weather patterns and the hourly resolution offered by the raw data is too fine a scale. All data points in the block of time 12:00:00am-11:59:59pm on a given day are averaged to form what will now be referred to as *daily data*. If that day contained less that 6 valid data points, then the entire day was considered to be an error and was removed from analysis. This operation had the effect of removing high frequency content present in the signal and producing a lower acquisition error frequency. The effects of producing the daily averages can be seen in Figure 5.2.

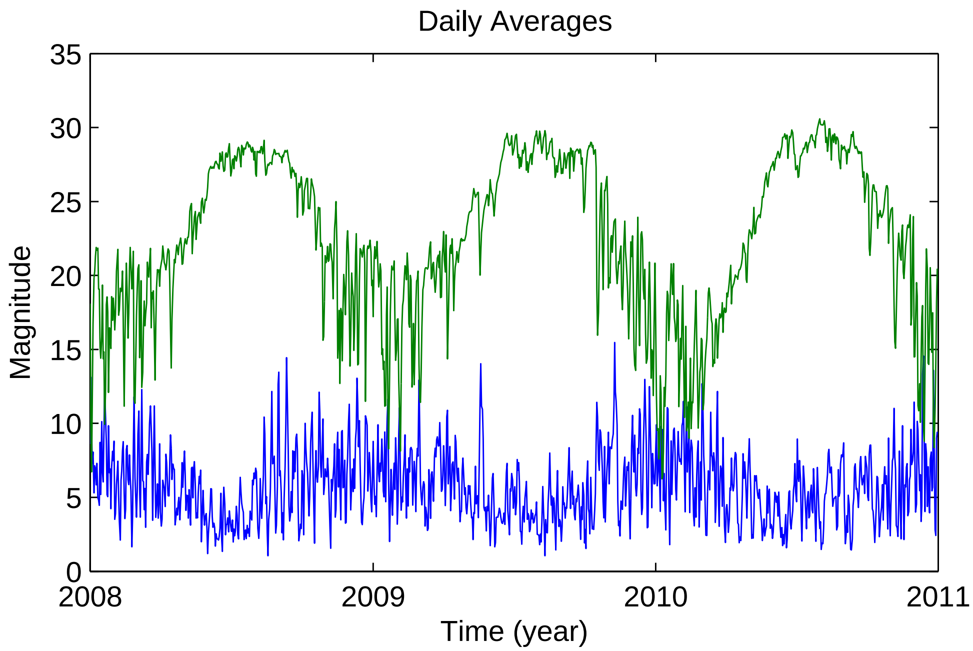

Figure 5.2: The hourly measurements are averaged over each 24 hour period to form daily averages or daily data. This operation reduces the high frequency content in the signal and reduces the occurrence of missing data points.

5.3.3 Seasonal Mean

The first component estimated is the seasonal mean and it is determined using least squares regression of a sinusoidal function onto the daily averages. Some of the the seasonal means in the data set are observed to deviate from a pure sinusoid. The seasonal means are modeled as forth order Fourier series to accommodate this deviation,

$$\mu_{1,t} = c + \sum_{k=1}^{3} a_k \sin\left(\frac{2\pi tk}{365}\right) + b_k \cos\left(\frac{2\pi tk}{365}\right). \tag{5.3}$$

Each dimension of the raw data is converted into a general linear model and the coefficients are determined using the Moore-Penrose pseudoinverse [138]. Figure 5.3 shows the mean function fit to the the daily data and the effects of subtracting that mean.

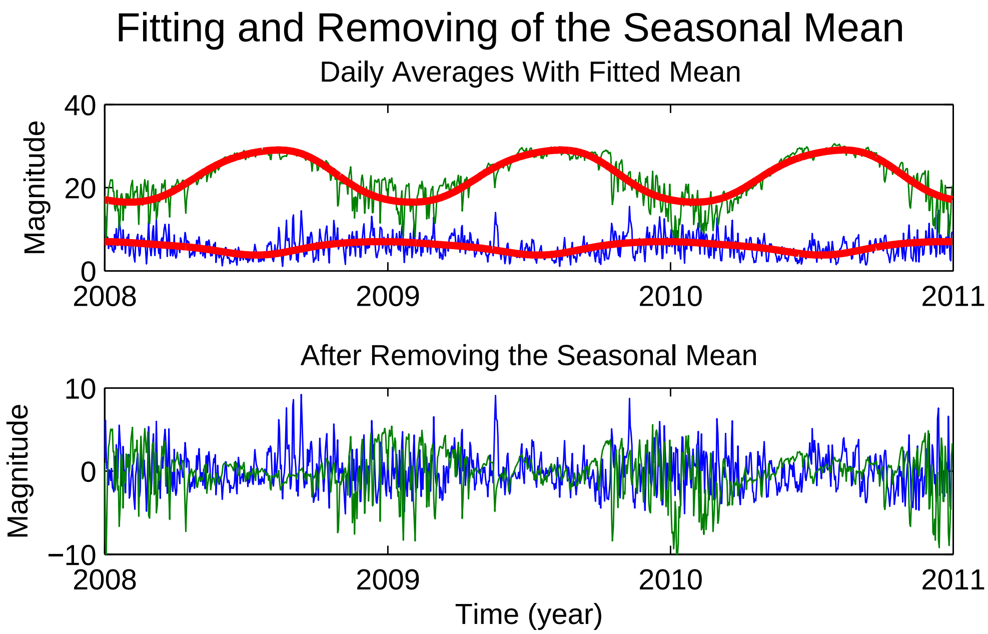

Figure 5.3: Daily averages on temperature and wind speed data fitted with their seasonal means (upper). Daily averages with the seasonal mean subtracted (lower). Notice the resulting series is clearly not stationary.

The seasonal mean appears to be removed, however, the data is clearly non-stationary and has a statistical variance that changes with time. This seasonal variance is estimated for each dimension.

5.3.4 Seasonal Standard Deviation

The **S** term in eq. (5.1) represents the signal's time varying standard deviation on any given day for any given year. The most direct way of computing this value is to compute the variance of data points corresponding to the same day for each year. For instance, one could compute the seasonal standard deviation for January 1st by collecting the data points from January 1st of each year and using them to estimate the standard deviation. This may be appropriate for very long signals (30 years or more). For the signals in this study that are only 5 years long, it results in very uncertain variance estimates (N=5 or less if there are missing data points).

One can improve the estimates by assuming that the seasonal standard deviation varies sinusoidally with a period of one year. This reduces the number of parameters to be estimated (for each signal) from 366 to 3. Additionally, even if there is a long data set, not imposing any assumptions about the form of $\mathbf{S}$ makes it extremely hard to differentiate the seasonal standard deviation from the standard deviation of the stochastic component since they are multiplied by one another. The seasonal standard deviation matrix is assumed to be of the form

$$\mathbf{S}_t = \operatorname{diag}\left[1 + a_1 \sin(2\pi t/365 + b_1),\ \cdots,\ 1 + a_i \sin(2\pi t/365 + b_i)\right] \tag{5.4}$$

which has periodic terms on the main diagonal and zeros elsewhere. Notice the effect this form has on the seasonal variance. If any of the a_i's are zero then the ith element of $\mathbf{U}_t$ will be multiplied by 1 and will not have a seasonal variance. However, non-zero a_i's will cause the ith element of $\mathbf{U}_t$ to be multiplied by a function that is oscillating between a value above 1 and a value below 1. This creates something of an undulating envelope around the stochastic portion of the series. The diagonal form implies that seasonal variance structure of series i will not influence series j when $i \neq j$.

The $\mathbf{S}_t$ term is estimated by processing the mean corrected data

$$\mathbf{X}_{mc,t} = \mathbf{X}_t - \hat{\boldsymbol{\mu}}_{\mathbf{X},t} \tag{5.5}$$

(defined similarly for the Y series) where the subscript mc refers to "mean corrected." Assuming the model form in eq. (5.1), the mean corrected data is

$$\mathbf{X}_{mc,t} = \mathbf{S}_{\mathbf{X},t}\mathbf{U}_{\mathbf{X},t} \tag{5.6}$$

where $\mathbf{U}_{\mathbf{X},t}$ is a zero-mean, stationary series. The variance of the mean corrected data is

$$\begin{aligned}
\text{Var}\left[(\mathbf{X}_{mc,t})(\mathbf{X}_{mc,t})'\right] &= E\left[(\mathbf{S}_{\mathbf{X},t}\mathbf{U}_{\mathbf{X},t})(\mathbf{S}_{\mathbf{X},t}\mathbf{U}_{\mathbf{X},t})'\right] \\
&= E\left[\mathbf{S}_{\mathbf{X},t}\mathbf{U}_{\mathbf{X},t}\mathbf{U}'_{\mathbf{X},t}\mathbf{S}'_{\mathbf{X},t}\right] \\
\mathbf{\Gamma}_{\mathbf{X},mc}(0) &= \mathbf{S}_{\mathbf{X},t}\mathbf{\Gamma}_{\mathbf{U},\mathbf{X}}(0)\mathbf{S}'_{\mathbf{X},t} \qquad (5.7)
\end{aligned}$$

where $\mathbf{\Gamma}_{\mathbf{U},\mathbf{X}}(0)$ is the lag-0 covariance of the $\mathbf{U}_{\mathbf{X},t}$ series.

Recall eq. (5.4) and examine the diagonal elements of eq. (5.7),

$$\gamma_{i,i,X,mc} = (1 + a_i sin(2\pi t/365 + b_i))^2 \gamma_{i,i,U}. \qquad (5.8)$$

where $\gamma_{i,i,X,mc}$ is computed from the mean corrected data and $\gamma_{i,i,U}$ is a constant representing the auto-covariance of a stationary time series. The square root of eq. (5.8),

$$s_{i,i,\mathbf{X},mc} = \sqrt{(1 + a_i sin(2\pi t/365 + b_i))^2}\ s_{i,i,\mathbf{U}} \qquad (5.9)$$

is fit with a three parameter curve using least squares regression. The three unknown parameters for each curve are a_i, b_i, $s_{i,i,\mathbf{U}}$. Figure 5.4 (upper) shows $s_{i,i,\mathbf{X},mc}$ for $i = [1, 2]$ (each component of that series). Notice the oscillating behavior. In the same figure, the two solid lines are the diagonal elements of $\mathbf{S}_t$ for that series.

There were some curve fitting difficulties worth mentioning. Equation 5.9 was used for curve fitting over eq. (5.8) due to the nature of the data. Not only does the value of the estimator $\hat{\gamma}_{i,i,\mathbf{X},mc}$ vary with time but so does its variance (see Figure 5.4). The varying, variance of $\hat{\gamma}_{i,i,\mathbf{X},mc}$ led to a particular problem with the least squares regression. The square of the errors is much larger in regions of large variance. These larger squared errors would cause the regression algorithm to weight more heavily the points in the region of larger variance at the expense of the data in the region of low variance. As a result, fitting the square root of eq. (5.9) lessened the effects of the varying variance.

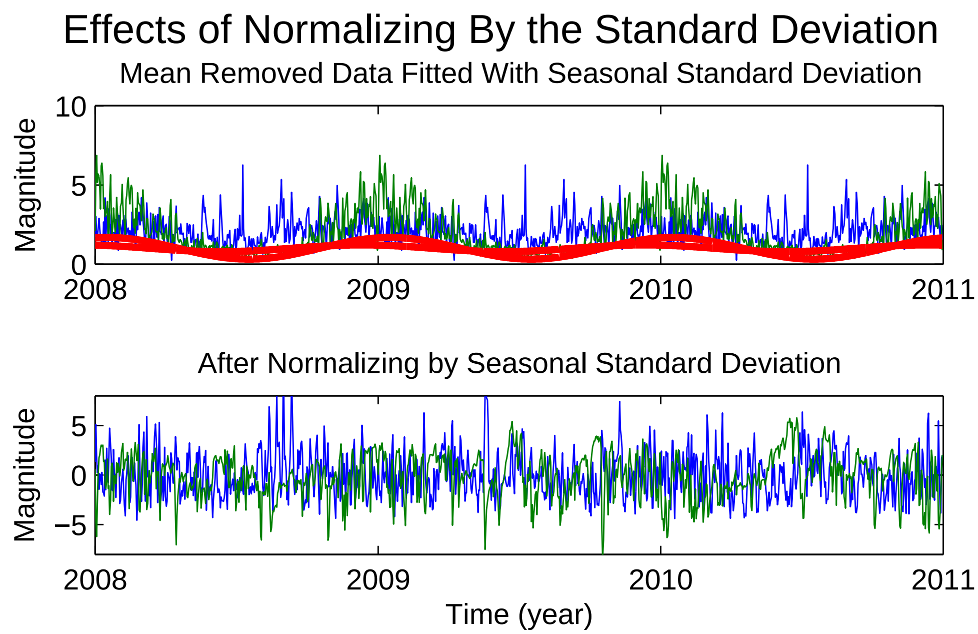

Figure 5.4: Daily standard deviations of mean-removed data (upper). Notice that these values vary in both their mean *and* variance. The thick line is meant to represent the time varying, multiplicative constant and not the least squares fit of the data itself. Mean corrected data that has been normalized by the seasonal standard deviation is shown at the bottom.

5.3.5 Stochastic Modeling

Once the deterministic components, $\boldsymbol{\mu}_{\mathbf{X},t}$ and $\mathbf{S}_{\mathbf{X},t}$, are estimated, the original series is processed to isolate the stochastic component of the time series

$$\mathbf{U}_{\mathbf{X},t} = \hat{\mathbf{S}}_{\mathbf{X},t}^{-1} \left(\mathbf{X}_t - \hat{\boldsymbol{\mu}}_{\mathbf{X},t} \right). \tag{5.10}$$

The resulting series is assumed to be VAR(1)

$$\mathbf{U}_{\mathbf{X},t} = \mathbf{\Phi}_{\mathbf{X}}\mathbf{U}_{\mathbf{X},t-1} + \mathbf{Z}_{\mathbf{X},t}, \quad \{\mathbf{Z}_{\mathbf{X},t}\} \sim \mathrm{WN}(\mathbf{0}, \mathbf{\Lambda}_{\mathbf{X}}) \tag{5.11}$$

and is modeled using multidimensional time series theory with two unknown parameters [3, 87]; the AR coefficient, $\mathbf{\Phi}_{\mathbf{X}}$, and the multidimensional white noise variance, $\mathbf{\Lambda}_{\mathbf{X}}$. The lag-1 covariance matrix

$$\mathbf{\Gamma}_{\mathbf{X}}(1) = E\left[\mathbf{X}_{t+1}\mathbf{X}_t'\right] = \mathbf{\Phi}_{\mathbf{X}}\mathbf{\Gamma}_{\mathbf{X}}(0), \tag{5.12}$$

and the auto-covariance matrix

$$\mathbf{\Gamma}_{\mathbf{X}}(0) = \mathrm{Var}(\mathbf{X}_t) = \mathbf{\Phi}_{\mathbf{X}}\mathbf{\Gamma}_{\mathbf{X}}(0)\mathbf{\Phi}_{\mathbf{X}}' + \mathbf{\Lambda}_{\mathbf{X}}, \tag{5.13}$$

constitute a system of two equations (eq. (5.12) and eq. (5.13)) and two unknowns ($\mathbf{\Phi}_{\mathbf{X}}$ and $\mathbf{\Lambda}_{\mathbf{X}}$). Estimates of the lag-0 and lag-1 covariance matrices,

$$\hat{\mathbf{\Gamma}}(0) = \frac{1}{N}\sum_{t=1}^{N}\mathbf{X}_t\mathbf{X}_t' \quad , \quad \hat{\mathbf{\Gamma}}(1) = \frac{1}{N}\sum_{t=1}^{N-1}\mathbf{X}_{t+1}\mathbf{X}_t', \tag{5.14}$$

may be used in eq. (5.12) and eq. (5.13) to estimate $\hat{\mathbf{\Phi}}_{\mathbf{X}}$ and $\hat{\mathbf{\Lambda}}_{\mathbf{X}}$.

5.4 Time Series Comparison

The signals are compared to one another by evaluating the probability that a pair of signals were "born" from the same stochastic process. This analysis involves performing a hypothesis test on the residuals of each series. A hierarchical clustering algorithm will be used to cluster the statistic values.

This proposed method is applied to both a simulation study as well as to multivariate climatological data.

5.4.1 Hypothesis Testing

After the model parameters have been estimated, a hypothesis test is designed to test signal equality [139]. Consider the hypothesis test

$$H_o : \; = \; \text{The two series have the same dynamics} \tag{5.15}$$

$$H_1 : \; = \; \text{Not } H_o \tag{5.16}$$

$$TS : \; = \; \text{dist}\left(\{\mathbf{X}_t\}, \{\mathbf{Y}_t\}\right) \tag{5.17}$$

$$RR : \; = \; \text{dist}\left(\{\mathbf{X}_t\}, \{\mathbf{Y}_t\}\right) > d + z_\alpha \sqrt{\frac{2d}{N}} \tag{5.18}$$

where

$$\text{dist}\left(\{\mathbf{X}_t\}, \{\mathbf{Y}_t\}\right) = \frac{1}{2N} \sum_{t=1}^{N} \left(\mathbf{R}_{\mathbf{X},t} - \mathbf{R}_{\mathbf{Y},t}\right)' \left(\mathbf{R}_{\mathbf{X},t} - \mathbf{R}_{\mathbf{Y},t}\right) \tag{5.19}$$

and $\{\mathbf{R}_{\mathbf{X},t}\}$ and $\{\mathbf{R}_{\mathbf{Y},t}\}$ are residuals computed from a time series model assuming H_o is true (discussed in Section 5.4.2). The null hypothesis states that the two series have the same dynamics and, therefore, have the same model parameters. Model parameters are estimated from each of the two series being tested and then averaged to form the assumed system model. That is,

$$\boldsymbol{\mu}_t = \frac{\boldsymbol{\mu}_{\mathbf{X},t} + \boldsymbol{\mu}_{\mathbf{Y},t}}{2}, \; \mathbf{S}_t = \frac{\mathbf{S}_{\mathbf{X},t} + \mathbf{S}_{\mathbf{Y},t}}{2}, \; \boldsymbol{\Phi} = \frac{\boldsymbol{\Phi}_{\mathbf{X}} + \boldsymbol{\Phi}_{\mathbf{Y}}}{2}, \; \boldsymbol{\Lambda} = \frac{\boldsymbol{\Lambda}_{\mathbf{X}} + \boldsymbol{\Lambda}_{\mathbf{Y}}}{2}. \tag{5.20}$$

5.4.2 Development of Statistic

The distance between two time series is based on the distribution of prediction errors for two series, $\{\mathbf{X}_t\}$ and $\{\mathbf{Y}_t\}$, assuming that they have the same modeling parameters (assuming H_o). The one step ahead predictor for $\{\mathbf{X}_t\}$ is denoted

$$P_t\mathbf{X}_{t+1} = P(\mathbf{X}_{t+1}|\mathbf{X}_t, \mathbf{X}_{t-1}, \ldots, \mathbf{X}_1). \tag{5.21}$$

This predictor is constructed as

$$P_t\mathbf{X}_{t+1} = \hat{\boldsymbol{\mu}}_{\mathbf{X},t+1} + \hat{\mathbf{S}}_{\mathbf{X},t+1}\hat{\mathbf{U}}_{\mathbf{X},t+1} \tag{5.22}$$

where

$$\hat{\mathbf{U}}_{\mathbf{X},t+1} = \hat{\boldsymbol{\Phi}}_{\mathbf{X}}\hat{\mathbf{U}}_{\mathbf{X},t} \tag{5.23}$$

and

$$\hat{\mathbf{U}}_{\mathbf{X},t} = \hat{\mathbf{S}}_{\mathbf{X},t}^{-1}\left(\mathbf{X}_t - \hat{\boldsymbol{\mu}}_{\mathbf{X},t}\right) \tag{5.24}$$

This construction allows the prediction of $\mathbf{X}_{t+1}$ from $\mathbf{X}_t$. The final prediction equation for the $\{\mathbf{X}_t\}$ series is

$$P_t\mathbf{X}_{t+1} = \hat{\boldsymbol{\mu}}_{\mathbf{X},t+1} + \hat{\mathbf{S}}_{\mathbf{X},t+1}\hat{\boldsymbol{\Phi}}_{\mathbf{X}}\left[\hat{\mathbf{S}}_{\mathbf{X},t}^{-1}\left(\mathbf{X}_t - \hat{\boldsymbol{\mu}}_{\mathbf{X},t}\right)\right] \tag{5.25}$$

and the predictor for the $\{\mathbf{Y}_t\}$ series can be expressed in a similar manner.

The mean and variance of the predictor in eq. (5.21) are

$$\begin{aligned}
E[\mathbf{X}_{t+1}-\hat{\mathbf{X}}_{t+1}] &= E\left[\mathbf{X}_{t+1}-\hat{\mathbf{X}}_{t+1}\right] \\
&= E\left[\mathbf{X}_{t+1}-\left[\hat{\boldsymbol{\mu}}_{\mathbf{X},t+1}+\hat{\mathbf{S}}_{\mathbf{X},t+1}\hat{\boldsymbol{\Phi}}_{\mathbf{X}}\left[\hat{\mathbf{S}}_{\mathbf{X},t}^{-1}\left[\mathbf{X}_t-\hat{\boldsymbol{\mu}}_{\mathbf{X},t}\right]\right]\right]\right] \\
&= E\left[\boldsymbol{\mu}_{t+1}+\mathbf{S}_{t+1}\mathbf{U}_{t+1}-\left[\hat{\boldsymbol{\mu}}_{\mathbf{X},t+1}+\hat{\mathbf{S}}_{\mathbf{X},t+1}\hat{\boldsymbol{\Phi}}_{\mathbf{X}}\hat{\mathbf{U}}_{\mathbf{X},t}\right]\right] \\
&= E\left[\mathbf{S}_{t+1}\mathbf{U}_{t+1}-\mathbf{S}_{\mathbf{X},t+1}\hat{\boldsymbol{\Phi}}_{\mathbf{X}}\hat{\mathbf{U}}_{\mathbf{X},t}\right] \\
&= E\left[\mathbf{S}_{\mathbf{X},t+1}\mathbf{Z}_{t+1}\right] \\
&= \mathbf{0} \qquad (5.26)
\end{aligned}$$

and

$$\begin{aligned}
\mathrm{Var}\left[P_t\mathbf{X}_{t+1}\right] &= \mathrm{Var}\left[\mathbf{X}_{t+1}-\hat{\mathbf{X}}_{t+1}\right] \\
&= \mathrm{Var}\left[\mathbf{X}_{t+1}-\left[\hat{\boldsymbol{\mu}}_{\mathbf{X},t+1}+\hat{\mathbf{S}}_{\mathbf{X},t+1}\hat{\boldsymbol{\Phi}}_{\mathbf{X}}\left[\hat{\mathbf{S}}_{\mathbf{X},t}^{-1}\left[\mathbf{X}_t-\hat{\boldsymbol{\mu}}_{\mathbf{X},t}\right]\right]\right]\right] \\
&= \mathrm{Var}\left[\boldsymbol{\mu}_{t+1}+\mathbf{S}_{t+1}\mathbf{U}_{t+1}-\left[\hat{\boldsymbol{\mu}}_{\mathbf{X},t+1}+\hat{\mathbf{S}}_{\mathbf{X},t+1}\hat{\boldsymbol{\Phi}}_{\mathbf{X}}\hat{\mathbf{U}}_{\mathbf{X},t}\right]\right] \\
&= \mathrm{Var}\left[\mathbf{S}_{t+1}\mathbf{U}_{t+1}-\mathbf{S}_{\mathbf{X},t+1}\hat{\boldsymbol{\Phi}}_{\mathbf{X}}\hat{\mathbf{U}}_{\mathbf{X},t}\right] \\
&= \mathrm{Var}\left[\mathbf{S}_{\mathbf{X},t+1}\mathbf{Z}_{t+1}\right] \\
&= \mathbf{S}_{\mathbf{X},t+1}\boldsymbol{\Lambda}\mathbf{S}'_{\mathbf{X},t+1} \qquad (5.27)
\end{aligned}$$

Assuming that the time series modeling approach is sufficient, the residuals are uncorrelated in time although they may still have lag-zero correlation. Let the one-step-ahead, normalized prediction error for the $\{\mathbf{X}_t\}$ series be

$$\mathbf{R}_{\mathbf{X},t} = \hat{\boldsymbol{\eta}}_{\mathbf{X}}^{-1/2}(\mathbf{X}_t - \hat{\mathbf{X}}_t) \qquad (5.28)$$

where

$$\hat{\boldsymbol{\eta}}_{\mathbf{X}} = \mathrm{Var}(\mathbf{X}_t - \hat{\mathbf{X}}_t) \tag{5.29}$$

The multi variable prediction errors are distributed standard, multivariate normal, $d \times d$,

$$\mathbf{R}_{\mathbf{X},t} \sim \mathbf{R}_{\mathbf{Y},t} \sim N(\mathbf{0}, \mathbf{I}_{d\times d}) \tag{5.30}$$

as such the difference of the errors will have the distribution

$$\mathbf{R}_{\mathbf{X},t} - \mathbf{R}_{\mathbf{Y},t} \sim N(\mathbf{0}, \mathbf{I}_{d\times d}) - N(\mathbf{0}, \mathbf{I}_{d\times d}) \sim N(\mathbf{0}, 2\mathbf{I}_{d\times d}). \tag{5.31}$$

The chi-squared variable can be constructed as

$$\frac{(\mathbf{R}_{\mathbf{X},t} - \mathbf{R}_{\mathbf{Y},t})'}{\sqrt{2}} \frac{(\mathbf{R}_{\mathbf{X},t} - \mathbf{R}_{\mathbf{Y},t})}{\sqrt{2}} \sim \chi^2(d) \tag{5.32}$$

with mean d and variance $2d$.

5.4.3 Distance

Equation 5.32 shows that the residuals (at every time t) can be used to construct a chi-squared random variable with with mean d and variance $2d$. Applying the central limit theorem to the sum in eq. (5.19) results in a random variable that is distributed asymptotically normal,

$$\mathrm{dist}\left(\{\mathbf{X}_t\}, \{\mathbf{Y}_t\}\right) \sim AN\left(d, \frac{2d}{N}\right) \tag{5.33}$$

allowing rejection of the null hypothesis when

$$\mathrm{dist}\left(\{\mathbf{X}_t\}, \{\mathbf{Y}_t\}\right) > d + z_\alpha \sqrt{\frac{2d}{N}} \tag{5.34}$$

This comparison is being carried out as a one sided test to accommodate correlation among the signals. When two signals demonstrate correlation, their prediction residuals tend to be less than when the relationship is purely structural. While this implies the signals may not be independent it should not take away from identifying these signals as similar. Not only are the signals similar in dynamics structure but they are actually correlated with one another. This means small distances (those significantly less than d) should be allowed to support signal equality rather than reject it. This concept is again demonstrated in Figure 4.2. Distances between the origin and the level-α cutoff are considered statistically insignificant even though excessively low values would suggest a relation that is beyond structural.

5.4.4 Noise Correlation

One of the statistical modeling assumptions is that the noise sequence, $\mathbf{Z}_t$, which drives the stochastic portion of the model, $\mathbf{U}_t$, is independent of the noise sequence in any other model. The residuals in eq. (5.28) are normally distributed and *assumed* to be uncorrelated with the residuals from any other series. However, $\mathbf{R}_{\mathbf{X},t}$ and $\mathbf{R}_{\mathbf{Y},t}$ have demonstrated correlation when the series $\{\mathbf{X}\}$ and $\{\mathbf{Y}\}$ are collected at the same time and from similar geographic regions. In this case, eq. (5.32) is not true. One could circumvent this problem by modifying eq. (5.32) as follows:

$$\frac{(\mathbf{R}_{\mathbf{X},t+s} - \mathbf{R}_{\mathbf{Y},t})'}{\sqrt{2}} \frac{(\mathbf{R}_{\mathbf{X},t+s} - \mathbf{R}_{\mathbf{Y},t})}{\sqrt{2}} \sim \chi^2(d) \tag{5.35}$$

where $s > p$, and p is the largest lag for which the covariance function is non-zero.

5.5 Series Clustering

Multidimensional signals with similar dynamic characteristics should be considered members of the same group while signals that do not share dynamic character should be excluded from that group. Determining these groupings is the subject of clustering. The hypothesis test developed in Section 5.4.1 produces a statistic with a well defined mean and variance. This statistic may be interpreted as a distance between two series; it is small when the series share the same dynamic structure and large otherwise.

5.5.1 Simulation Study

A simulation study was performed to evaluate the performance of the proposed pattern recognition scheme. Six different series of daily averages are simulated in accordance with the model described in Section 5.2 and subsequently processed and clustered as described in Section 5.3 and Section 5.4.

The series are simulated to resemble the actual buoy data. The six simulated series will have a mean of the form

$$\boldsymbol{\mu}_t = \begin{bmatrix} \mu_{1,t} \\ \mu_{2,t} \\ \vdots \\ \mu_{N,t} \end{bmatrix} = \begin{bmatrix} a_1 + b_1 \sin(\omega t + c_1) \\ a_2 + b_2 \sin(\omega t + c_2) \\ \vdots \\ a_N + b_N \sin(\omega t + c_N) \end{bmatrix} \tag{5.36}$$

Table 5.1: Simulated series parameters

		Series Number					
Parameter	Component	1	2	3	4	5	6
a	1	8	8	8	8	8	8
a	2	30	30	50	50	50	50
b	1	1	1	1	1	1	1
b	2	10	10	10	10	20	20
c	1	0	0	0	0	0	0
c	2	0	0	0	0	0	0
e	1	0.4	0.4	0.3	0.3	0.1	0.1
e	2	0.1	0.1	0.2	0.2	0.7	0.7
f	1	π	π	π	π	π	π
f	2	π	π	π	π	π	π

and a standard deviation of the form

$$\mathbf{S}_t = \text{diag} \begin{bmatrix} 1 + e_1 \sin(\omega t + f_1) \\ 1 + e_2 \sin(\omega t + f_2) \\ \vdots \\ 1 + e_N \sin(\omega t + f_N) \end{bmatrix} \tag{5.37}$$

where the constants are defined in Table 5.1

The autoregressive and noise parameters for the simulated series are defined as follows

$$\text{Group 1: } \mathbf{\Phi}_1 = \mathbf{\Phi}_2 = \begin{bmatrix} 0.7 & 0.1 \\ -0.1 & 0.3 \end{bmatrix}, \quad \mathbf{\Lambda}_1 = \mathbf{\Lambda}_2 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \tag{5.38}$$

$$\text{Group 2: } \mathbf{\Phi}_3 = \mathbf{\Phi}_4 = \begin{bmatrix} 0.5 & 0.1 \\ -0.1 & 0.3 \end{bmatrix}, \quad \mathbf{\Lambda}_3 = \mathbf{\Lambda}_4 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \tag{5.39}$$

$$\text{Group 3: } \mathbf{\Phi}_5 = \mathbf{\Phi}_6 = \begin{bmatrix} 0.7 & 0.1 \\ -0.1 & 0.9 \end{bmatrix}, \quad \mathbf{\Lambda}_5 = \mathbf{\Lambda}_6 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \tag{5.40}$$

These parameter values form 3 distinct groups. Group 1 consists of series 1 & 2, group 2 consists of series 3 & 4, and group 3 consists of series 5 & 6. Additionally, one may consider groups 2 & 3 to be similar to one another based on the relatively small difference in parameter values when compared to group 1. This structure is expected to show up in the clustering dendrogram. A plot of the daily averages for series 1 is shown in Figure 5.5.

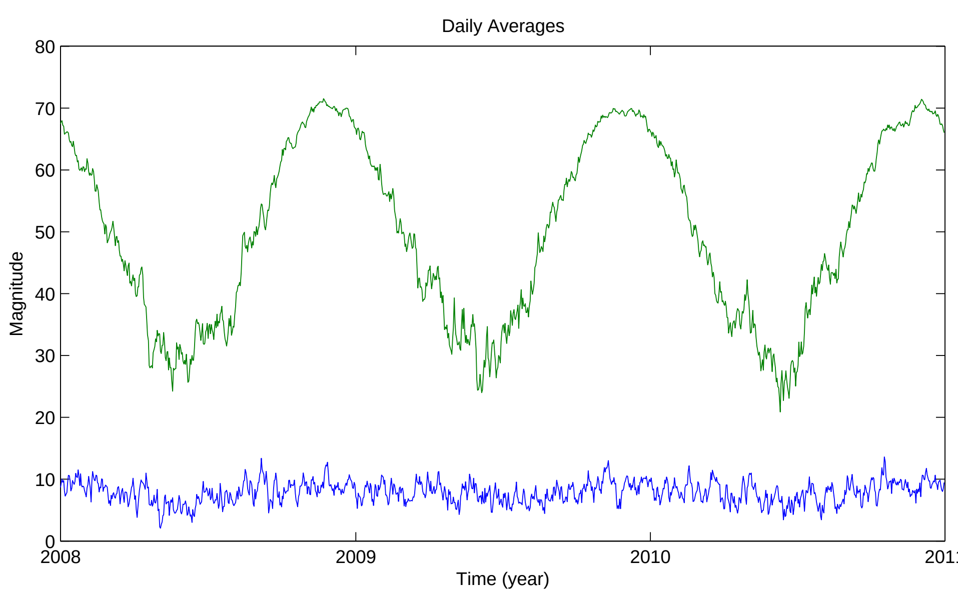

Figure 5.5: Plot of the simulated daily averages used in the analysis with the temperature shown above and the wind speed shown below.

The results of the modeling and clustering scheme described in Sections 5.3 and Section 5.4 are shown with the dendrogram Figure 5.6. Samples are arranged on the abscissa axis and the height of a horizontal bar above two samples indicates their relative distance to one another. The agglomerative clustering scheme starts with each sample in its own group and merges groups based on their proximity to one another. Once two groups are merged, the distance of the resulting group to the remaining groups is computed and used to form the next branch of the dendrogram. This process is repeated until all groups have been merged into a single group.

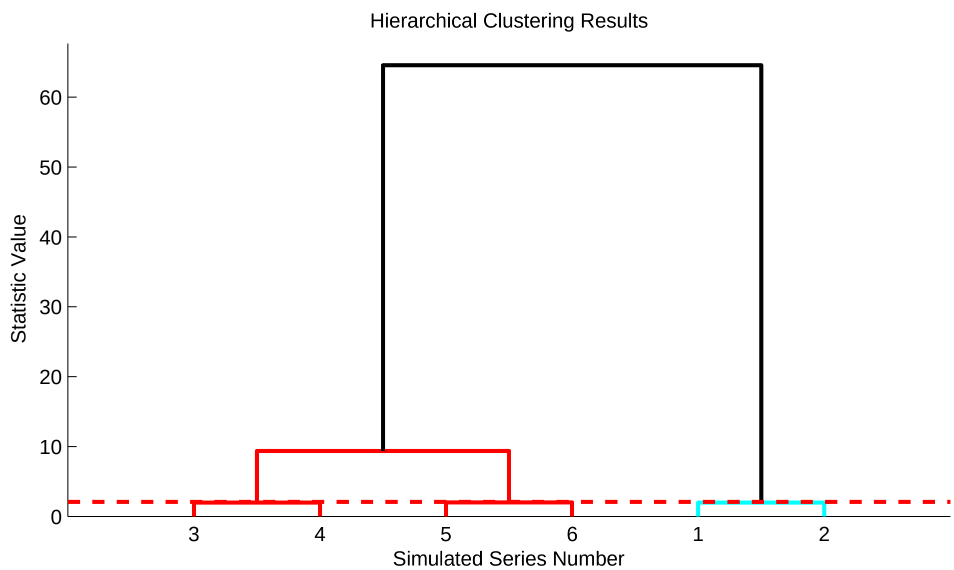

Figure 5.6: Results of hierarchical clustering using the raw statistic as the distance measure for the simulated data.

For two series generated from the same statistical process, the expected value of the distance defined in eq. (5.19) is d. Indeed Figure 5.6 shows that the value of the statistic is very small between series 1 & 2, 3 & 4, and 5 & 6. Also, the distance between the clusters consisting of series 3 & 4 (group 2) and series 5 & 6 (group 3) is small when compared to the distance to group 1 (which consists of series 1 & 2).

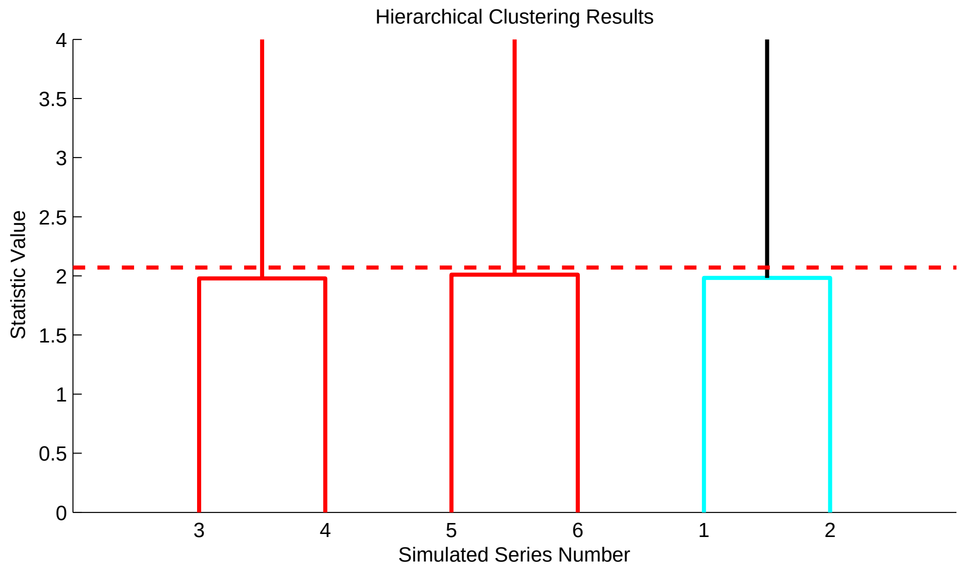

Figure 5.7: A reprint of Figure 5.6 with rescaled ordinate axis for clarity.

5.5.2 Multidimensional Climatological Data Clustering

Attention is now turned to the experimental data. Climatological data represents the behavior of weather events as they evolve over time. Successfully identifying weather patterns with dynamics may be useful in exploiting weather dependent power sources such as wind. Data made available from The National Oceanic and Atmospheric Administration (NOAA) National Data Buoy Center (NDBC) was used in this analysis. The NDBC collects and makes available climatological data from numerous buoy weather stations located around the world.

This study focused on data collected from eleven buoys in the United States' coastal regions whose exact locations can be found in Table 5.2. These buoys collect multiple measurements and report their values hourly.

Figure 5.8 and Figure 5.9 show the result of performing hierarchical clustering of statistic values resulting from pairwise comparison of the buoy data. Figure 5.8 makes use of 4 years worth of data while Figure 5.9 makes use of 6 years worth of

Table 5.2: Buoy locations

Buoy #	NDBC #	Lat/Long	General Location
1	46054	34.274N 120.459W	West of Santa Barbara, CA
2	46025	33.749N 119.053W	West Southwest of Santa Monica, CA
3	46086	32.491N 118.034W	San Clemente Basin, CA
4	46042	36.785N 122.469W	West of Monterey Bay, CA
5	42039	28.791N 86.008W	East Southeast of Pensacola, FL
6	42020	26.966N 96.695W	Southeast of Corpus Christi, TX
7	42036	28.500N 84.517W	West Northwest of Tampa, FL
8	41008	31.402N 80.869W	Southeast of Savannah, GA
9	41004	32.501N 79.099W	Southeast of Charleston, SC
10	44025	40.250N 73.167W	South of Islip, NY
11	44008	40.502N 69.247W	Southeast of Nantucket, MA

data. The dendrograms show two well defined groups evidenced by the long primary branches. The different signal lengths were used to show how differing signal lengths will affect the shape of the dendrogram.

The clusters defined in Figure 5.8 and Figure 5.9 identify clusters that are loosely related to geographic location. Buoys 1-4 are all from the coast of southern California. Buoys 5-9 are are form the Gulf of Mexico and eastern Florida and Buoys 10-11 are from the New England coast. Somewhat counter intuitively, however, is the fact that buoys form the coast of New England tested very similar to buoys from the west coast and that buoys form inside the Gulf of Mexico tested similar to those on the eastern cost of Florida. Being from very different water masses (Northern Atlantic vs. Pacific and Gulf vs. Atlantic), one may expect the temperature and wind character to be quite different but the analysis would suggest otherwise.

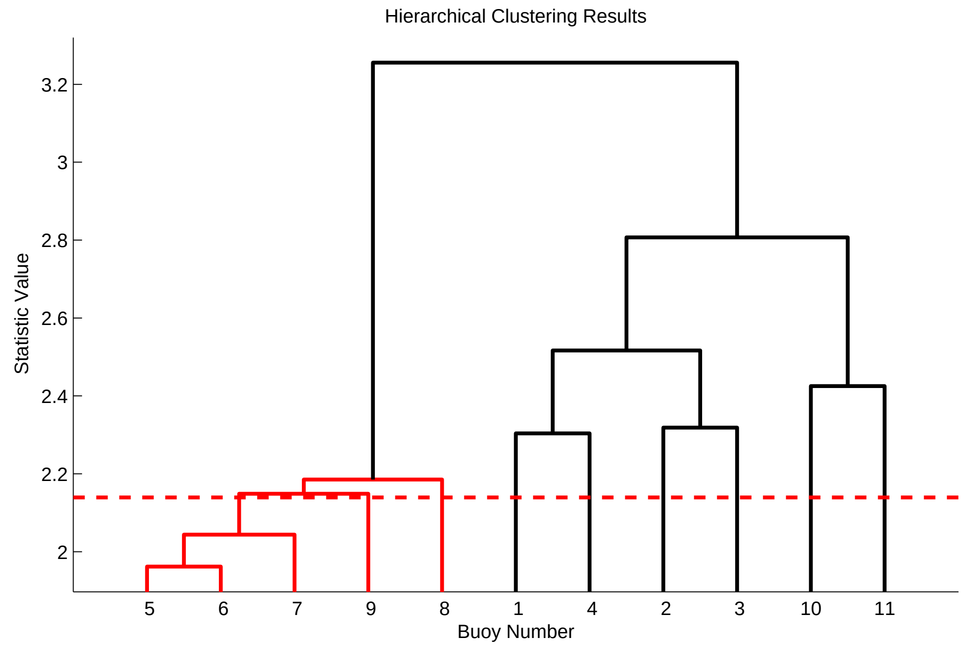

Figure 5.8: Results of hierarchical clustering using the raw statistic as the distance measure based on 4 years of climatological data.

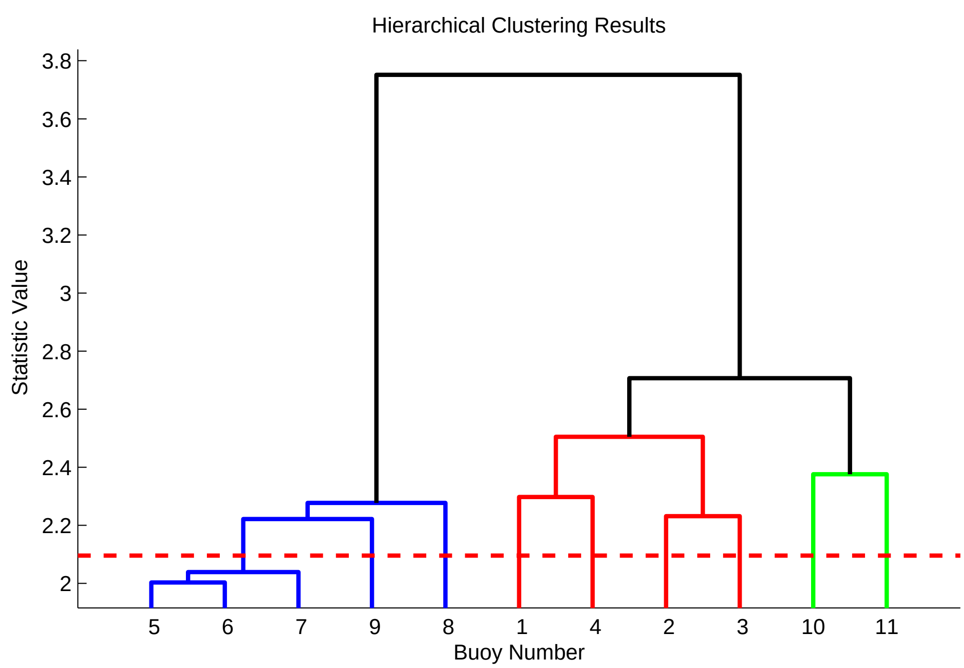

Figure 5.9: Results of hierarchical clustering using the raw statistic as the distance measure based on 6 years of climatological data.

5.6 Summary

This chapter investigated the capability of a traditional hierarchical clustering algorithm where the distance metric was replaced by a non-traditional, multi-dimensional statistic. First, a model for non-stationary, but seasonal, climatological data was created. Second, a statistic was formed that made use of the particular model structure. This statistic had desirable properties in the sense that it behaved predictably with varying signal length. A simulation study based on the developed models was performed demonstrating the utility of the clustering scheme. Finally, the techniques were applied to climatological data made available by the The National Oceanic and Atmospheric Administration (NOAA) National Data Buoy Center (NDBC) which revealed a consistent clustering pattern.

The results of the modeling and clustering efforts were favorable. The techniques were able to determine, at level-α, when two multivariate series displayed the same dynamics. For those series that appear to favor the alternative hypothesis, a question seems to remain open. To what degree is the alternative hypothesis true?

Generally speaking, clustering and classification is concerned with "pattern similarity" [83] as opposed to pattern equality like hypothesis testing. Pattern recognition paradigms allow for members of distinctly different groups to be grouped together based on their similarity even if they are not the same.

The statistic or distance described in eq. (5.19) has properties that are well understood when the null hypothesis is true. "Well understood" refers to the derivation outlined in Section 5.4.2. But one would still like to draw conclusions regarding signal similarity when the *alternative* hypothesis is true. The statistical behavior of the distance is not known exactly when the signals favor H_1 and leaves open the opportunity for future work.

Chapter 6

Conclusions, Contributions, and Future Work

The research in this dissertation is reviewed from a global perspective. The main activities are discussed making note of their individual results and contributions to the literature. Additionally, opportunities for future research that have emerged as part of this dissertation are discussed.

6.1 Conclusions and Contributions

This research primarily contained three main parts. Chapter 3 was an investigation of the effectiveness of a periodogram based test statistic over a more traditional covariance based test. The hypothesis was that the periodogram based test would be more effective in discriminating signals with differing dynamics. The reasoning was that the periodogram based test used more information (the entire covariance function) to test similarity than the covariance test which only used the lag-0 covariance value.

When the signals being analyzed were white noise and did not have a covariance structure beyond the lag-0 covariance matrix, the traditional covariance based test was more decisive in distinguishing signals with different dynamics. On the other hand, when the signals had a non-zero covariance structure, the periodogram based test was more effective at identifying differences in the dynamic structure. The AR, MA, and ARMA series fell into this category of signals with non-zero covariance structures. The take-away from this portion of the research was that a "one size fits all" approach is not best and that signal character is an important consideration. If the signals were stochastic time series with non-trivial covariance structures then the periodogram test offered improved performance. Otherwise, the traditional covariance based test was appropriate. The most appropriate test was the one that aligns with the signals dynamics.

Chapter 4 investigated the use of a periodogram based metric as the primary distance measure in two common clustering and classification algorithms. The algorithms were used to analyze data sets in the frequency domain. Numerous series were simulated, analyzed and clustered as part of a simulation study. The simulated series were stationary and linear processes but came from a variety of dynamic structures. Results were favorable and showed that a periodogram based measure performed well in the algorithms.

Chapter 5 explored time series clustering in the time domain with an application specific to climatological data. In addition to simulated series, data made available by The National Oceanic and Atmospheric Administration (NOAA) National Data Buoy Center (NDBC) `http://www.ndbc.noaa.gov/` was analyzed. This data was not simulated and provided a nice complement to the simulation based work presented as part of this research. Once again, the results were favorable for both the simulated series and the real data, but the real data helped to uncover a shortcomings of the approach. It appeared that the null hypothesis of signal equality may be overly restrictive for clustering and classification algorithms.

In total, contributions to multivariate signal discrimination and clustering were made in both the time and frequency domains. These techniques expand on existing techniques most noticeably by considering dimensions greater than 1 and by considering the auto-correlation structure in a signal's random component.

Additionally, contributions in engineering education and course design were documented in Appendix A. This study cataloged successful improvements made to the department of mechanical engineering's undergraduate laboratories. After shortcomings were identified, improvements were proposed and then implemented. These improvements had positive effect that were measured and recorded with student surveys.

6.2 Future Work

6.2.1 Wind Speed Modeling

ARMA based models have been used for modeling wind speed variation with time [140–142], yet it is widely agreed upon that wind speed variation is both non-linear and non-stationary. These two conditions violate properties of the ARMA model. One feature of wind speed variation that stands out as inconsistent with ARMA modeling is asymmetry. Wind speed (magnitude), $v(t)$, must be greater than or equal to zero and is (assumed) unbounded,

$$0 \leq v(t) \leq \infty. \tag{6.1}$$

This is consistent with the fact that wind speed variation is routinely observed to be Weibull distributed [7, 8, 143]. Not only is the number of observations above the mean different from those below the mean, but the regions on either side of the mean have different domains. ARMA processes do not model asymmetric signals well. Furthermore, ARMA models have no way of implementing the natural floor at zero observed in with speed variation. Opportunities exist for the creation of models that better mimic and predict the nonlinear and non-stationary behavior of wind speed variation.

6.2.2 Signals of Different Length

All the tests described this research required the signals to be the same length. This is a restrictive constraint. However, the literature concerning hypothesis testing of signals that are of different lengths is scarce.

Signal lengths had to be the same for frequency domain analysis because the analysis is a point wise comparison of periodogram ordinates. If the signals are different lengths then the ordinates are calculated for different Fourier frequencies and a one-to-one comparison is not possible. There have been efforts to address different signal length by interpolating periodogram ordinates, but the results leave room for improvement.

This research also performed analysis in the time domain where prediction errors or residuals formed the basis of comparison. At every instance of time, t, the residuals from two different models were computed and analyzed. Here, the two signals could have different lengths but comparisons were made only at instances of time when both series had a valid data point. The time domain of the resulting signal was the union of the valid time domains of the individual signals. This situation is different form the length constraints in the frequency domain analysis but is still a limitation of the proposed techniques.

Future work is required to determine how best to map a stochastic series onto a space where that mapping is independent of the series it is being compared to. This space would serve as an intermediary. Ideally, this space would also accommodate series of different lengths. The longer the stochastic series, the more confident one can be about the series' location in that space. Confidence intervals, or confidence hyperspheres could give an indication of how confident one is with the calculated distance.

6.2.3 Noise Similarity

As discussed in Section 5.4.4, an observation was made during the course of the research that some of the stochastic series being analyzed appeared to be correlated with one another. This may be explained by correlation in the noise sequence of the two series. For example, consider the two seres

$$x_t = \phi_x x_{t-1} + z_t \tag{6.2}$$

$$y_t = \phi_y y_{t-1} + q_t \tag{6.3}$$

where z_t and q_t are the white noise components of their respective series. Usually it is assumed that the two noise sequences are uncorelated,

$$\text{cov}(z, q) = 0 \tag{6.4}$$

making the two series, x and y, uncorrelated with one another.

Chapter 5 considered signals that were collected at the same time from different geographic locations. Depending on the proximity of these locations one could argue for correlation among the signals. The limiting case for this scenario is that the temperature and wind speed measurements are made from the exact location in space and time. In this case, it would be easy to believe that the two signals are in fact the *same*. As the geographic locations begin to separate, perhaps 100 miles, it would be more reasonable to assume that the two series are independent from one another.

Another area where this scenario may arise is in fault detection and diagnosis of redundant sensors. The name "redundant sensor" implies that more than one sensor is being used to measure the same phenomenon. However, depending on the actual proximity of the sensors to one another, the recorded signals may only be correlated with one another as opposed to being the same. Being able to assess the degree to which these acquired signals are correlated is pertinent to properly analyzing the signals. Being able to assess the dynamic similarity of signals that exhibit correlation in their noise processes is a relevant problem and should be investigated further.

Appendix A

An Improved Undergraduate Mechanical Engineering Laboratory Structure and Curriculum: Design and Assessment

The mechanical engineering department at Clemson University re-evaluated their undergraduate laboratory experience and focused on improving various aspects of the three required laboratory courses. The faculty believe that these laboratory courses are a defining feature of the bachelor of science degree as many graduates accept entry level manufacturing positions or pursue graduate studies. The mechanical engineering laboratory courses at Clemson are stand alone offerings in the undergraduate program in contrast to other schools which attach the labs to select courses. This structure allows a variety of experiments to be offered during each course which can encompass various scientific and engineering topics. This chapter reviews various changes to the laboratories, describes their implementation, and presents assessment results as to their effectiveness. Some of the improvements include the development of printed student and teaching assistant manuals, the development of a unified training program for the teaching assistants, introduction of new laboratory equipment & experiments and revision of the current laboratory documentation. To evaluate the effectiveness of the implemented changes survey results were analyzed. Overall, student satisfaction with the course has improved significantly as evidenced by the survey results.

A.1 Introduction and Laboratory Evolution

In recent history, the undergraduate mechanical engineering laboratory sequence at Clemson University was comprised of four (1995 to 2006) and three (2006 to present) required courses (six credit hours total) as part of the accredited mechanical engineering bachelor of science curriculum. These courses were stand alone classes and were not required to be taken concurrently with any particular core (non-laboratory) course. This type of course structure has been discussed by Roppel et al. [144]. Traditionally, laboratory courses are offered as co-requisites to be completed alongside a lecture course. Also, the advancement in personal computing and internet availability has encouraged some universities to offer laboratories on-line [145–149].

Prior to 1995, the laboratory courses were offered in conjunction with specific lecture courses. In August 1995, the department faculty adopted a new laboratory curriculum which was comprised of four sequential laboratory courses (ME 221, ME 322, ME 323 & ME 424) where each new laboratory course was to include experiments in dynamic systems, materials processing, solid mechanics and thermal fluid sciences.

ME 221 was to be taken during the second semester of the sophomore year, ME 322 & 323 to be taken the junior year, and ME 424 was to be taken the first semester senior year. ME 221 was first offered during the spring semester 1997 so that the class of 1999 completed the sequence first.

ME 221 came to be known, informally, as the *discovery laboratory* where students were exposed to basic mechanical principles and guided through critical analysis activities with instructor-provided questions. ME 322 & ME 323 were two closely related laboratories focused on the steady-state behavior of thermo-fluid and mechanical dynamic systems. ME 424, the terminal laboratory, required students to undertake more experimental design and to determine their own procedures as opposed to having it listed for them.

In September 2004, the mechanical engineering faculty decided on another curriculum change affecting the undergraduate laboratories. Under the newly proposed curriculum, the number of undergraduate laboratory courses was reduced to three, a sophomore level ME 222, a junior level ME 333 and a senior level ME 444. The new curriculum was to be implemented during the 2005-2006 academic year.

ME 222 was to remain the discovery laboratory while ME 333 would focus on the Thermal Fluid Sciences (TFS) and ME 444 would be reserved for Dynamic Systems and Controls (DSC). The content would be delivered through a problem based learning approach and evaluations would be performed via written technical documents. The complexity and sophistication of these technical documents would be expected to keep pace with the students' developing academic maturity.

Today, ME 222 is a hands on laboratory focused on exposing sophomore engineering students to basic mechanical systems as well as the fundamentals of technical writing. ME 333 investigates thermal-fluid systems while expecting more detailed and insightful reports. ME 444 takes the most open ended approach to investigating dynamic system and control type experiments requiring students to design, execute and analyze their activities from start to finish.

A.2 Catalyst of Change

The department of mechanical engineering regularly administers end-of-the-semester student surveys in the laboratory as part of both self- and external-review requirements. The accrediting agencies the Accreditation Board for Engineering and Technology (ABET) and the Southern Association of Colleges and Schools/Commission on Colleges (SACS/COC), periodically evaluate the mechanical engineering curriculum and university, including the laboratories, to make decisions regarding accreditation. The survey presented in this chapter is part of the evaluation process.

Internally, the department will use those same survey results as part of a self-evaluation process performed at the end of each semester. This evaluation is not part of an accreditation process rather it is used to assess strengths and weaknesses within the department on an ongoing basis. As a goal, the department has set an 80% positive response rate for the survey questions. Circa 2007, the percentages of positive responses for most questions were well below 80% and therefore unsatisfactory. Particularly troubling for the department were the response rates concerning writing and statistics.

These poor survey results indicated a less than satisfactory laboratory experience for undergraduates, prompting the department to take action. In January 2008, a number of changes were proposed (refer to Table A.1) based on faculty, teaching assistant and student feedback to address the under performing laboratories. These actions should ideally improve the quality of the laboratories which would then be evidenced by more positive survey responses. The following paragraph briefly describes the proposed changes.

Teaching Assistant (TA) training helps the TAs to become familiar with the experiments/equipment, grow accustomed to teaching and interact more effectively with the students. The printed manuals (both student and TA) contain the information necessary to complete the laboratories which reduces confusion and keeps students "on the same page." Clearly stating (printing on the first page of each laboratory assignment) the learning objectives for each laboratory experiment helps students put their activities in context and gives them an academic compass. Specifically, students understand not only how to perform the experiment but why they are performing the experiment. Requiring students to produce comprehensive laboratory reports gives them an opportunity to organize and then defend their thought process. Designing the experiments to be progressively more and more open ended challenges students throughout the laboratory experience. Incorporating a statistics and uncertainty component into the experiments helps students to understand the limitations of their analysis. Finally, integrating more modern laboratory equipment allows students to gain hands on experience with industry standard hardware and software.

The objective of these laboratory course improvements is to help students better apply the concepts and skills from lectures to out-of-context engineering problems drawn from real world applications.

The laboratory experience encourages students to talk with one another, use their hands, use equipment to solve problems, and think about what they are doing. This engagement among students, or lack thereof, has been addressed by [150]. The proposed improvements are primarily aimed at student involvement. The reformed laboratory offers an environment where the students are given a problem and the required support to address and, hopefully, solve that problem. Part of these sugges-

Table A.1: Proposed improvements

Number	Improvement Description
1	Implement a formal TA training program
2	Formalize (print and bind) all student materials into a student manual
3	Develop a comprehensive set of TA course lecture notes (TA manual)
4	List the learning objectives associated with each laboratory
5	Require more thorough reports as the final deliverable for each experiment
6	Incorporate statistics and uncertainty component into many of the experiments
7	Design the experiments to be progressively more and more open ended as the students mature
8	Integrate more modern laboratory equipment

Table A.2: Intended outcomes

Number	Outcome Description
1	Students develop more in depth understanding of the lecture and laboratory material
2	Students gain exposure to modern engineering tools (hardware and software)
3	Improved student communication skills
4	Students appreciate some of the limitations of engineering theory and of experimental work
5	Students are better able to design and conduct experiments

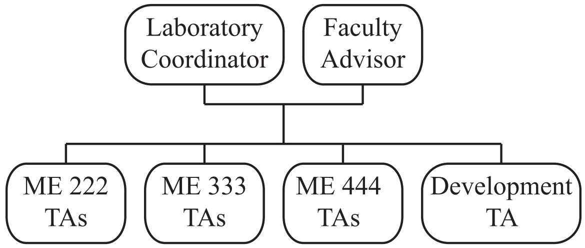

Figure A.1: Organization chart for the undergraduate laboratories showing the relationship between the faculty members/coordinators and the teaching assistants

tions include incorporating engineering innovations into the classroom/laboratory as discussed by [151]. In addition to students successfully leaning the material presented in laboratory, the students should *want* to learn the material. The issue of motivation in the laboratory was studied by [152] who believe that the laboratory experiments become more engaging when they address practical problems.

The laboratory is managed and delivered by 22 individuals; 1 laboratory coordinator, 1 faculty advisor, 18 TAs and 1 laboratory development TA. The organization chart for these individuals is shown in Figure A.1.

The laboratory coordinator is a faculty member whose primary responsibility is the undergraduate laboratories. The faculty advisor is a professor whose responsibilities include an undergraduate laboratory course. The TAs interact with the students six hours a week and deliver the laboratories. Lastly, the laboratory development TA is a doctoral graduate student who assists with programming and critical development activities associated with the laboratories.

A.3 Description of Laboratory Content and Course Modifications

The proposed course changes were implemented by revising the individual laboratories, producing a printed and bound student manual, producing a printed and bound TA manual, standardizing and revising specific deliverables, purchasing and using modern laboratory equipment, phasing in new laboratory experiments and incorporating a statistics and uncertainty component into each experiment. The details of these activities and objectives will be described in this section.

A.3.1 Course Content

The demands of the sophomore, junior and senior level laboratories grow with the progression of the students through the BSME program; the overall objectives align with those outlined in the literature [153, 154]. The sophomore (ME 222) level laboratory focuses on exposing students to engineering concepts through hands on experiences. As the students move to the junior level laboratory (ME 333) they are met with greater demands in terms of drawing from theoretical content developed in other courses and organizing their ideas in laboratory reports. And in the final/senior level laboratory (ME 444), the students are given the most open-ended problems and the least amount of direct guidance.

The experiments performed during each one of these laboratory courses are briefly outlined in Tables A.3, A.4 & A.5.

Table A.3: ME 222 course outline

Laboratory	Description
Metrology	Students are introduced to high precision measuring devices (vernier caliper, micrometer, etc.) and asked to assess tolerances on an internal combustion engine.
Machine Shop	Students are introduced to the basic elements of a machine shop and required to become proficient on those devices.
Reverse Engineering	Students are exposed to the process of reverse engineering whereby they disassemble, analyze and comment on a commercial vacuum cleaner.
Calibration	Students observe the effects of proper and improper calibration and the effects it has the resulting measurements.
Flow Loop	Students analyze the fluid flow characteristics of a liquid level system composed of tanks, pumps, valves and flow meters.
Deformation of Materials	Students study the concepts of material deformation and observe the influence of geometric discontinuities on the distribution of stress.
Tensile Testing	Students perform a standard tensile test and perform statistical analysis on their results.
Torsion	Students perform a standard torsion test and perform statistical analysis on their results.

Table A.4: ME 333 course outline

Laboratory	Description
Introduction to Data Acquisition	Students use software to create a graphical user interface for data acquisition and to measure and store voltage signals.
Data Acquisition System I: Temperature Sensors	Students identify the strengths and weaknesses of thermocouples and thermistors by analyzing their sensitivity, resolution and time response.
Data Acquisition System II: Cam Follower	Student evaluate commonly utilized numerical methods; identify and explain the importance of sampling rate in measuring dynamic signals; and apply numerical methods to analyze data.
Cylinder Experiment: Stationary and Rotating	Students discuss lift generation on a rotating cylinder by manipulating the flow; they identify and explain the flow characteristics of a viscous fluid flowing over a stationary and rotating cylinder-shaped body.
Heat Exchanger	Students identify the parameters affecting the convective heat transfer coefficient by performing empirical determinations.
Wings Lab: Airfoil & Delta Wing	Students identify and explain the flow characteristics of a viscous fluid flowing over an airfoil and a delta wing.
HVAC	Identify the major components of conventional HVAC drying units and discuss the underlying analysis and design assumptions.

Table A.5: ME 444 course outline

Laboratory	Description
Introduction to PLCs: Home Security System	Students apply and program programmable logic controllers to recreate a home security system with ladder logic.
A Dynamic Vibration Absorber: Modeling, Test and Analysis	Students analyze and manipulate a two degree of freedom vibratory system and tune parameters to design a vibration absorber.
Fatigue Testing	Students perform a fatigue test and share data to compile a laboratory database for analysis against published values.
Analysis of a Convection Cooled Electronic System Enclosure	A heater-container combination emulates an enclosed electronic device and students model and analyze the thermal properties of the device and make recommendations regarding its cooling efficiency.
Computer Numerically Controlled (CNC) Machining	Students create tool paths to be executed by a 3-axis HAAS milling machine with attention focused on overall efficiency and cutting profile.

A.3.2 Course Modifications

A methodological approach has been pursued over the past four years to modify the laboratory courses. The continued delivery of the laboratory courses required those changes to be gradually introduced and then refined based on evaluation results.

A.3.2.1 Training & Continuing Education

One of the proposed changes was to implement a formal training program for new and returning TAs, which at the time did not exist. This training program now consists of an intensive one and a half week training exercise that takes place just before the start of the fall semester. Further, weekly meetings that are held between the TAs and the appropriate faculty coordinator continue the training throughout the academic year.

The pre-semester training is particularly effective because returning TAs are involved in delivering the training rather than just receiving it. This arrangement allows returning TAs to recall and sharpen their skills by delivering the training while simultaneously demonstrating to new TAs what is expected of them. Additionally this arrangement minimizes the weekly demands on the undergraduate laboratory coordinator and the faculty coordinator in delivering the courses.

The weekly meetings keep everyone on the same page. Ideally, students receive the same, high level of instruction regardless of what TAs is teaching the course. This level of uniformity is very difficult to achieve, especially if there is little communication among TAs. By meeting on a weekly basis the TAs are able to make sure they are presenting similar material and following a common time line. These meetings are overseen and approved by the course instructor of record.

A.3.2.2 Publications

The mechanical engineering department had gradually moved away from a printed laboratory manual over the last decade in favor of delivering laboratory materials through the on line course management tool such as Blackboard®. More recently, however, the department has favored printed laboratory materials and has decided to produce a spiral bound laboratory manual. This document (specific to each laboratory) is referred to as the *Student Manual* and includes all the material the students will need throughout the course.

The department also commissioned the creation of what came to be known as the laboratory *TA Manuals*. These documents are a collection of structured course notes that the TAs may use to conduct each three hour laboratory session. In the past it was each TA's responsibility to produce course notes. The TA manual allows for the information to be uniformly distributed among all the TAs. Also, the notes are revised and enhanced at the end of each semester to address any comments or suggestions made by the TAs using the notes. The TA manual is very much a "living" document.

A.3.2.3 Design of Deliverables for Each Experiment

The deliverables for each laboratory remain somewhat constant throughout the three courses but the expectations rise between the sophomore, junior and senior level laboratory. The general deliverable for each laboratory is a written lab report that introduces or sets up the activity, describes the procedure and analysis techniques, presents results, discusses results and finally draws conclusions from those results. These reports are prepared in small groups.

While a report is expected for each experiment the TAs must be realistic with their expectations. Reports prepared in the sophomore level laboratory will generally contain less breadth and depth than those prepared in the junior level laboratory and even more so when compared to the senior level laboratory.

The laboratory reports require the students to present a complete discussion of the subject matter in much the same way a technical journal paper is expected to present scientific findings. The report should contain all necessary background and theory the reader needs to follow the report. The report should explain the outcome in terms of the presented theoretical background. Finally, the report should address the question that was asked at the beginning of the lab. Every effort should be made for these report to be stand alone documents and readable by anyone with an undergraduate engineering degree.

The level of complexity in the laboratory reports is primarily due to the students' academic maturity but also the manner in which the laboratories are presented. There is a noticeable, decreasing level of guidance present as the student ascend through the laboratories. The sophomore level laboratory is focused on *observation* or *exploration* and is delivered with a relatively high amount of guidance. This is meant to demonstrate proper procedure and analysis techniques to the students.

As the students progress from the sophomore level to the junior level laboratory they are given less procedural guidance while being held to higher standards regarding the final deliverable: the laboratory report. The junior level laboratory is meant to reinforce *understanding* and offers students the opportunity for a more in-depth investigation laboratory content.

The senior level laboratory lab is concerned with *explanation and forecasting.* Students are given the least amount of direct guidance and are held to the highest standards in terms of the completeness and thoroughness of their reports. As the students move through the three laboratories they will answer (in this order), "How things behave," "Why they behave that way," and "What does this behavior mean for the investigator?" It is a gradual approach to nurturing inquisitive students.

A.3.2.4 Updating Laboratory Systems and Experiment Evolution

Updated laboratory equipment allows the students to gain hands on experience with both hardware and software prior to graduation. A sampling of the laboratory equipment includes Programmable Logic Controllers (PLCs) hardware and software, LabVIEW® data acquisition hardware and software, HAAS® computer numerically controlled machines and associated software as well as MATLAB™ and Simulink® software packages for simulation and post-processing of data.

In addition to providing hands on experience with industry standard hardware and software, an effort was made to phase-in new experiments. The new experiments focus on both fundamental engineering principles as well as modern implementation and were introduced at a rate of no more than one per semester. This gradual evolution represents, in part, the laboratory's continual improvement efforts.

A.4 Assessment Strategy and Results

End-of-semester surveys were administered to the enrolled students in each laboratory course and analyzed to gauge the effectiveness of the implemented changes. The surveys were administered through Blackboard® and the responses were anonymous. The authors feel that the responses to the survey questions indicate a realization of the desired outcomes listed in Table A.2.

A.4.1 Assessment Surveys

The student survey is an effective means of collecting diagnostic information [155–157]. Surveys were administered at the end of each semester to determine how successful the ME 222, ME 333 and ME 444 courses were at achieving their goals.

The surveys ask a total of eight questions regarding four different aspects (categories) of the course: report writing, software, statistics, and design of experiments. The first question in each group is answered with either extensive coverage, moderate coverage or minimal coverage and the second question is answered with either strongly agree, agree, disagree or strongly disagree. The questions are as follows:

Q1) Report Writing: Writing effective reports concerning experimental procedures and results.

Q2) "My report writing skills and ability to discuss results and draw conclusions have been improved."

Q3) Software: Use of software to acquire, analyze, and present data.

Q4) "My skills in the use of software for data analysis, plotting and presentation have been improved by experiences in this course."

Q5) Statistics: Application of statistics in the analysis of engineering data.

Q6) "I have increased my knowledge of statistics with engineering applications including uncertainty analysis."

Q7) Design of Experiments: Design of test procedures, selection of instruments, randomization, and calibration.

Q8) I have increased my knowledge and experience in designing and conducting experiments."

A.4.2 Discussion of Responses to Survey Questions

The assessment results in Figures A.2 through A.9 will be discussed with recognition that each category contained two questions that inquired about the coverage level and the student's self-measured improvement. The first category explores report writing. In Figure A.2, the extensive and moderate coverage levels per question Q1 have exceeded 90% for the Fall 2008 through Spring 2011 time periods which reflects an acceptable amount of writing coverage. The students reported improvements in report writing in question Q2 at the strongly agree and agree levels generally exceeding the 80% threshold per Figure A.3. The 80% threshold represents the department of mechanical engineering's acceptance level for laboratory assessment questions Q2, Q4, Q6, and Q8. It should be noted that some fluctuations exist semester-by-semester which may be attributed to a particular teaching assistant and their strengths with guiding laboratory report writing activities.

The undergraduate laboratory courses maintain a strong emphasis on computer-based data acquisition (currently National Instruments™) so that students can integrate sensors with digital hardware to collect test data. In addition, the undergraduate program requires the use of simulation tools (typically MATLAB™ or Simulink®) for homework and project assignments in the various courses. Consequently, questions Q3 and Q4 have been answered by students with a smaller percentage of responses to "extensive coverage" and "strongly agree." However, the number of individuals rating the coverage as extensive and moderate continually exceeds 85% in Figure A.4. Similarly, the 80% threshold in Figure A.5 has been satisfied in four of the six semesters with the other two semesters missing the target by a maximum of 3%. Overall, the student observations on software usage and improvement in the laboratory are acceptable.

The statistical and uncertainty analysis efforts in the laboratory as measured by student assessment measures Q5 and Q6 have historically been low. In Figures A.6 and A.7, the results displayed for the Fall 2008 through Fall 2010 semesters were not acceptable but reflect the positive trend of real improvements in the courses with respect to this category. An important aspect of the department of mechanical engineering continual assessment process is the recognition of deficiencies and the commencement of corrective actions to resolve and improve the situation. For this particular case, the Spring 2011 semester truly represents the first instance in which the coverage (extensive and moderate) has exceeded 90% and the corresponding improvement in statistics has passed 80% by a wide margin. Although students are required to complete an undergraduate statistics course, they do not necessarily understand the opportunity to put theory into practice in the laboratory courses without significant guidance. The laboratory team believes that the continual emphasis on statistics and uncertainty analysis throughout the experimental assignments has produced the deliverables stated for the three courses.

The final assessment category concerns the design of experiments. As shown in Figure A.8, the percentage of students who rate the coverage level as extensive and/or moderate has continually increased over the past five semesters. In the most recent assessment period, the coverage was 97% per Q7 which clearly demonstrates that the students responded favorably to those laboratory tasks which required experimental design activities. Finally, the last question, Q8, inquires whether students improved their knowledge and experience in designing and conducting experiments. In Figure A.9, the 80% threshold was exceeded in Fall 2009 and has been maintained at a satisfactory assessment level for the past three semesters. Overall, the students demonstrate an ability to design, conduct, and report on the results of their experimental investigations which complement the theoretical underpinnings from their traditional lecture classes.

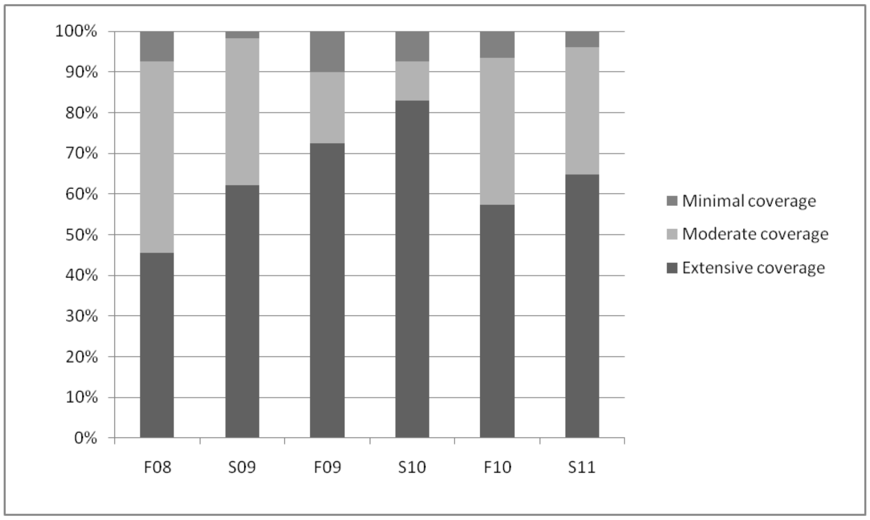

Figure A.2: Q1: Regarding report writing: "Writing effective reports concerning experimental procedures and results."

A.4.3 Analysis of Laboratory Outcomes

Table A.6 offers a concise summery of the information displayed in Figures A.2-A.9. The table quantifies by what percent positive survey responses increased over the six semester time period. For those questions where the possible responses were extensive coverage, moderate coverage, or minimal coverage; extensive coverage was considered a "positive response." For those questions where the possible responses were strongly agree, agree, disagree, or strongly disagree; strongly agree and agree were considered "positive response." As can be observed in Table A.6 every one of the survey questions experienced an increase in positive responses over the course of the study. The smallest survey question increase was 11% while the largest exceeded 47%.

The first intended outcome was to help students develop a more in-depth understanding of the lecture and laboratory material. Responses to Q1 of the survey

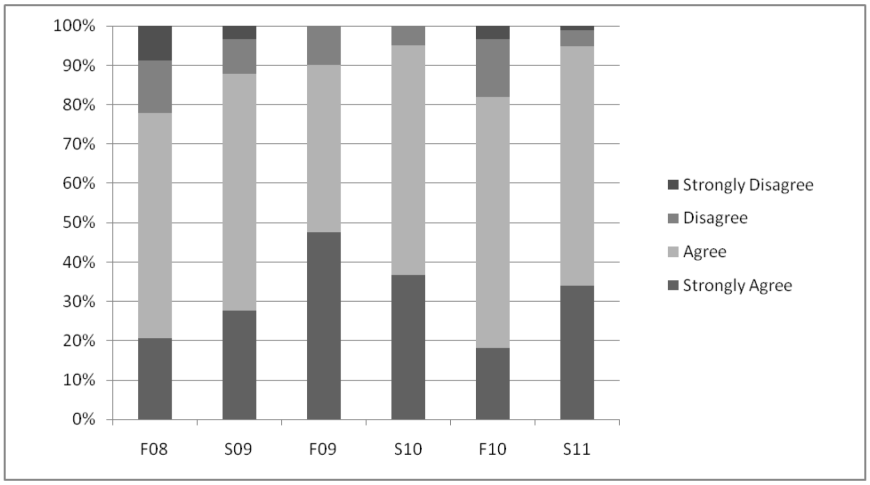

Figure A.3: Q2: Regarding report writing: "My report writing skills and ability to discuss results and draw conclusions have been improved."

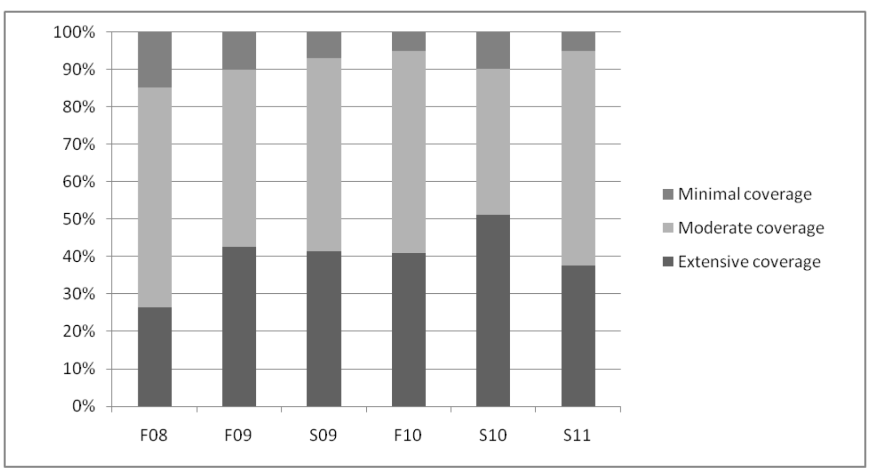

Figure A.4: Q3: Regarding software: "Use of software to acquire, analyze, and present data."

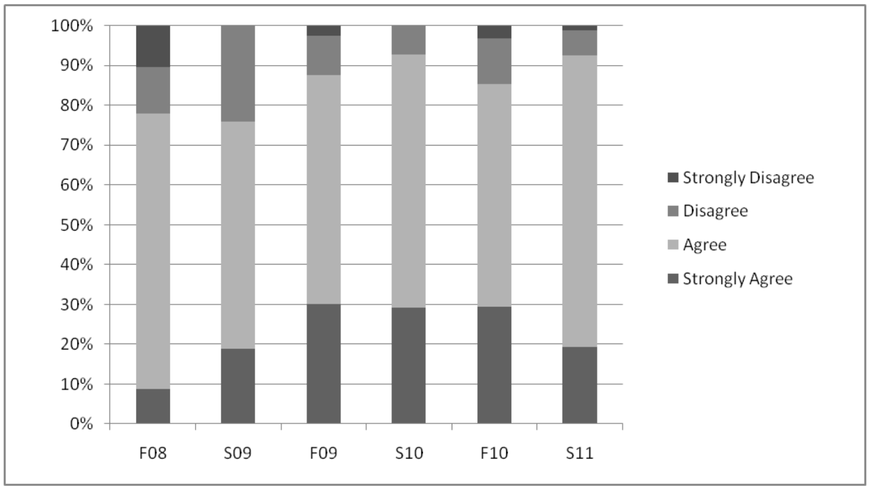

Figure A.5: Q4: Regarding software: “My skills in the use of software for data analysis, plotting and presentation have been improved by experiences in this course.”

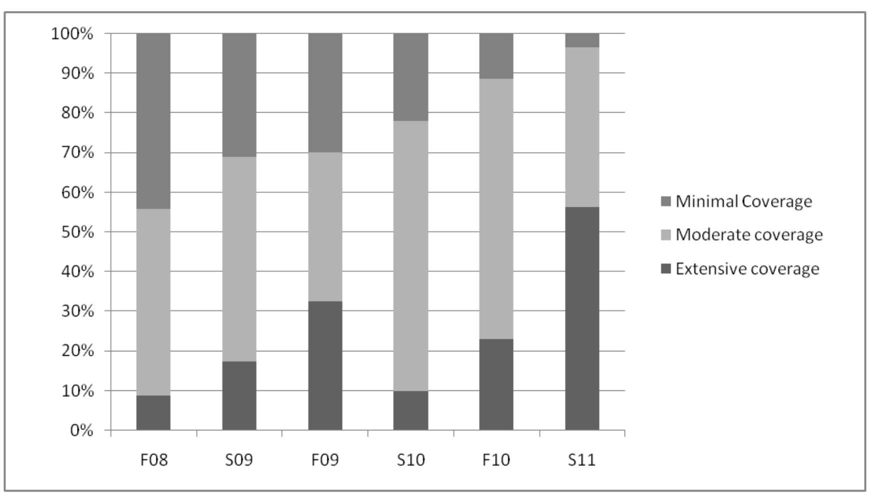

Figure A.6: Q5: Regarding statistics: “Application of statistics in the analysis of engineering data.”

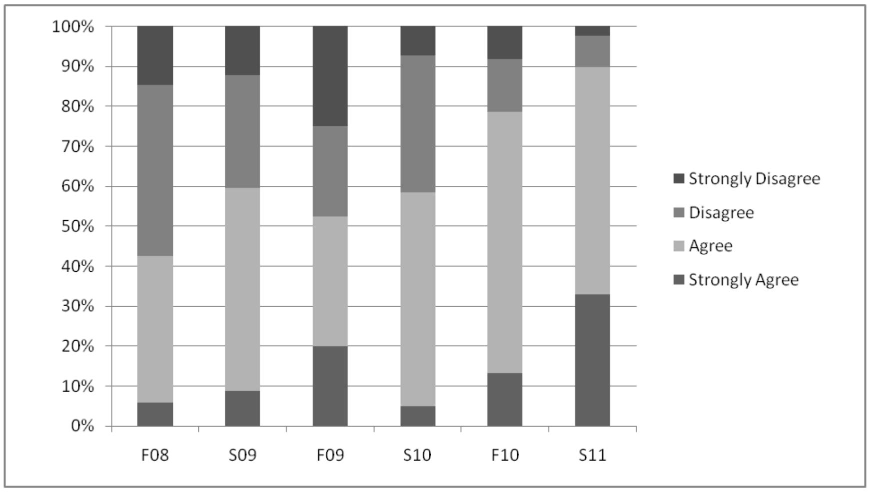

Figure A.7: Q6: Regarding statistics: "I have increased my knowledge of statistics with engineering applications including uncertainty analysis."

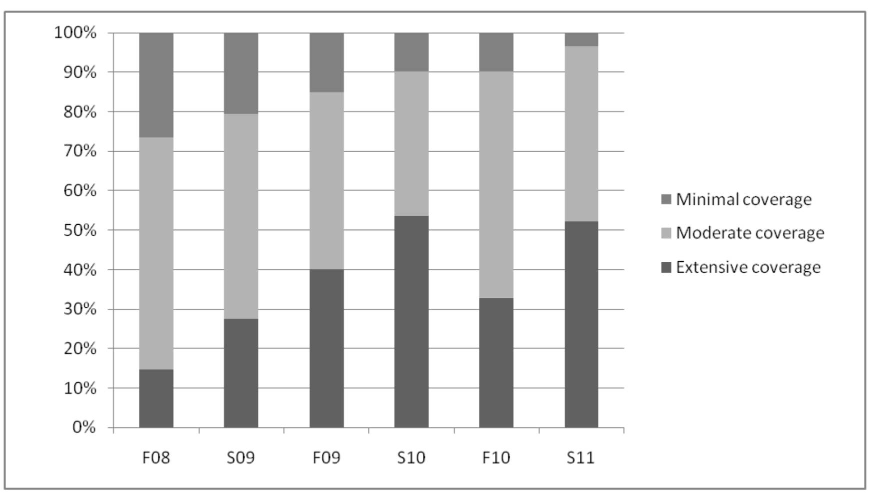

Figure A.8: Q7: Regarding design of experiments: "Design of test procedures, selection of instruments, randomization, and calibration."

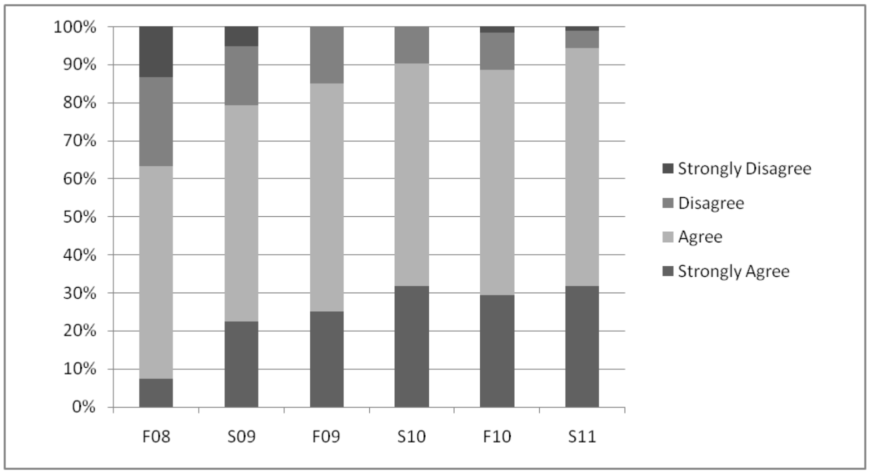

Figure A.9: Q8: Regarding design of experiments: "I have increased my knowledge and experience in designing and conducting experiments."

indicate that the students feel they are getting more extensive coverage of effective report writing. Furthermore, the increase in positive responses to Q2 implies that the extra effort spent in covering report writing has been effective in improving the students writing skills.

Exposing students to modern engineering tools (both hardware and software) was the second intended outcome. The changes made to the laboratories help increase the percentage of positive responses to Q3 and Q4 which directly address part of that particular outcome. Students reported more extensive coverage of software used to acquire, analyze, and present data while at the same time indicating that their software skills have been improved throughout the course. Again, an increase in positive responses to both Q3 and Q4 is more convincing than an increase of either question by itself.

The third intended outcome, improved communication skills, is most directly addressed by Q1 and Q2. Students reported experiencing more extensive coverage of report writing skills (Q1) as well as believing that their report writing skills have been improved (Q2). Response to these two questions were particularly important because well developed writing skills are essential to success both during and after college. While many engineers seem to loath writing the truth is that, "...college graduates spend an average of 20 to 30 per cent [sic] of their time in the workplace on writing tasks, and even more as they advance through the ranks." (p. 151. [158])

Understanding the limitations of experimental work was the fourth intended outcome of the changes made to the laboratories. We believe that the introduction of statistics and uncertainty material into the laboratories helped in accomplishing this goal and that the impact of these actions can me measured with the responses to Q5 and Q6. Again, we had students report that the application of statistics in the analysis of their data was more extensive by the end of the study and that they have increased their knowledge of statistics with engineering applications including uncertainty analysis.

Lastly, we wanted to improve the student's ability to design and conduct experiments. This intended outcome was most directly addressed by Q7 and Q8. By the end of the study students were reporting that they had more extensive coverage in designing of test procedures, selection of instruments, randomization, and calibration. Also, students report that they have increased their knowledge and experience in designing and conducting experiments.

A.5 Conclusion

The undergraduate laboratory structure in the mechanical engineering department at Clemson University is atypical in that the laboratories are offered in stand alone courses as opposed to being integrated into corresponding lecture courses. However, this structure allows for a small number of faculty members (laboratory coordinator, faculty advisory) and a single graduate student (laboratory development TA)

Table A.6: Percent change of positive survey responses from the beginning to the end of the study based on Figures A.2-A.9.

Question	Percentage of positive responses, Fall 2008	Percentage of positive responses, Spring 2011	Overall Result
Q1	45.6%	64.8%	↑ 19.2%
Q2	77.9%	94.9%	↑ 16.9%
Q3	26.5%	37.5%	↑ 11.0%
Q4	77.9%	92.6%	↑ 14.7%
Q5	8.8%	56.3%	↑ 47.4%
Q6	42.6%	89.8%	↑ 47.1%
Q7	14.7%	52.3%	↑ 37.6%
Q8	63.2%	94.3%	↑ 31.1%

to oversee and work with all aspects of all the laboratories.

The undergraduate laboratories have been documented and discussed in this chapter. The redesigned laboratory courses provided a coordinated sequence of experiences for the students that gradually prepares them to be competent engineers and investigators. The laboratories start off by exposing the students to engineering/physical phenomena, then asks them to explain the phenomena and finally asks the students how to use this information. The implemented changes have resulted in achieving all of the desired outcomes evidenced by positive survey results. For those institutions struggling with laboratory delivery, a methodical approach to experiments has yielded great results.

Bibliography

[1] R. Isermann, *Fault-Diagnosis Systems: An Introduction from Fault Detection to Fault Tolerance*. Germany: Springer, 2006.

[2] J. J. Gertler, *Fault Detection and Diagnosis in Engineering Systems*. New York: Marcel Dekker, Inc., 1998.

[3] P. J. Brockwell and R. A. Davis, *Time Series: Theory and Methods*, 2nd ed. New York: Springer-Verlag, 1991.

[4] U.S. Department of Energy, "20% Wind Energy by 2030 Increasing Wind Energys Contribution to U.S. Electricity Supply Executive Summary," U.S. Department of Energy, Tech. Rep. NREL TP-500-42864; DOE/GO-102008-2578, December 2008.

[5] ——, "20% Wind Energy by 2030 Increasing Wind Energys Contribution to U.S. Electricity Supply," U.S. Department of Energy, Tech. Rep. NREL TP-500-41869; DOE/GO-102008-2567, July 2008.

[6] C. McGowin, "Condition Monitoring of Wind Turbines: Technology Overview, Seeded-Fault Testing, and Cost-Benefit Analysis," EPRI, Palo Alto, CA, Tech. Rep. 1010419, 2006.

[7] J. Manwell, *Wind Energy Explained: Theory, Design and Application*. New York: John Wiley & Sons, Inc, 2002.

[8] T. Burton, D. Sharpe, N. Jenkins, and E. Bossanyi, *Wind Energy Handbook*. West Sussex, England: Wiley, 2011.

[9] P. Gipe, *Wind Power: Renewable Energy for Home, Farm, and Business*. Vermont: Chelsea Green Publishing, 2003.

[10] I. Woofenden, *Wind Power for Dummies*. New Jersey: Wiley Pub, 2009.

[11] Z. Hameed, Y. Hong, Y. Cho, S. Ahn, and C. Song, "Condition monitoring and fault detection of wind turbines and related algorithms: A review," *Renewable and Sustainable Energy Reviews*, vol. 13, no. 1, pp. 1–39, 2009.

[12] Y. Amirat, M. Benbouzid, B. Bensaker, and R. Wamkeue, "Condition monitoring and fault diagnosis in wind energy conversion systems: A review," in *Proceedings of IEEE International Electric Machines and Drives Conference, IEMDC 2007*, vol. 2, Antalya, Turkey, 2007, pp. 1434–1439.

[13] X. Zhang, S. he, P. Zho, and W. Wang, "Summerization [sic] and study of fault diagnosis technology of the main components of wind turbine generator system," in *Sustainable Energy Technologies. ICSET 2008. IEEE International Conference on*, Piscataway, NJ, 2008, pp. 1223–6.

[14] M. Twerdochlib, R. C. Miller, T. A. Lemak, R. L. Osborne, C. C. Repplier, D. R. Morrison, and D. J. Boomgaard, "Two recent developments in monitors for large turbine generators." *IEEE Transactions on Energy Conversion*, vol. 3, no. 3, pp. 653–659, 1988.

[15] E. Echavarria, B. Hahn, G. J. W. van Bussel, and T. Tomiyama, "Reliability of wind turbine technology through time," *Journal of Solar Energy Engineering*, vol. 130, no. 3, pp. 031 005–1–031 005–8, 2008.

[16] M. Mohsin Khan, M. Tariq Iqbal, and F. Khan, "Reliability and condition monitoring of a wind turbine," in *Canadian Conference on Electrical and Computer Engineering*, Saskatoon, Canada, 2005, pp. 1978–1981.

[17] E. Wiggelinkhuizen, T. Verbruggen, H. Braam, L. Rademakers, J. Xiang, and S. Watson, "Assessment of condition monitoring techniques for offshore wind farms," *Journal of Solar Energy Engineering*, vol. 130, no. 3, pp. 031 004–1–031 004–9, 2008.

[18] J. Ribrant and L. M. Bertling, "Survey of failures in wind power systems with focus on swedish wind power plants during 1997-2005," *IEEE Transactions on Energy Conversion*, vol. 22, no. 1, pp. 167–173, March 2007.

[19] L. Popa, B. Jensen, E. Ritchie, and I. Boldea, "Condition monitoring of wind generators," in *Conference Record of the Industry Applications Conference, 2003. 38th IAS Annual Meeting*, vol. 3, Salt Lake City, UT, October 2003, pp. 1839–1846.

[20] D. McMillan and G. W. Ault, "Quantification of condition monitoring benefit for offshore wind turbines," *Wind Engineering*, vol. 31, no. 4, pp. 267–285, 2007.

[21] W. Yang, P. Tavner, and M. Wilkinson, "Wind turbine condition monitoring and fault diagnosis using both mechanical and electrical signatures," in *IEEE/ASME International Conference on Advanced Intelligent Mechatronics. AIM 2008*, Xi'an, China, 2008, pp. 1296–1301.

[22] E. Becker and P. Poste, "Keeping the blades turning. Condition monitoring of wind turbine gears," *Refocus*, vol. 7, no. 2, pp. 26–28, 2006.

[23] M. R. Wilkinson, F. Spinato, and P. J. Tavner, "Condition monitoring of generators and other subassemblies in wind turbine drive trains," in *2007 IEEE International Symposium on Diagnostics for Electric Machines, Power Electronics and Drives, SDEMPED*, Cracow, Poland, 2007, pp. 388–392.

[24] E. Echavarria, T. Tomiyama, and G. van Bussel, "Fault diagnosis approach based on a model-based reasoner and a functional designer for a wind turbine. an approach towards self-maintenance," *Journal of Physics: Conference Series*, vol. 75, no. 1, pp. 012 078–1–012 078–9, 2007.

[25] M. Artioli, G. Capolino, F. Filippetti, and A. Yazidi, "A general purpose software for distance monitoring and diagnosis of electrical machines," in *SDEMPED 2003 - Symposium on Diagnostics for Electric Machines, Power Electronics and Drives*, Atlanta, GA, August 2003, pp. 272–276.

[26] M. C. Garcia, M. A. Sanz-Bobi, and J. del Pico, "SIMAP: Intelligent system for predictive maintenance. Application to the health condition monitoring of a windturbine gearbox," *Computers in Industry*, vol. 57, no. 6, pp. 552–568, 2006.

[27] H. Camblong, M. Vidal, and J. Puiggali, "Principles of a simulation model for a variable-speed pitch-regulated wind turbine," *Wind Engineering*, vol. 28, no. 2, pp. 157–175, 2004.

[28] J. Carta and D. Mentado, "A continuous bivariate model for wind power density and wind turbine energy output estimations," *Energy Conversion and Management*, vol. 48, no. 2, pp. 420–32, 2007.

[29] P. Delarue, A. Bouscayrol, A. Tounzi, X. Guillaud, and G. Lancigu, "Modelling, control and simulation of an overall wind energy conversion system," *Renewable Energy*, vol. 28, no. 8, pp. 1169–1185, 2003.

[30] C. Eisenhut, F. Krug, C. Schram, and B. Klockl, "Wind-turbine model for system simulations near cut-in wind speed," *IEEE Transactions on Energy Conversion*, vol. 22, no. 2, pp. 414–420, 2007.

[31] J. Ekanayake, L. Holdsworth, X. Wu, and N. Jenkins, "Dynamic modeling of doubly fed induction generator wind turbines," *IEEE Transactions on Power Systems*, vol. 18, no. 2, pp. 803–809, 2003.

[32] C. Jauch and T. Cronin, "Simulation model of a wind turbine pitch controller for grid frequency stabilisation," *Wind Engineering*, vol. 29, no. 4, pp. 377–387, 2005.

[33] C. Jauch, P. Serensen, and B. Jensen, "Simulation model of a transient fault controller for an active-stall wind turbine," *Wind Engineering*, vol. 29, no. 1, pp. 33–47, 2005.

[34] C. Leclerc and C. Masson, "Wind turbine performance predictions using a differential actuator-lifting disk model," *Transactions of the ASME Journal of Solar Energy Engineering*, vol. 127, no. 2, pp. 200–208, 2005.

[35] M. Mansouri, M. Mimouni, B. Benghanem, and M. Annabi, "Simulation model for wind turbine with asynchronous generator interconnected to the electric network," *Renewable Energy*, vol. 29, no. 3, pp. 421–431, 2004.

[36] S. Santoso and H. T. Le, "Fundamental time-domain wind turbine models for wind power studies," *Renewable Energy*, vol. 32, no. 14, pp. 2436–2452, 2007.

[37] A. Yazidi, H. Henao, G. Capolino, D. Casadei, F. Filippetti, and C. Rossi, "Simulation of a doubly-fed induction machine for wind turbine generator fault analysis," in *SDEMPED 2005 - International Symposium on Diagnostics of Electric Machines, Power Electronics, and Drives*, Vienna, Spain, 2005, pp. 1–6.

[38] E. Muljadi, C. Butterfield, H. Romanowitz, and R. Yinger, "Self-excitation and harmonics in wind power generation," *Transactions of the ASME Journal of Solar Energy Engineering*, vol. 127, no. 4, pp. 581–587, 2005.

[39] A. Baumgart, "A mathematical model for wind turbine blades," *Journal of Sound and Vibration*, vol. 251, no. 1, pp. 1–12, 2002.

[40] E. G. Harokopos and R. W. Mayne, "Motor characteristics in the control of a compliant load," *Journal of Guidance, Control, and Dynamics*, vol. 9, no. 1, pp. 113–118, 1986.

[41] A. Hansen, P. Sorensen, F. Blaabjerg, and J. Becho, "Dynamic modelling of wind farm grid interaction," *Wind Engineering*, vol. 26, no. 4, pp. 191–210, 2002.

[42] G. M. McNerney and R. D. Richardson, "The statistical smoothing of power delivered to utilities by multiple wind turbines," *IEEE Transactions on Energy Conversion*, vol. 7, no. 4, pp. 644–647, 1992.

[43] A. Shafiu, O. Anaya-Lara, G. Bathurst, and N. Jenkins, "Aggregated wind turbine models for power system dynamic studies," *Wind Engineering*, vol. 30, no. 3, pp. 171–186, 2006.

[44] C. McGowin, "Tennessee Valley Authority Buffalo Mountain Wind Power Project First- and Second-Year Operating Experience: 2001-2003: U.S. Department of Energy-EPRI Wind Turbine Verification Program," EPRI, Palo Alto, CA, U.S. Department of Energy, Washington DC, and Tennessee Valley Authority, Chattanooga, TN, Tech. Rep. 1008031, 2003.

[45] ——, "U.S. Department of Energy/Electric Power Research Institute (EPRI) Utility Wind Turbine Verification Program," EPRI, EPRI, Palo Alto, CA, U.S. Department of Energy, Washington, D.C., Tech. Rep. 1014488, 2008.

[46] J. Nilsson and L. Bertling, "Maintenance management of wind power systems using condition monitoring systems-life cycle cost analysis for two case studies," *IEEE Transactions on Energy Conversion*, vol. 22, no. 1, pp. 223–229, 2007.

[47] W. Jeffries, J. Chambers, and D. Infield, "Experience with bicoherence of electrical power for condition monitoring of wind turbine blades," *IEE Proceedings-Vision, Image and Signal Processing*, vol. 145, no. 3, pp. 141–148, 1998.

[48] P. C. Sen, *Principles of Electric Machines and Power Electronics*, 2nd ed. New York: John Wiley & Sons, Inc., 1997.

[49] S. J. Chapman, *Electric Machinery and Power System Fundamentals.* New York: McGraw-Hill, Inc., 2002.

[50] A. R. Daniels, *Introduction to Electrical Machines.* Australia: The Macmillan Press Ltd., 1976.

[51] J. Dron, L. Rasolofondraibe, C. Couet, and A. Pavan, "Fault detection and monitoring of a ball bearing benchtest and a production machine via autoregressive spectrum analysis," *Journal of Sound and Vibration*, vol. 218, no. 3, pp. 501–525, 1998.

[52] J. Stack, T. Habetler, and R. Harley, "Bearing fault detection via autoregressive stator current modeling," *IEEE Transactions on Industry Applications*, vol. 40, no. 3, pp. 740–747, May-June 2004.

[53] B. Yazici and G. Kliman, "An adaptive statistical time-frequency method for detection of broken bars and bearing faults in motors using stator current," *Industry Applications, IEEE Transactions on*, vol. 35, no. 2, pp. 442–452, March-April 1999.

[54] M. Benbouzid, "Bibliography on induction motors faults detection and diagnosis," *IEEE Transactions on Energy Conversion*, vol. 14, no. 4, pp. 1065–1074, December 1999.

[55] M. E. H. Benbouzid, “A review of induction motors signature analysis as a medium for faults detection,” *Industrial Electronics, IEEE Transactions on*, vol. 47, no. 5, pp. 984–993, October 2000.

[56] R. Tallam, S. B. Lee, G. Stone, G. Kliman, J. Yoo, T. Habetler, and R. Harley, “A survey of methods for detection of stator-related faults in induction machines,” *Industry Applications, IEEE Transactions on*, vol. 43, no. 4, pp. 920–933, July-August 2007.

[57] D. Casadei, F. Filippetti, A. Stefani, C. Rossi, A. Yazidi, and G. Capolino, “Experimental fault characterization of doubly fed induction machines for wind power generation,” in *SPEEDAM 2006 - International Symposium on Power Electronics, Electrical Drives, Automation and Motion, 2006*, Taormina, Italy, 2006, pp. 1281–1286.

[58] F. Filippetti, G. Franceschini, C. Tassoni, and P. Vas, “Recent developments of induction motor drives fault diagnosis using AI techniques,” *Industrial Electronics, IEEE Transactions on*, vol. 47, no. 5, pp. 994–1004, October 2000.

[59] K. Kim and A. Parlos, “Induction motor fault diagnosis based on neuropredictors and wavelet signal processing,” *Mechatronics, IEEE/ASME Transactions on*, vol. 7, no. 2, pp. 201–219, June 2002.

[60] S.-H. Lee, S. Kim, J. M. Kim, C. Choi, J. Kim, S. Lee, and Y. Oh, “Extraction of induction motor fault characteristics in frequency domain and fuzzy entropy,” in *Electric Machines and Drives. IEEE International Conference on*, San Antonio, TX, May 2005, pp. 35–40.

[61] A. Parlos, K. Kim, and R. Bharadwaj, “Detection of induction motor faults-combining signal-based and model-based techniques,” in *American Control Conference. Proceedings of the 2002*, vol. 6, Anchorage, AK, 2002, pp. 4531–4536.

[62] A. Stefani, A. Yazidi, C. Rossi, F. Filippetti, D. Casadei, and G.-A. Capolino, “Doubly fed induction machines diagnosis based on signature analysis of rotor modulating signals,” *IEEE Transactions on Industry Applications*, vol. 44, no. 6, pp. 1711–1721, 2008.

[63] S. Gupta, A. Ray, S. Sarkar, and M. Yasar, “Fault detection and isolation in aircraft gas turbine engines. part 1: underlying concept,” *Proceedings of the Institution of Mechanical Engineers, Part G: Journal of Aerospace Engineering*, vol. 222, no. 3, pp. 307–318, 2008.

[64] M. J. Roemer, C. S. Byington, and G. J. Kacprzynski, "An overview of selected prognostic technologies with application to engine health management," in *Proceedings of GT2006 ASME Turbo Expo 2006: Power for Land, Sea, and Air*, no. GT2004-54135, Barcelona, Spain, May 2006, pp. 1–9.

[65] M. J. Roemer and C. S. Byington, "Prognostics and health management software for gas turbine engine bearings," in *Proceedings of GT2007 ASME Turbo Expo 2007: Power for Land, Sea, and Air*, no. 4790X. Montreal, Canada: ASME, May 2007, pp. 795–802.

[66] P. Vecer, M. Kreidl, and R. Smid, "Condition indicators for gearbox condition monitoring systems," *Acta Polytechnica, Czech Technical University in Prague*, vol. 45, no. 6, pp. 35–43, 2005.

[67] A. Mohanty and C. Kar, "Fault detection in a multistage gearbox by demodulation of motor current waveform," *Industrial Electronics, IEEE Transactions on*, vol. 53, no. 4, pp. 1285–1297, June 2006.

[68] M. J. Watson, C. S. Byington, and A. Behbahani, "Very high frequency monitoring system for engine gearbox and generator health management," in *Proceedings of the SAE AeroTech Congress & Exhibition*, no. 2007-01-3878, Los Angeles, CA, September 2007, pp. 1–10.

[69] S. E. Oraby, A. F. Al-Modhuf, and D. R. Hayhurst, "A diagnostic approach for turning tool based on the dynamic force signals," *Journal of Manufacturing Science and Engineering*, vol. 127, no. 3, pp. 463–475, 2005.

[70] S. Thanagasundram, S. Spurgeon, and F. S. Schlindwein, "A fault detection tool using analysis from an autoregressive model pole trajectory," *Journal of Sound and Vibration*, vol. 317, no. 3-5, pp. 975–993, 2008.

[71] R. Sekhon, H. Bassily, and J. Wagner, "A comparison of two trending strategies for gas turbine performance prediction," *Journal of Engineering for Gas Turbines and Power*, vol. 130, no. 4, pp. 041 601–1–041 601–10, July 2008.

[72] M. Watson, C. Byington, D. Edwards, and S. Amin, "Dynamic modeling and wear-based remaining useful life prediction of high power clutch systems," *STLE Tribology Transactions*, vol. 48, no. 2, pp. 208–217, April-June 2005.

[73] B. R. Upadhyaya, M. Kitamura, and T. W. Kerlin, "Signature monitoring of nuclear power plant dynamics - stochastic modeling and case studies," in *Decision and Control including the Symposium on Adaptive Processes. 19th IEEE Conference on*, vol. 19, Albuquerque, NM, December 1980, pp. 121–126.

[74] R. Patton, P. Frank, and R. Clark, Eds., *Fault Diagnosis in Dynamic Systems: Theory and Application.* New Jersey: Prentice-Hall, Inc., 1989.

[75] J. Gertler, "Survey of model-based failure detection and isolation in complex plants," *IEEE Control Systems Magazine*, vol. 8, no. 6, pp. 3–11, December 1988.

[76] R. J. Patton, P. M. Frank, and R. N. Clark, Eds., *Issues of Fault Diagnosis for Dynamic Systems*. London: Springer, 2000.

[77] C. Byington, M. Watson, D. Edwards, and P. Stoelting, "A model-based approach to prognostics and health management for flight control actuators," in *IEEE Aerospace Conference Proceedings*, vol. 6, Big Sky, MT, March 2004, pp. 3551–3562.

[78] G. Vachtsevanos, F. L. Lewis, M. Roemer, A. Hess, and B. Wu, *Intelligent Fault Diagnosis and Prognosis for Engineering Systems*. New Jersey: John Wiley & Sons, Inc., 2006.

[79] R. K. Mobley, *An Introduction to Predictive Maintenance*, 2nd ed. New York: Butterworth-Heinemann, 2002.

[80] P. Bloomfield, *Fourier Analysis of Time Series*. New York: Wiley, 1976.

[81] G. M. Jenkins and D. G. Watts, *Spectral Analysis and Its Applications*, G. M. Jenkins and E. Parzen, Eds. California: Holden-Day, 1968.

[82] R. A. Fisher, "Tests of significance in harmonic analysis," *Proceedings of the Royal Society of London. Series A, Containing Papers of a Mathematical and Physical Character*, vol. 125, no. 796, pp. 54–59, 1929. [Online]. Available: http://www.jstor.org/stable/95247

[83] R. Schalkoff, *Pattern Recognition: Statistical, Structural and Neural Approaches*. New York: John Wiley & Sons, Inc., 1992.

[84] R. H. Shumway, "Time-frequency clustering and discriminant analysis," *Statistics & Probability Letters*, vol. 63, no. 3, pp. 307–314, July 2003.

[85] R. Lund and B. Li, "Revisiting climate region definitions via clustering," *Journal of Climate*, vol. 22, no. 7, pp. 1787–800, 2009.

[86] H. Huang, H. Ombao, and D. S. Stoffer, "Discrimination and classification of nonstationary time series using the slex model," *Journal of the American Statistical Association*, vol. 99, no. 467, pp. 763–774, September 2004.

[87] P. J. Brockwell and R. A. Davis, *Introduction to Time Series and Forecasting*, 2nd ed. New York: Springer, 2002.

[88] D. R. Brillinger, *Time Series Data Analysis and Theory*, expanded ed. California: Holden-Day, Inc., 1981.

[89] R. H. Shumway and D. S. Stoffer, *Time Series Analysis and Its Applications With R Examples*, 2nd ed. New York: Springer, 2006.

[90] J. W. Tukey, "Discussion, emphasizing the connection between analysis of variance and spectrum analysis," *Technometrics*, vol. 3, no. 2, pp. 191–219, 1961. [Online]. Available: http://www.jstor.org/stable/1266112

[91] P. Welch, "The use of fast fourier transform for the estimation of power spectra: A method based on time averaging over short, modified periodograms," *Audio and Electroacoustics, IEEE Transactions on*, vol. 15, no. 2, pp. 70–73, June 1967.

[92] J. Caiado, N. Crato, and D. Pena, "A periodogram-based metric for time series classification," *Computational Statistics & Data Analysis*, vol. 50, no. 10, pp. 2668–2684, 2006.

[93] J. Caiado, N. Crato, and D. Peña, "An interpolated periodogram-based metric for comparison of time series with unequal lengths," 2006, published in the Munich Personal RePEc Archive. MPRA Paper No. 2075, posted 07. November 2007 / 02:12. [Online]. Available: http://mpra.ub.uni-muenchen.de/2075/

[94] H. Hotelling, "The generalization of student's ratio," *Annals of Mathematical Statistics*, vol. 2, no. 3, pp. 360–378, 1931.

[95] ——, "Multivariate quality control," in *Techniques of Statistical Analysis*, 1st ed., C. Eisenhart, M. W. Hastay, and W. A. Wallis, Eds. New York, NY: McGraw-Hill Book Company, Inc., 1947.

[96] J. E. Jackson, "Multivariate quality control," *Communications in Statistics - Theory and Methods*, vol. 14, no. 11, pp. 2657–2688, 1985.

[97] J. A. Mauricio, "Computing and using residuals in time series models," *Computational Statistics & Data Analysis*, vol. 52, no. 3, pp. 1746–1763, 2008.

[98] D. S. Coates and P. J. Diggle, "Tests for comparing two estimated spectral densities," *Journal of Time Series Analysis*, vol. 7, no. 1, pp. 7–20, 1986.

[99] R. Lund, H. Bassily, and B. Vidakovic, "Testing equality of stationary autocovariances," *Journal of Time Series Analysis*, vol. 30, no. 3, pp. 332–348, 2009.

[100] Y. Kakizawa, R. H. Shumway, and M. Taniguchi, "Discrimination and clustering for multivariate time series," *Journal of the American Statistical Association*, vol. 93, no. 441, pp. 328–340, March 1998. [Online]. Available: http://www.jstor.org/stable/2669629

[101] H. F. Bassily, "A comparative fault diagnosis methodology based on time series analysis of system's signals," Ph.D. dissertation, Clemson University, 2007. [Online]. Available: http://etd.lib.clemson.edu/documents/1202498718/

[102] H. Bassily, R. Lund, and J. Wagner, "Fault detection in multivariate signals with applications to gas turbines," *IEEE Transactions on Signal Processing*, vol. 57, no. 3, pp. 835–42, 2009.

[103] I. Jolliffe, *Principal Component Analysis*, 2nd ed. New York: Springer, 2002.

[104] C. Zhang, G. He, and S. Liang, "Pca-based analog fault detection by combining features of time domain and spectrum," in *Intelligent Systems and Applications, 2009. ISA 2009. International Workshop on*, Wuhan, China, May 2009, pp. 1–4.

[105] J. E. Jackson and G. S. Mudholkar, "Control procedures for residuals associated with principal component analysis," *Technometrics*, vol. 21, no. 3, pp. 341–349, August 1979.

[106] B. Moore, "Principal component analysis in linear systems: controllability, observability, and model reduction," *IEEE Transactions on Automatic Control*, vol. AC-26, no. 1, pp. 17–32, 1981.

[107] W. Ku, R. H. Storer, and C. Georgakis, "Disturbance detection and isolation by dynamic principal component analysis," *Chemometrics and Intelligent Laboratory Systems*, vol. 30, no. 1, pp. 179–196, 1995.

[108] L. Ljung, *System Identification*, 2nd ed. New Jersey: Prentice Hall PTR, 1999.

[109] A. Gut, *An Intermediate Course in Probability.* New York: Springer-Verlag, 1995.

[110] S. Ross, *A First Course in Probability*, 7th ed. New Jersey: Pearson Prentice Hall, 2006.

[111] S. M. Ross, *Introduction to Probability Models*, 9th ed. New York: Academic Press, 2007.

[112] T. W. Anderson, *An Introduction to Multivariate Statistical Analysis*, 2nd ed. New York: Wiley, 1984.

[113] G. H. Golub and C. F. Van Loan, *Matrix Computations.* Maryland: Johns Hopkins University Press, 1983.

[114] C. E. Shannon and W. Weaver, *The Mathematical Theory of Communication.* Illinois: The University of Illinois Press: Urbana, 1949.

[115] S. Kullback and R. A. Leibler, "On information and sufficiency," *The Annals of Mathematical Statistics*, vol. 22, no. 1, pp. 79–86, 1951.

[116] H. Chernoff, "A measure of asymptotic efficiency for tests of a hypothesis based on the sum of observations," *The Annals of Mathematical Statistics*, vol. 23, no. 4, pp. 493–507, December 1952.

[117] A. Rényi, "On measures of entropy and information," in *Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability*, J. Neyman, Ed., Statistical Laboratory of the University of California, Berkeley, June-July 1960, pp. 547–561. [Online]. Available: http://projecteuclid.org/euclid.bsmsp/1200512181

[118] N. Ye, *The Handbook of Data Mining*. New Jersey: Lawrence Erlbaum Assoc, 2003.

[119] *NIST/SEMATECH e-Handbook of Statistical Methods*, 2010. [Online]. Available: http://www.itl.nist.gov/div898/handbook/

[120] A. C. Davison, *Statistical Models*. Cambridge, UK: Cambridge University Press, 2003.

[121] G. Casella and R. Berger, *Statistical Inference*, 2nd ed. California: Duxbury Press, 2001.

[122] V. Monbet, P. Ailliot, and M. Prevosto, "Survey of stochastic models for wind and sea-state time series," *Probabilistic Engineering Mechanics*, vol. 22, pp. 113–126, 2007.

[123] P. J. Diggle and N. I. Fisher, "Nonparametric comparison of cumulative periodograms," *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, vol. 40, no. 3, pp. 423–434, 1991. [Online]. Available: http://www.jstor.org/stable/2347522

[124] G. M. Jenkins, "General considerations in the analysis of spectra," *Technometrics*, vol. 3, no. 2, pp. 133–166, May 1961. [Online]. Available: http://www.jstor.org/stable/1266110

[125] R. H. Shumway and A. N. Unger, "Linear discriminant functions for stationary time series," *Journal of the American Statistical Association*, vol. 69, no. 348, pp. 948–956, December 1974. [Online]. Available: http://www.jstor.org/stable/2286169

[126] P. De Souza and P. Thomson, "Lpc distance measures and statistical tests with particular reference to the likelihood ratio," *Acoustics, Speech and Signal Processing, IEEE Transactions on*, vol. 30, no. 2, pp. 304–315, April 1982.

[127] E. Maharaj, "Cluster of time series," *Journal of Classification*, vol. 17, pp. 297–314, 2000.

[128] D. Piccolo, "A distance measure for classifying arima models," *Journal of Time Series Analysis*, vol. 11, no. 2, pp. 153–164, 1990.

[129] H. Tong and P. Dabas, "Cluster of time series models: An example," *Journal of Applied Statistics*, vol. 17, no. 2, pp. 187–198, 1990.

[130] J. W. H. Swanepoel and J. W. J. Van Wyk, "The comparision of two spectral density functions using the bootstrap," *Journal of Statistical Computation and Simulation*, vol. 24, no. 3-4, pp. 271–282, 1986.

[131] C. T. Shaw and G. P. King, "Using cluster analysis to classify time series," *Physica D: Nonlinear Phenomena*, vol. 58, no. 1-4, pp. 288–298, 1992.

[132] T. W. Liao, "Clustering of time series data - a survey," *Pattern Recognition*, vol. 38, pp. 1857–1874, 2005.

[133] E. Keogh and S. Kasetty, "On the need for time series data mining benchmarks: A survey and empirical demonstration," *Data Mining and Knowledge Discovery*, vol. 7, no. 4, pp. 349–371, October 2003, special Issue: Selected Papers from the Eighth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining - Part II.

[134] E. Keogh and J. Lin, "Clustering of time series subsequences is meaningless: Implications for previous and future research," *Knowledge and Information Systems*, vol. 8, no. 2, pp. 154–177, August 2005.

[135] E. A. Maharaj, "Comparison of non-stationary time series in the frequency domain," *Computational Statistics & Data Analysis*, vol. 40, no. 1, pp. 131–141, 2002.

[136] ——, "Comparison and classification of stationary multivariate time series," *Pattern Recognition*, vol. 32, no. 7, pp. 1129–1138, 1999.

[137] R. Lund, H. Hurd, P. Bloomfield, and R. Smith, "Climatological time series with periodic correlation," *J. Climate*, vol. 8, no. 11, pp. 2787–2809, November 1995.

[138] A. Ben-Israel and T. N. Greville, *Generalized Inverses: Theory and Applications.* New York: John Wiley & Sons, 1973.

[139] W. Navidi, *Statistics for engineers and scientists.* Boston: McGraw-Hill Higher Education, 2008.

[140] S. Bivona, G. Bonanno, R. Burlon, D. Gurrera, and C. Leone, "Stochastic models for wind speed time series: A case study," *Acta Physica Polonica B*, vol. 41, no. 5, pp. 1083–1092, May 2010.

[141] P. Chen, T. Pedersen, B. Bak-Jensen, and Z. Chen, "Arima-based time series model of stochastic wind power generation," *Power Systems, IEEE Transactions on*, vol. 25, no. 2, pp. 667–676, May 2010.

[142] B. T. Ewing, J. B. Kruse, J. L. Schroeder, and D. A. Smith, "Time series analysis of wind speed using VAR and the generalized impulse response technique," *Journal of Wind Engineering and Industrial Aerodynamics*, vol. 95, no. 3, pp. 209 –219, March 2007.

[143] J. F. Manwell, J. G. McGowan, and A. L. Rogers, *Wind Energy Explained: Theory, Design and Application*, 2nd ed. Great Britain, UK: John Wiley & Sons, Inc, 2009.

[144] T. Roppel, J. Hung, S. Wentworth, and A. Hodel, "An interdisciplinary laboratory sequence in electrical and computer engineering: curriculum design and assessment results," *IEEE Transactions on Education*, vol. 43, no. 2, pp. 143–152, May 2000.

[145] Y. Amigud, G. Archer, J. Smith, M. Szymanski, and B. Servatius, "Assessing the quality of web-enabled laboratories in undergraduate education," in *32nd ASEE/IEEE Frontiers in Education Conference*, Boston, MA, November 2002, pp. 3–12.

[146] L. Gomes and S. Bogosyan, "Current trends in remote laboratories," *IEEE Transactions on Industrial Electronics*, vol. 56, no. 12, pp. 4744–4756, December 2009.

[147] E. D. Lindsay and M. C. Good, "Effects of laboratory access modes upon learning outcomes," *IEEE Transactions on Education*, vol. 48, no. 4, pp. 619–631, November 2005.

[148] J. L. Watson, G. Bibel, K. Ebeling, J. Erjavec, H. Salehfar, and M. Zahui, "On-line laboratories for undergraduate distance engineering students," in *34th ASEE/IEEE Frontiers in Education Conference*, Savannah, GA, October 2004, pp. T3C–1–T3C–6.

[149] H. Wu, Y. Yang, W. Qingquan, and S. Zhu, "An internet-based control engineering laboratory for undergraduate and graduate education," in *2006 IEEE International Conference on Automation Science and Engineering, CASE*, Shanghai, China, October 2007, pp. 546–550.

[150] R. S. Heller, C. Beil, K. Dam, and B. Haerum, "Student and faculty perceptions of engagement in engineering," *Journal of Engineering Education*, vol. 99, no. 3, pp. 253–261, July 2010.

[151] M. Borrego, J. E. Froyd, and T. S. Hall, "Diffusion of engineering education innovations: A survey of awareness and adoption rates in U.S. engineering departments," *Journal of Engineering Education*, vol. 99, no. 3, pp. 185–207, July 2010.

[152] A. Saavedra Montes, H. Botero Castro, and J. Hernandez Riveros, "How to motivate students to work in the laboratory: A new approach for an electrical machines laboratory," *IEEE Transactions on Education*, vol. 53, no. 3, pp. 490–496, August 2010.

[153] J. Allison and F. Benson, "Undergraduate projects and their assessment," *Physical Science, Measurement and Instrumentation, Management and Education - Reviews, IEE Proceedings A*, vol. 130, no. 8, pp. 402–419, November 1983.

[154] G. Carter, D. Armour, L. Lee, and R. Sharples, "Assessment of undergraduate electrical engineering laboratory studies," *Physical Science, Measurement and Instrumentation, Management and Education - Reviews, IEE Proceedings A*, vol. 127, no. 7, pp. 460–474, September 1980.

[155] M. Perez, P. Herrero, F. Sanchez, and V. Robles, "Are web self-assessment tools useful for training?" *IEEE Transactions on Education*, vol. 48, no. 4, pp. 757–763, November 2005.

[156] J. Weaver and S. Das, "Overhaul of an undergraduate mechanical measurements laboratory," in *29th ASEE/IEEE Frontiers in Education Conference*, San Juan, Puerto Rico, November 1999, pp. 10–16.

[157] S. W. Rogers and R. K. Goktas, "Exploring engineering graduate student research proficiency with student surveys," *Journal of Engineering Education*, vol. 99, no. 3, pp. 263–278, July 2010.

[158] L. Nilson, *Teaching at Its Best*, 2nd ed. Massachusetts: Anker Pub. Co, 2003.