

APPLICATIONS OF MULTITAPER SPECTRAL ANALYSIS
TO NONSTATIONARY DATA

by

KARIM JOHN RAHIM

A thesis submitted to the
Graduate Program in Statistics
in conformity with the requirements for
the degree of Doctor of Philosophy

Queen's University
Kingston, Ontario, Canada

October, 2014

Copyright Karim John Rahim, 2014

Abstract

This thesis is concerned with changes in the spectrum over time observed in Holocene climate data as recorded in the Burgundy grape harvest date series. These changes represent nonstationarities, and while spectral estimation techniques are relatively robust in the presence of nonstationarity—that is, they are able to detect significant contributions to power at a given frequency in cases where the contribution to power at that given frequency is not constant over time—estimation and prediction can be improved by considering nonstationarity. We propose improving spectral estimation by considering such changes. Specifically, we propose estimating the level of change in frequency over time, detecting change-point(s) and sectioning the time series into stationary segments. We focus on locating a change in frequency domain in time, and propose a graphical technique to detect spectral changes over time. We test the estimation technique in simulation, and then apply it to the Burgundy grape harvest date series. The Burgundy grape harvest date series was selected to demonstrate the introduced estimator and methodology because the time series is equally spaced, has few missing values, and a multitaper spectral analysis, which the methodology proposed in this thesis is based on, of the grape harvest date series was recently published. In addition, we propose a method using a test for goodness-of-fit of autoregressive estimators to aid in assessment of change in spectral properties over time.

This thesis has four components: (1) introduction and study of a level-of-change estimator for use in the frequency domain change-point detection, (2) spectral analysis of the Burgundy grape harvest date series, (3) goodness-of-fit estimates for autoregressive processes, and (4) introduction of a statistical software package for multitaper spectral analysis. We present four results. (1) We introduce and demonstrate the feasibility of a level-of-change estimator. (2) We present a spectral analysis and coherence study of the Burgundy grape harvest date series that includes locating a change-point. (3) We present a study showing an advantage using multitaper spectral estimates when calculating autocorrelation coefficients. And (4) we introduce an R software package, available on the Comprehensive R Archive Network (CRAN), to perform multitaper spectral estimation.

Acknowledgments

I would like to thank my advisor, David Thomson, for sharing his knowledge and interest, and, perhaps most importantly, for his kindness, insight, encouragement, and honesty in working with me in this endeavour. I would like to thank the following past and current students of David Thomson who have provided helpful discussion along the way: Wesley Burr, Charlotte Haley, Kyle Lepage, Ian Moore, Joshua Pohlkamp-Hartt, David Riegert, and Aaron Springford. In addition, I would also like to thank Maja-Lisa Thomson, Valdimar Tasnov, and Jim Diamond for helpful discussions, I would like to thank Jennifer Reid for making this department a comfortable place to be. Finally yet importantly, I would like to thank my family for their patience, kindness, and encouragement along the way.

Co-authorship

Chapters 3, 4, and 5 are co-authored with David J. Thomson. Appendix A1, which discusses the multitaper R software package, is co-authored with Wesley S. Burr and David J. Thomson. The multitaper R software package is co-authored with Wesley S. Burr and David J. Thomson.

Table of Contents

Abstract	ii
Acknowledgments	iv
Co-authorship	v
Table of Contents	vi
List of Tables	x
List of Figures	xiii
List of Abbreviations	xxv
List of R Package Function Calls	xxvii
Chapter 1:	
Introduction	1
Chapter 2:	
Background	3
2.1 Time Series Analysis	4
2.2 Stochastic Processes	4

2.3	Stationary Time Series	6
2.4	Several Definitions	8
2.5	Spectral Density Function	10
2.6	Spectral Estimation	10
2.7	Spectral Representation of a Stationary Process	15
2.8	Nonstationary Harmonizable Process	19
2.9	Multitaper Spectral Estimation Overview	20
2.10	Aliasing	25
2.11	Zero Padding Spectral Estimates	26
2.12	Jackknife Estimates	27
2.13	Coherence	28
2.14	Spectrograms	30

Chapter 3:

	Frequency-domain Change-point Detection	32
3.1	Introduction	33
3.2	Change-points Problem Overview	34
3.3	Literature Review of Change-point Techniques	36
3.4	Additional Preliminaries	39
3.5	Level-of-change in Frequency-domain	41
3.6	Simulation Study of Estimator	46
3.7	Suggested Methodology	73
3.8	Summary and Comments	74

Chapter 4:

Burgundy Grape Harvest Dates	76
4.1 Introduction	77
4.2 Initial Analysis	78
4.3 Spectrograms and Level-of-change	93
4.4 Summary and Concluding Remarks	98

Chapter 5:

Goodness-of-fit in AR Processes	99
5.1 Introduction	100
5.2 Calculation of AR Coefficients	102
5.3 Cautionary Notes on Using AR Spectral Estimates	113
5.4 Comparison of Methods for Finding AR Coefficients	114
5.5 Goodness-of-fit Test for Autoregressive Processes	115
5.6 Simulations of Goodness-of-fit	119
5.7 Burgundy Grape Harvest Dates	120
5.8 Conclusions and Future Work	122

Chapter 6:

Concluding Remarks	126
-------------------------------------	------------

Bibliography	129
-------------------------------	------------

Appendix A:

Multitaper R Package	149
A.1 Appendix Overview	150

A.2	Introduction	150
A.3	The Theory of Multitaper Spectral Estimation	154
A.4	Addressing Statistical Significance with Multitaper Tools	162
A.5	Bivariate Time Series: Magnitude-squared Coherence	168
A.6	Complex Demodulation	172
A.7	Additional Tools and Extending Functionality	177
A.8	Summary	179

List of Tables

3.1	Random samples of spectral means were generated of size 2048, 4096, 8192, then multitaper adaptively weighted block spectrograms were constructed by using block lengths of 128, 256, and 512 respectively. The table gives sample means found using simulation for the periodogram and nonadaptive weighted multitaper spectral estimates with time-bandwidth parameters, $NW = 2, 3, 4$ and 5	50
3.2	Variances of random samples were generated and multitaper spectrograms constructed as in Table 3.2. Observed sample variances constructed using adaptive weighting are higher than both theoretical variances and simulated variances constructed from multitaper spectrograms without adaptive weighting.	50
3.3	Sample means of the level-of-change estimator from an $N(0, 1)^3$ distribution. 4000-run simulations were made, each having 16 blocks in length. The bottom row gives the approximations derived in Section 3.5.2.	52
3.4	Sample variances from simulated level-of-change estimator from an $N(0, 1)^3$ distribution. 4000-run simulations were made, each having 16 blocks in length. The bottom row gives the approximations derived in Section 3.5.2.	53

3.5	Average across blocks and frequencies of the standard error matrix of the level-of-change estimator, using adaptive weights, with $NW = 5$, and $K = 9$, from 4000 simulations of the autoregressive moving average (ARMA)(4,2) process.	62
3.6	Average across blocks and frequencies of the standard sample mean of the level-of-change estimator, using adaptive weights, $NW = 5$ and $K = 9$, from 4000 simulations of the ARMA(4,2) process.	63
3.7	Cutoffs for controlling Type I error for the level-of-change estimator based on maximum values in each level-of-change matrix for 4000 simulation and a $N(0, 1)$ process.	70
3.8	A sample of potential block sizes, selected by using the criterion that data at the end points not be discarded. In general, when the offset size is small, the options for block size increase, and the trade-off occurs when block size and offset are close and thus minimizing the overlap.	72
5.1	Comparisons of estimates of $\phi_{4,4}$ from 100,000 run simulations using the Yule-Walker equations with the biased autocovariance estimator, an autocovariance estimator using one Slepian taper with $NW = 5$, an adaptive weighted multitaper spectral estimate with $NW = 5$, and $k = 8$, and the partial autocovariance estimator made using Burg's method.	115

5.2 Shape and rate parameters with their respective standard errors, abbreviated SE, for the fitted Gamma distributions shown in Figure 5.2. Both the shape and rate parameters are considerably higher for the case where the simulated autoregressive (AR) model did not match the theoretical model. 121

5.3 Maximum absolute deviation (max abs dist) of the observed grape harvest date (GHD) standardized integrated spectrum to the theoretical standardized integrated spectrum for the various models and approximate p-values based on simulations testing the null hypothesis that the maximum absolute deviation is small enough for the model to be appropriate. 124

List of Figures

3.1	Multitaper spectrogram plot with adaptive weighting of white noise data using 16 non-overlapped blocks of length 128—that is, the total length is $N = 16 \times 128 = 2048$. The multitaper spectral estimates use the parameters $NW = 5$ and $K = 9$	48
3.2	Level-of-change estimator between block pairs based on the spectrogram of white noise shown in Figure 3.1, using multitaper parameters $NW = 5$, and $K = 9$. The frequency range is reduced as we omit frequencies within W of the zeroth and Nyquist (0.5). In this example, we use a cutoff value of 4.16, giving a 5% error rate for the complete matrix, and this matrix exhibits no change-points.	49
3.3	Multitaper spectrogram of a realization of the stationary AR(2) process. Each block is 128 samples long, and the multitaper parameters used are $NW = 5$ and $K = 9$. The spectral estimate in each block is not well resolved with 128 sample block sizes.	55
3.4	Level-of-change estimator for the AR(2) example shown in Figure 3.3. This realization indicates the potential for a false detect. This risk can be reduced by recognizing a higher likelihood of false detect around the unresolved peak in the spectrum.	56

3.5	Multitaper adaptively weighted spectrogram of a realization of the ARMA(4,2) process. Each block is 128 samples long, and the multitaper parameters used are $NW = 5$ and $K = 9$	58
3.6	Multitaper adaptively weighted spectrum estimate all 2048 samples from the same realization of the ARMA(4,2) process using multitaper parameters $NW = 5$ and $K = 9$	59
3.7	Level-of-change estimator, constructed without adaptive weights, $NW = 5$, and $K = 9$, for the ARMA(4,2) example shown in Figure 3.5. This plot has h	60
3.8	Level-of-change estimator, constructed using adaptive weights, with $NW = 5$ and $K = 9$ for the ARMA(4,2) example shown in Figure 3.5. This image is less noisy than the one without adaptive weights. This plot has some high valued false detects which require further examination. 61	61
3.9	Level-of-change estimator plot showing only values above the 4.16 cut-off constructed using adaptive weights, with $NW = 5$, and $K = 9$ for the ARMA(4,2) example shown in Figure 3.5. The only detected values are in a region where false detects are expected due to the low resolution of each block.	62

3.10 Multitaper spectrogram plot of simulated data containing two sinusoidal frequencies, with one that considerably damps down at the halfway point. In this case the nonstationarity is clearly visible in the spectrogram. The black line segment in the upper left indicates the bandwidth, $2W$. The first half of the data has a sinusoid of amplitude $A_{1a} = 1$ at $f_{1a} = .09$, and a sinusoid of amplitude $A_2 = 0.6$ at $f_2 = 0.2$. The second half has a sinusoid of amplitude $A_{1b} = 0.2$ at $f_{1b} \approx 0.0526$. The background noise has constant variance of one. The multitaper parameters used were $NW = 5$, and $K = 9$. The ≈ 0.0526 low-amplitude frequency is not distinguishable at this block length. 65

3.11 We plot the level-of-change estimator between adjacent blocks, trimming the blocks by w at the frequency edges (zero and Nyquist frequencies). Note that we visually detect a level-of-change estimator between blocks 8 and 9 at a frequency of approximately 0.091 (1/11). 66

3.12 Bartlett M-test for this change-point example. This test shows nonstationarity at the frequency where there is a change in amplitude and change in frequency. The line segment in the below the legend indicates the bandwidth, $2W$, and the two dashed lines indicate the chi-squared expected value and the 95% value. The multitaper parameters used were $NW = 5$, and $K = 9$ 68

3.13	Average eighth block pair level-of-change column over 4000 simulations. This figure shows that the average observed level-of-change over the 4000 simulations is considerably higher in the frequency range where the change-point occurs. The multitaper parameters used were $NW = 5$ and $K = 9$	69
3.14	Plots of densities of the level-of-change estimator for a model with a change-point and a model without. These are based on 4000 simulations comparing maximum values of a model with a change-point to one without. The intersection point is 0.68.	71
4.1	(a) Burgundy GHD plotted as number of days after September 1 st . Five additional series are also shown: (b) Swiss GHD as days after September 1 st . There are several large gaps in the first part of this series. (c) Central England Temperature (CET) annual temperature series. (d) Annual phase of the CET series in (angular) degrees. (e) Estimated total solar irradiance (TSI) in watts per square metre. (f) Three reconstructions of the El Niño—southern oscillation (ENSO) cycle shown in normalized degrees Celsius.	79
4.2	Multitaper spectra of GHD series. Multitaper spectral estimates were made with $NW = 3, 4, 5$ and 6 , and with $K = 5, 7, 9$ and 10 , respectively starting at the top left. The crosses at approximately 0.135 cycles/year indicates the passband bandwidth, $2W$, and height of the approximate theoretical 95% confidence interval based on the χ^2_{2k} distribution. Note that the peak at a period of 3.9 years almost agrees with Tourre et al. (2011).	81

4.3	This figure shows the harmonic F -test statistic for the harvest dates. The parameter values used are $NW = 3, 4, 5$ and 6 , with $K = 5, 7, 9$ and 10 respectively. The red dashed line indicates a $1 - 1/N$ level of significance where $N = 634$, in keeping with the rule of thumb for the harmonic F -test (see Section 3.6.1). We note that the most significant peak occurs at a period of 4.14 years, which is close to the reported period of 3.9 years reported in Tourre et al. (2011, p. 247).	82
4.4	Overlapping section of the Swiss and Burgundy GHD series consisting of years 1550 to 2003; no prewhitening has been applied, and magnitude-squared coherence (MSC) is presented in the next plot. We note that the Swiss harvest is on average ~ 14 days after the Burgundy harvest.	84
4.5	MSC between Swiss and Burgundy GHDs. The coherence is constructed from overlapped years 1550 to 2003 and is based on the multi-taper spectral estimates with parameters $NW = 4$ and $K = 7$. The y-axis indicates a normalized MSC; a hyperbolic inverse tangent transform is known to transform the MSC to a standard normal distribution (Thomson and Chave, 1991b). The dashed red line indicates the inherent bias in the estimate; specifically, it shows that a coherence of 0.14 will be observed for estimated values of uncorrelated samples. The faint dashed line on the coherence plot represents the lower of a one standard deviation jackknife confidence interval. The two dashed blue lines indicate a significance of 95% and 99% , corresponding to an MSC of 0.39 and 0.54 respectively.	85

4.6	Phase coherence between Burgundy and Swiss GHDs. Coherence is defined in (2.67) and based on the multitaper cross-spectrum in (2.68). In these equations the Burgundy series is represented by x and the Swiss series is represented by y . Two standard deviation confidence intervals are indicated on the plots; the green line represents multitaper jackknife confidence intervals, and the blue line represents approximate theoretical confidence intervals (Bendat and Piersol, 2011, p. 306). It may be observed that these agree well. The phase is generally consistent with zero, excluding the low-frequency part, and no phase unwrapping was required. Between periods of ~ 208 and ~ 90 years there is a sharp drop to -69 degrees. Both edge frequencies are well known in the climate literature: 208 years is one of the main “Suess cycles” (Thomson, 1990b), and 90 years is very close to the upper peak, 91.5 years, of the ~ 88 year Gleissberg cycle triplet (Peristyk and Damon, 2003). The linear regression line (in grey) has a negative intercept and a positive slope. This indicates that the Swiss series leads the Burgundy series by ~ 9 days.	88
4.7	Plots of the Burgundy GHD and the CET annual series for overlapping years 1661 to 2003.	89

4.8	MSC between Central England average annual temperature and the Burgundy harvest dates from 1661 to 2003. The parameters used are $NW = 6.5$ $K = 11$. The dashed red line indicates the bias value of 0.09, and the dashed blue lines indicate MSC of 0.173 and 0.201. The coherence is modest, particularly at low frequencies. The association between GHD and April to August temperatures in Burgundy have been established (Chuine et al., 2004; Krieger et al., 2011).	89
4.9	Phase coherence between the Burgundy GHD and the average annual temperature of Central England series and for years 1661 to 2003. This figure is based on (2.67) with the Burgundy series is represented by x and the Central England series is represented by y . The multitaper parameters are: $NW = 6.5$, $K = 11$. The linear regression line (in red) has a positive intercept and a positive slope. This indicates that the Burgundy series leads the Central England series by ~ 18 days. .	90
4.10	Plot of the Burgundy GHD Series and the Central England phase constructed from three years of monthly data. The phase was first corrected for the three day offset. A discussion of obtaining the phase plot is given in Appendix A.6.1.	90

4.11 MSC between annual phase of the Central England temperature series and Burgundy GHD for years 1661 to 2003. The multitaper parameters are: $NW = 6.5$, $K = 11$. The coherence is modest at low frequencies. The annual phase of the Central England temperature series was calculated with zeroth order Slepian complex demodulation technique with a length of $N = 36$, 3 years of monthly data, with $NW = 4.5$. The three-day offset for years 1661 to 1752, originally reported in Thomson (1995), discussed on page 174, was applied. The dashed red line indicates the bias value of 0.091, and the dashed blue lines indicate a MSC of 0.17 and 0.21. 91

4.12 Phase coherence between the Burgundy GHD and annual phase of the Central England temperature series calculated over three years. This figure is based on (2.67) with the Burgundy series is represented by x and the annual phase of the Central England temperature series represented by y . The intercept is positive and the slope is ~ 300 degrees per year indicating the Burgundy GHD leads phase of the CET series by ~ 305 days. The multitaper parameters are: $NW = 6.5$, $K = 11$ 92

4.13 Multitaper spectrogram with considerable overlap. In this case the block length is 74, there are 71 blocks, and the offset is 8 years. This indicates an overlap of about 89%, but it allows for higher-frequency resolution. The vertical line segment on the left indicates the bandwidth, $2W$, and one can see the spectral estimates evolve over time. The centre line indicates where are analysis selects to section the series. 94

4.14	Bartlett M-test for stationarity using block sizes with 2.5% (little) overlap. The expected value (green dashed line) and the 95% significance level (red dotted line) are on the graph. The multitaper parameters used were $NW = 3$ $K = 5$, with 8 blocks, each of length 81 with an offset of 79. The line segment in the top right of the plot indicates the bandwidth. Nonstationary components are approximately between the frequencies of 0.1 and 0.18 cycles/year, and between 0.2 and 0.24 cycles/year	95
4.15	We plot the level-of-change between blocks in the spectrogram for the GHD. If we restrict ourselves to the frequency of interest, 0.10 to 0.18, based on the Bartlett M-test, we see that considerable change occurs at approximately the centre of the series.	96
4.16	Multitaper spectra of the GHD before (top) and after (bottom) the year 1675.5. The crosses indicate 95% confidence levels and the width of bandwidth parameter, $2W$. On the upper plot, the dashed lines indicate a period of 10.6 years (0.94 cycles/year), and 7.5 years (0.133 cycles/year)for the date up to the year 1675. On the lower plot, the dashed line indicates a period of 3.9 years (0.278 cycles/year). It appears that a change in the spectral properties of the GHD series occurs when the data is sectioned at the year 1675.	97

5.1	<p>Estimated fourth-reflection coefficient based on a 100000-run simulation of an AR(4) process with coefficients 2.7607, -3.8106, 2.6535, -0.9238. Levinson-Durbin estimate using: (a) the default estimate—i.e., using the autocovariance sequence (acvs) from unwindowed Fourier transforms; (b) one discrete prolate spheroidal sequence (DPSS) taper with $NW = 5$; and (c) the use of an adaptive multitaper estimate with $k = 8$. The dashed line indicates -0.9238, the true value. Mean estimates were -0.425, -0.914, and -0.920 respectively. The distribution of the Burg estimator is very similar to the multitaper spectral estimator and is not shown.</p>	116
5.2	<p>This figure shows the observed maximum absolute distance observed from 40000 simulations. The top left plot compares a simulated AR(4) to the theoretical AR(4), the top right plot compares a simulated AR(2) to the theoretical AR(2), the bottom left plot compares a simulated AR(4) to the theoretical AR(2), and the bottom right plot compares a simulated AR(2) to a theoretical AR(4). Note the changing y-axis scales.</p>	120

5.3	We ran 40000 simulations each comparing a simulated AR(4) to the theoretical AR(4), top left, a simulated AR(2) to the theoretical AR(2), top right, a simulated AR(2) to the theoretical AR(2), bottom left, and a simulated AR(4) to the theoretical AR(2), bottom right. The top two plots indicate the <i>worst fit</i> of the 40000 runs when the simulations were from the same model as the theoretical AR, and the bottom two plots indicate the <i>best fit</i> of the 40000 runs when the simulations are from a model than different from the theoretical AR.	122
5.4	Adaptive multitaper spectrum of the GHD series. The parameters used are: $NW = 3$ and $k = 5$. Plotted over the spectrum, we have the standard AR(1) spectrum in red, the standard AR(8) spectrum in green, the DPSS tapered AR(8) spectrum in blue, and the multitaper AR(8) spectrum in cyan. The multitaper AR(8) in cyan and the standard AR(8) follow closely except between the frequencies 0.2 and 0.3 (cycles/year), where the multitaper estimate has slightly higher power and appears to follow the spectral estimate more closely.	123
A.1	Adaptive MTM of AR(4) time series	160
A.2	First six years of the CET daily series.	166
A.3	Spectrum of CET series, zoomed to region around 1 cycle/year ($31.69n\text{Hz} = 31.69 \times 10^{-9} \text{ Hz}$), with 95% jackknifed confidence intervals.	180
A.4	Harmonic F -test statistic for the CET series, zoomed to low frequencies using the function <code>dropFreqs</code>	180
A.5	CO ₂ concentration time series in parts-per-billion with trend lines fitted.	181

A.6	Temperature deviations time series in degrees Celsius with trend lines fitted.	181
A.7	MSC between monthly CO ₂ measurements from Mauna Loa, and the global temperature series during 1958–2007. The Arctanh transform normalizes the MSC and each integer value on this scale represents approximately one standard deviation (Thomson and Chave, 1991b).	182
A.8	Central England monthly temperature Phase	183

List of Abbreviations

acvf autocovariance function

acvs autocovariance sequence

AIC Akaike information criterion

AR autoregressive

ARMA autoregressive moving average

CDF cumulative distribution function

CET Central England Temperature

CRAN Comprehensive R Archive Network

DFT discrete Fourier transform

DPSS discrete prolate spheroidal sequence

dpswf discrete prolate spheroidal wave function

ENSO El Niño—southern oscillation

FFT fast Fourier transform

GHD grape harvest date

GMM generalized method-of-moments

MA moving average

MSC magnitude-squared coherence

MSE mean-squared-error

SDF spectral density function

TSI total solar irradiance

WOSA Welch's (windowed) overlapped segment average

List of R Package Function Calls

A.1	spec.mtm	161
A.2	spec.mtm (Ftest and Jackknife)	167
A.3	dropFreqs	167
A.4	plot.mtm (Jackknife Confidence Intervals)	167
A.5	plot.mtm (Ftest)	168
A.6	multitaperTrend	171
A.7	mtm.coh	172
A.8	plot.mtm.coh	172
A.9	demod.dpss	176

Chapter 1

Introduction

We present an application of multitaper spectral analysis to nonstationarity climate data, and we present a novel analysis of the Burgundy grape harvest date (GHD) series which includes a change-point detection. Spectral analysis is robust to structural nonstationarity; specifically, spectral analysis detects harmonic components that are present for a fraction of the time series. However robust, spectral analysis may be, we propose to improve harmonic analysis by considering change in spectral properties over time. We focus on changes to the dynamic spectrum—i.e., changes in the spectrum over time—and present a methodology for assessing the level-of-change in the spectrum and locating change-points using multitaper spectral estimates.

This thesis is structured as follows: Chapter 2 contains a description of the multitaper spectral estimator and a general literature review. In Chapter 3, a spectral methodology for examining level-of-change and locating change-points in time series is introduced. This chapter includes an overview of the current time series change-point literature, a description of the introduced level-of-change estimator, a discussion of its statistical properties, an assessment using simulations, and a methodology using

this estimator, in conjunction with other existing tools. Chapter 4 is dedicated to the analysis of the Burgundy GHD series introduced by Chuine et al. (2004), which is used as a proxy measure for Western European climate fluctuations (Tourre et al., 2011). This chapter establishes a coherence between the GHD series and other climate proxy measures and uses the methodology introduced in Chapter 3 to detect a change-point that is consistent with known climate fluctuations. In Chapter 5, we consider a test for goodness-of-fit for autoregressive processes and test the fits of several autoregressive processes to the GHD time series. This chapter is based on a paper submitted to conference proceedings at the 2013 Joint Statistical Meeting (Rahim and Thomson, 2013). In Chapter 6 we present some concluding remarks and discuss possible future work. Finally, we include a paper on an R statistical software package that implements the multitaper method. We submitted this paper to the *Journal of Statistical Software* in July 2012, and it was recommended for publication pending a major revision. The major revision included changes to the paper, the software and the software documentation. This paper is presented with the majority of revisions incorporated as appendix A.

Chapter 2

Background

2.1 Time Series Analysis

A time series is a sequence of ordered observations. The observations are usually ordered by time, but they can be ordered by other variables such as depth or distance. Chatfield (2004) provides a good introduction to time series analysis. Percival and Walden (1993), hereinafter abbreviated as P&W93, provide an overview of spectral estimation techniques, including the multitaper method of spectral estimation. Priestley (1981) provides a reference on frequency domain techniques, Brockwell and Davis (1991) provide a theoretical overview of many of the time series techniques, and Shumway and Stoffer (2010) provide a modern reference, including R code and many examples but unfortunately omitting the multitaper technique. Anderson (1971, p. 1) notes the key difference between time series and other statistical analysis as follows:

The feature of time series analysis which distinguishes it from other statistical analyses is the explicit recognition of the importance of the order in which the observations are made. While in many problems the observations are statistically independent, in time series successive observations may be dependent, and the dependence may depend on the positions in the sequence. The nature of a series and the structure of its generating process may also involve in other ways the sequence in which the observations are taken.

2.2 Stochastic Processes

We are interested in time series that generally come under the category of “stochastic processes,” which are described as *a statistical phenomenon that evolves in time*

according to probabilistic laws (Chatfield, 2004, p. 27). The theory of stochastic processes is well developed and is beyond the scope of this work. An introductory book on the subject that covers the frequency domain is Papoulis and Pillai (2001), a discussion of stationary stochastic processes is presented in Grenander and Rosenblatt (1984), and Cox and Miller (1965, ch. 7) provide suitable supplementary information.

If the sample space is the ensemble of all possible realizations, then at any fixed time, we can define a random variable as a function from the sample space of all possible outcomes to the real line for a real-valued random variable, $X(t)$, that describes the outcome of the experiment at time t . A stochastic process, $\{X(t) : t \in T\}$ is the family of random variables indexed by t , where t belongs to some given set T .

Most statistical problems are concerned with estimating the properties of the population based on a sample. An investigator typically determines sample size and how randomness is incorporated into the sample. In time series analysis, the observations are determined by time, and it is rarely possible to take more than one sample at a given time. While it may be possible to increase the sample size—i.e., the length of the series—there will only be one sample at each time t . We can imagine an infinite set of time series, an ensemble, where every member of the ensemble is a possible realization of a stochastic process and the time series is a particular realization of the ensemble. Ergodic theorems, discussed briefly in Section 2.4.1, seek to address this theoretically.

2.2.1 Gaussian Distribution

The Gaussian, or *normal*, distribution is used frequently in statistics and is defined as

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}, \quad (2.1)$$

where μ is the arithmetic mean, and σ is the standard deviation.

2.3 Stationary Time Series

A time series is said to be *stationary* if there is no systematic trend, if the variance does not change systematically, and if all strictly periodic components have been removed (Chatfield, 2004, p. 13). Briefly, a process is stationary if its statistics do not depend on the origin.

A process $X(t)$ is strictly stationary if, for all $N \geq 0$, for any t_0, t_1, \dots, t_{N-1} contained in the index set, and for any τ , such that $t_0 + \tau, t_1 + \tau, \dots, t_{N-1} + \tau$ are also contained in the index set, the joint cumulative distribution function (CDF) of $X(t_0), X(t_1), \dots, X(t_{N-1})$ is the same as that of $X(t_0 + \tau), X(t_1 + \tau), \dots, X(t_{N-1} + \tau)$. The probabilistic structure of a strictly stationary process is invariant under a time shift (P&W93).

Second-order stationarity (also called weak, wide-sense, or covariance stationarity) is defined when for all $n \geq 0$, for any t_0, t_1, \dots, t_{N-1} contained in the index set and for any τ , such that $t_0 + \tau, t_1 + \tau, \dots, t_{N-1} + \tau$ are also contained in the index set, all the joint moments of orders one and two of $X(t_0), X(t_1), \dots, X(t_{N-1})$ exist, are finite, and are equal to the corresponding joint moments of $X(t_0 + \tau), X(t_1 + \tau), \dots, X(t_{N-1} + \tau)$.

τ) (P&W93). In this paper, stationary refers second-order stationary process unless otherwise specified.

Third- and fourth-order stationarity can also describe a process. Their can be defined in a manner similar to the defining of second-order stationarity. Higher-order stationarity implies lower-order stationarity (Priestley, 1981, p. 113). In the case of a real-valued Gaussian process, second-order stationarity implies strict stationarity, but this is not generally the case for a complex-valued Gaussian second-order stationary process (Cramér and Leadbetter, 1967, pp. 122–123).¹

In weak stationary processes, the mean, μ , and variance σ^2 are constant and do not depend on time,

$$\mu(t) = \mu \tag{2.2}$$

$$\sigma^2(t) = \sigma^2. \tag{2.3}$$

The autocovariance function (acvf) is defined as

$$\begin{aligned} R(\tau) &= \text{Cov}(X(t), X(t + \tau)) \\ &= \text{E}\{[X(t) - \mu][X(t + \tau) - \mu]\}, \end{aligned} \tag{2.4}$$

and it depends only on the lag τ . Restricting τ to discrete time steps (2.4) becomes the autocovariance sequence (acvs), denoted R_τ . The autocorrelation sequence is the standardized autocovariance sequence; this is discussed in Section 5.2.1.

Autocorrelation—that is, correlations between samples in the same series at different times—were first used in Cave-Browne-Cave (1905) while studying meteorological data. She had earlier worked with Karl Pearson and computed correlations between

¹Here, in addition to the ordinary autocovariance $r(t - u) = \text{E}\{X(t), X^*(u)\}$, one must have the “outer covariance” $q(t, u) = \text{E}\{X(t), X(u)\} = q(t - u)$. Note that $X(u)$ is not conjugated.

stations (Cave-Browne-Cave and Pearson, 1902); thus autocorrelations were invented shortly after Pearson invented correlation. The terms “autocorrelated” and “serial correlated” are used interchangeably.

It is known that the climate system is not stationary (Thomson, 1990b), but Tukey (1961) cautions not to allow this to prevent harmonic analysis of the data.

“The assumption of stationarity is one at which the innocent boggle, sometimes even to the extent of failing to learn what the data would tell them if asked. I have yet to meet anyone experienced in the analysis of time series (Gwilym Jenkins is the outstanding exception) who is over-concerned with stationarity.”

We propose a method to enhance the analysis by considering nonstationarity while recognizing Tukey’s concern. Spectral analysis is robust to nonstationarity; the technique frequently detects harmonic components that are present for only a significant fraction of the time series. The introduced methodology recognizes this and adds precision by (1) estimating the level-of-change across time and (2) locating change-points based on detecting structural (harmonic component) changes over time.

2.4 Several Definitions

2.4.1 Ergodic Theorems

This may not be immediately obvious, but it is possible to obtain a consistent estimate of the properties of a stationary process from a single finite realization. *Ergodic theorems* show that, for most stationary processes met in practice, sample moments of

finite length N converge in mean square to the population moments as $N \rightarrow \infty$ (Chatfield, 2004, p. 49). A more detailed explanation is given in Yaglom (1987b, pp. 210–224). This thesis does not further address the concept of ergodicity, which has been found to be of limited use in practical applications in the physical sciences (P&W93, p. 190).

2.4.2 Cumulants and Polygamma Functions

These terms will be used in Section 3.5.2 and are defined here. The *cumulants* κ_n of a random variable X are defined with the cumulant-generating function

$$g(t) = \log \mathbb{E}(e^{tX}). \quad (2.5)$$

The cumulants are obtained from the power series expansion of the cumulant generating function

$$g(t) = \sum_{n=1}^{\infty} \kappa_n \frac{t^n}{n!}. \quad (2.6)$$

The *gamma* function is defined for all complex numbers except the negative integers and zero. For complex numbers with a positive real part, the gamma function is defined as

$$\Gamma(t) = \int_0^{\infty} x^{t-1} e^{-x} dx. \quad (2.7)$$

The *polygamma* function of order m , $\psi^{(m)}(z)$, of complex value z is defined as the $m + 1$ derivative of the logarithm of the gamma function:

$$\psi^{(m)}(z) = \frac{d^{m+1}}{dz^{m+1}} \ln \Gamma(z). \quad (2.8)$$

2.5 Spectral Density Function

$S(f)$ is the spectral density function. Assuming the $S(f)$ is square integrable, and if $R(\tau)$ is square summable, then for a time step of $\Delta t = 1$ the following holds in mean-square:

$$S(f) = \sum_{\tau=-\infty}^{\infty} R(\tau)e^{-i2\pi f\tau}, \quad (2.9)$$

and conversely

$$R(\tau) = \int_{-1/2}^{1/2} S(f)e^{i2\pi f\tau} df. \quad (2.10)$$

2.6 Spectral Estimation

This section introduces the spectrum. The paleoclimate datasets under study are real-valued, and we restrict attention here to real-valued time series. We will use the multitaper spectrum estimates (Thomson, 1982, 2001; Thomson et al., 2007; Park et al., 1987; Lindberg and Park, 1987). Other methods, such as classical Blackman-Tukey, classical periodogram, autoregressive, maximum likelihood, Prony and Pisarenko methods are reviewed in Kay and Marple (1981). In general, the multitaper method provides improved bias and variance properties over earlier estimators at a computational cost. A comparison of multitaper spectral estimation and Welch's (windowed) overlapped segment average (WOSA) indicates the multitaper method has a performance advantage (Bronez, 1992).

2.6.1 Periodogram

Given N observations $x(t)$ for $t = 0, 1, \dots, N-1$ equally spaced in time at $\Delta t = 1$, the question is how to estimate the spectrum. The original solution was the *periodogram* which is the square of the discrete Fourier transform (DFT) scaled by $1/N$,²

$$P(f) = \frac{1}{N} \left| \sum_{t=0}^{N-1} x(t) e^{-i2\pi ft} \right|^2. \quad (2.11)$$

The periodogram was suggested in Stokes (1879), then named and analyzed in Schuster (1898). Einstein introduced the concept of power spectrum without using the term (Einstein, 1987; Yaglom, 1987a).

The periodogram is the Fourier transform of the sample autocovariance sequence,

$$P(f) = \sum_{\tau=-(N-1)}^{N-1} \hat{R}_\tau e^{-i2\pi f\tau}, \quad (2.12)$$

where \hat{R}_τ is the sample acvs. The frequency domain of such estimates is in the Nyquist band $-1/2 \leq f < 1/2$. This is a stationary version of the Einstein-Wiener-Khintchine theorem (Einstein, 1987; Khintchine, 1934).

Thomson (1977a) has given an example where the periodogram was in error by a factor of greater than 10^{10} over most of the frequency range. Periodograms have two major problems, variance and bias. The periodogram is an inconsistent estimator, as the variance does not decrease with sample size; this was first pointed out by Rayleigh (1903). Brillinger (2001) has shown that if $x(t)$ is a strictly stationary process with an acvs such that

$$\sum_{\tau=-\infty}^{\infty} |\tau R_\tau| < \infty, \quad (2.13)$$

²This can also be considered single tapered spectral estimate with a constant taper.

then the rate of decrease for the bias is given by

$$\mathbb{E}\{P(f)\} = P(f) + O\left(\frac{1}{N}\right). \quad (2.14)$$

The regularity condition implies that the acvs decays to zero quickly. Alternatively, it implies the true spectrum, $S(f)$, is a smooth function; $S(f)$ has a continuous first derivative (P&W93).

2.6.2 Direct Spectral Estimator

The unsmoothed direct spectral estimator that followed is defined as

$$S_{\mathbb{D}}(f) = \left| \sum_{t=0}^{N-1} x(t)\mathbb{D}(t)e^{-i2\pi ft} \right|^2. \quad (2.15)$$

In this case, $\mathbb{D}(t)$ is a data window or data taper. If we let $\mathbb{D}(t) = 1/\sqrt{N}$, (2.15) becomes (2.11). There are multiple window choices available, and many are described in Harris (1978), Kay and Marple (1981) and (P&W93). These windows were developed to have particular frequency response. As a window down weights or “tapers” the edges of data, some data is lost as a result of tapering, and each window has an associated variance inflation factor. A bias variance trade-off exists in window development. Once a window is selected, variance can be reduced by averaging adjacent frequency terms, averaging overlapped segments, and, if possible, making use of multiple orthogonal windows. A common standard window choice is the Hanning window, introduced by Julius von Hann (P&W93, p. 210), which is defined as a 100% cosine window,

$$\frac{C}{2} \left[1 - \cos\left(\frac{2\pi t}{N+1}\right) \right], \quad \text{for } 1 \leq t \leq \frac{\lfloor N \rfloor}{2}. \quad (2.16)$$

In (2.16), C is a constant used to normalize the window such that

$$\sum_t \mathbb{D}(t)^2 = 1. \quad (2.17)$$

Often the direct spectral estimator is convolved³ with a *smoothing window*. An example of convolution with a smoothing window is taking a running mean of five adjacent points⁴. The direct spectral estimate is the sum of two squares, the imaginary and real part of the discrete Fourier transform, and it has a chi-squared distribution with 2 degrees of freedom (Blackman and Tukey, 1959).

2.6.3 Multitaper Spectral Estimator

Multitaper spectral estimates, first introduced in Thomson (1982), make use of multiple direct spectral estimators with discrete prolate spheroidal sequence (DPSS) (Slepian, 1978), also called Slepian sequences, described in Section 2.9.1 as the windowing function. In this procedure, one selects an analysis bandwidth W , such that $0 < W \leq 1/4$, often $NW \approx 4$ to 6 (Thomson, 2001). One then selects $K \approx 2NW$ Slepian sequences to use as tapers. For each taper one computes the *eigencoefficients*,

$$y_k(f) = \sum_{t=0}^{N-1} x(t)v_t^{(k)}(N, W)e^{-i2\pi ft}, \quad (2.18)$$

where $v_t^{(k)}(N, W)$ is the k^{th} Slepian sequences for parameters N, W and $k = 0, 1, \dots, K - 1$. The crudest multitaper spectrum estimator is an average of the eigencoefficients,

$$\bar{S}(f) = \frac{1}{K} \sum_{k=0}^{K-1} |y_k(f)|^2. \quad (2.19)$$

³The convolution product, sometimes called the “resultant” or “Faltung” of two functions f and g , is defined as $(f * g)(x) = \int_0^{2\pi} f(x-t)g(t)dt$ (Davis, 1963).

⁴The terms data window and data taper refer to a window applied prior to a Fourier transform, whereas a smoothing window is applied to a dataset or a spectral estimate.

An alternative to averaging eigencoefficients is to weight them with their associated eigenvalue, λ_k from (2.44), giving

$$\bar{S}(f)_\lambda = \left(\sum_{k=0}^{K-1} \lambda_k \right)^{-1} \sum_{k=0}^{K-1} \lambda_k |y_k(f)|^2. \quad (2.20)$$

This is similar to (2.19) as the eigenvalues λ_k used are close to one.

The advantages of the multitaper spectral estimator are summarized in P&W (1993, pp. 331–332). We consider the practical advantages as (1) it allows the use of the Slepian tapers with the superior concentration properties, and (2) it provides an estimator with increased degrees-of-freedom without increased bandwidth. The multitaper spectral estimate is chi-squared distributed with $2k$ degrees of freedom as long as f is not too close to the zeroth, or Nyquist frequency. The edge frequencies, zero and Nyquist, have half the degrees-of-freedom.

2.6.4 Thomson F -test

The multitaper method provides a harmonic F -test for periodic components in coloured noise (Thomson, 1982). To describe the F -test, we first define $U_k(N, W; 0)$ as the discrete prolate wave function, which is also the Fourier transform of the Slepian sequence $v_t^{(k)}(N, W)$, taken with $f = 0$. Then the harmonic F -test is defined as

$$F(f) = \frac{(K-1)|\hat{\mu}|^2 \sum_{k=0}^{K-1} U_k(N, W; 0)^2}{\sum_{k=0}^{K-1} |y_k(f) - \hat{\mu}(f)U_k(N, W; 0)|^2}, \quad (2.21)$$

with 2, and $2K - 2$ degrees of freedom. Our mean, $\hat{\mu}(f)$, is estimated by

$$\hat{\mu}(f) = \frac{\sum_{k=0}^{K-1} U_k(N, W; 0)y_k(f)}{\sum_{k=0}^{K-1} U_k^2(N, W; 0)}. \quad (2.22)$$

As the Fourier transform when $f = 0$, is simply a sum, in practice $U_k(N, W; 0)$ is the sum of the Slepian sequences of even order, as the odd-order Slepian sequences are

known to sum to zero. We can use this along with multitaper spectra to locate and assess the significance of harmonic components found in a time series.

2.7 Spectral Representation of a Stationary Process

Papers explaining the multitaper spectral estimate generally begin with the *Cramér Spectral Representation* (Cramér, 1940; Thomson, 1990b, 1982; Park et al., 1987; Thomson, 2001). In this section, we motivate the spectral representation theorem following the procedure in P&W(1993).

Initially we consider the spectral representation theorem for real-valued discrete time harmonic process

$$X_t = \sum_{l=1}^L D_l \cos(2\pi f_l t + \phi_l), \quad t = 0, \pm 1, \pm 2, \dots, \quad (2.23)$$

where $L \geq 1$; D_l and f_l are real-valued constants. The f_l 's are distinct, $f_l > 0$, and the terms ϕ_l are independent random variables having a rectangular distribution on $[-\pi, \pi]$. This is a zero-mean harmonic process. Assume the frequencies are ordered such that $0 < f_l < f_{l+1} \leq 1/2$. Here the Nyquist frequency is $1/2$ and $\Delta t = 1$. Using the definition

$$\cos(\theta) = \frac{e^{i\theta} + e^{-i\theta}}{2}, \quad (2.24)$$

we can rewrite (2.23) as

$$X_t = \sum_{l=-L}^L C_l e^{i2\pi f_l t}, \quad (2.25)$$

where $C_l = D_l e^{i\phi_l}/2$ and $C_{-l} = D_l e^{-i\phi_l}/2$, for $l = 1, \dots, L$; $C_0 = 0$; $f_0 = 0$; and $f_{-l} = f_l$. Note: $-l$ refers to the complex conjugate, specifically $C_{-l} = C_l^*$.⁵ The $2L+1$ random variables $C_{-L}, C_{-L+1}, \dots, C_L$ are mutually uncorrelated, with moments

$$\mathbb{E}\{C_l\} = 0, \quad \text{and} \quad (2.26)$$

$$\text{Var}\{C_l\} = \mathbb{E}\{|C_l|^2\} = D_l^2/4. \quad (2.27)$$

If we define $D_0 = 0$, and $D_{-l} = D_l$, we can find the variance of X_t defined in (2.23) as

$$\text{Var}\{X_t\} = \sum_{l=-L}^L D_l^2/4. \quad (2.28)$$

The variance of the stationary process can be decomposed into a sum of components $\mathbb{E}\{|C_l|^2\}$. We can define a variance spectrum by

$$S^{(V)}(f) = \begin{cases} D_l^2/4, & \text{if } f = f_l, l = 0, \pm 1, \dots, \pm L, \\ 0, & \text{otherwise.} \end{cases} \quad (2.29)$$

We can define the complex-valued stochastic process

$$Z(f) = \sum_{j=0}^l C_j, \quad f_l \leq f_{l+1} \text{ with } l = 0, \dots, L, \quad (2.30)$$

where $f_{L+1} = 1/2$, and $Z(0) = 0$. $Z(f)$ is a “jump” process on the interval $[0, 1/2]$ with a random complex-valued jump at each f_l . Then

$$Z(f) = \begin{cases} 0, & \text{for } 0 \leq f \leq f_1, \\ C_1, & \text{for } f_1 < f \leq f_2, \\ C_1 + C_2, & \text{for } f_2 < f \leq f_3, \end{cases} \quad (2.31)$$

⁵ $\text{Cov}(Z_1, Z_2) = \mathbb{E}\{[Z_1 - E[Z_1]]^* [Z_2 - E[Z_2]]\}$. A superscripted * denotes complex conjugate.

and so forth.

We now define an orthogonal increment process as

$$dZ(f) = \begin{cases} Z(f + df) - Z(f), & 0 \leq f < 1/2, \\ 0, & f = 1/2, \text{ and} \\ dZ^*(-f), & -1/2 < f < 0. \end{cases} \quad (2.32)$$

In this case, df is a small increment such that $0 < f + df < 1/2$ when $0 < f < 1/2$.

For $l \geq 0$ we have

$$\begin{aligned} dZ(f_l) &= Z(f_l + df) - Z(f_l) \\ &= \sum_{j=0}^l C_j - \sum_{j=0}^{l-1} C_j \\ &= C_l. \end{aligned} \quad (2.33)$$

For any $f \neq f_l$ for some l , $dZ(f) = 0$ for df that are sufficiently small. As $E\{C_l\} = 0$ and $dZ(f)$ is either 0 or C_l , $E\{dZ(f)\} = 0$. Next, provided that f, f', df, df' , are such that the intervals $[f, f + df]$, and $[f', f' + df']$ do not intersect, the random variables $dZ(f)$ and $dZ(f')$ are uncorrelated,

$$\text{Cov}\{dZ(f), dZ(f')\} = E\{dZ^*(f), dZ(f')\} = 0. \quad (2.34)$$

We can show that the random variables are uncorrelated if we consider the following cases:

1. Neither interval covers a jump point, f_l
2. One interval covers a jump point and
3. Both intervals cover different jump points.

Based on (2.34), the process $\{Z(f)\}$ has *orthogonal increments* and is called an *orthogonal process*. Note that $\text{Var}\{dZ(f)\} = D_f^2/4$.

Let $g(f)$ be a continuous function over the interval $[-1/2, 1/2]$, and let $H(f)$ be a step function defined over the same intervals with jumps at

$$-1/2 < a_1 < a_2 < \dots < a_N < 1/2$$

with finite sizes b_1, b_2, \dots, b_N . Using a Riemann-Stieltjes integral, we can write

$$\int_{-1/2}^{1/2} g(f) dH(f) = \sum_{k=1}^N g(a_k) b_k. \quad (2.35)$$

Next, let $g(f) = e^{i2\pi ft}$ and $H(f) = Z(f)$ and we can rewrite (2.25) as

$$X_t = \int_{-1/2}^{1/2} e^{i2\pi ft} dZ(f). \quad (2.36)$$

This is a stochastic version of a Reimann-Stieltjes integral, and (2.36) is the *spectral representation* for the stationary process (2.25). The spectral representation theorem (2.36) is accurately named; it is a convenient form used primarily to establish properties of estimates. We cannot think of an example in which it represents reality.

We now state, without proof, the *spectral representation theorem for a continuous parameter stationary process* (Priestley, 1981). Let $\{X(t)\}$ be a real-valued discrete parameter stationary process with zero mean. There exists an orthogonal process, $\{Z(f)\}$, defined on $[-1/2, 1/2]$ such that

$$X(t) = \int_{-\infty}^{\infty} e^{i2\pi ft} dZ(f) \quad (2.37)$$

for all t . Note the difference in the limits of integration between (2.37) and (2.36). The process $\{Z(f)\}$ has the following properties:

1. $E\{dZ(f)\} = 0$ for all f ;

2. $E\{|dZ(f)|^2\} = dS^{(I)}(f)$ for all f , where the *integrated spectrum* $S^{(I)}(f)$ is bounded and nondecreasing; and
3. Any two distinct frequencies, f and f' on $[-1/2, 1/2]$, are uncorrelated.

This theorem holds for stochastically continuous processes.⁶ If $S^{(I)}(f)$ is differentiable everywhere,

$$E\{|dZ(f)|^2\} = dS^{(I)}(f) = S(f)df. \quad (2.38)$$

2.8 Nonstationary Harmonizable Process

We now consider the spectral representation theorem for nonstationary *harmonizable* processes (Loève, 1946, p. 464). We begin with N contiguous samples of a continuous-time process $x(t)$, $t = 0, \dots, N - 1$. In this case the Cramér spectral representation theorem, as before, is

$$x(t) = \int_{-\infty}^{\infty} e^{i2\pi\nu t} dX(\nu). \quad (2.39)$$

In (2.39), $dX(\nu)$ is a complex-valued increment process (or the generalized Fourier transform) of the process $x(t)$. The difference between (2.39) and (2.37) is that $dX(\nu)$ has *non-orthogonal* increments (Hanssen and Scharf, 2002). The covariance function of the process is

$$\begin{aligned} \text{Cov}(t_1, t_2) &= E\{x(t_1), x^*(t_2)\} \\ &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} e^{i2\pi(t_1 f_1 - t_2 f_2)} E\{X(f_1), X^*(f_2)\}. \end{aligned} \quad (2.40)$$

Thomson (2001) points out

⁶ $X(t)$ is stochastically continuous at $t = t_0$ if and only if $\lim_{t \rightarrow t_0} E\{[X(t) - X(t_0)]^2\} = 0$ (Priestley, 1981, p. 151).

“... the essential feature of a nonstationary process, namely, that there is a correlation between different frequencies. If the process is stationary, the correlation must depend only on $t_1 - t_2$ and not explicitly on t_1 or t_2 .”

This means that, in the case of a stationary process, we are only interested in one column of the variance covariance matrix, as all other columns are cyclic shifts—that is, the matrix is Toeplitz. In a harmonizable process, we must consider the entire matrix.

2.9 Multitaper Spectral Estimation Overview

Multitaper spectral estimation differs from WOSA estimates (Welch, 1967a) in that instead of using a single Hamming window⁷ on overlapped segments, one will use multiple orthogonal Slepian sequence tapers on the entire length of the time series, N . One performs several multitaper spectral estimates on sections, or *blocks*, of the time series. When these block estimators are plotted sequentially in colour, a spectrogram is formed. In constructing a spectrogram, one must consider the appropriate length, bandwidth, and allowable overlap in selecting block size.

Three important variables are used in this approach (Thomson, 2001).

1. The bandwidth parameter W , which must be chosen on the basis of fundamentals (principles of physics, for example) of the problem. Thomson (2001) cautions that the assumptions about the physical properties could be wrong and that one should investigate several choices.

⁷This Hamming window is similar to the Hanning window introduced on page 12, with the difference being that the edges of this one do not taper to zero. It is defined as $\alpha - \beta \cos\left(\frac{2\pi t}{N+1}\right)$.

2. The block size N_b or, equivalently the time-bandwidth product. Typically $N_bW \approx 4$ or 6 is a good starting point. If N_bW is too small, the estimator will have a poor sidelobe performance, and if N_bW is too large the estimates will have poor frequency resolution (Thomson, 1990a).
3. The offset between blocks, Δd . Often $N_b/2$ is used as the offset between blocks. WOSA estimates use a 50% offset which is equivalent to $\Delta d = N/2$.

Welch was timely—his estimator was introduced just after the introduction of the fast Fourier transform (FFT), but averaging periodograms over different times was also mentioned in Schuster (1898).

Details of the multitaper spectral estimator for full-length time series follow. These details apply to a single section, or block, of the series if N is replaced with N_b . Multitaper estimates of the spectrum are based on approximately solving the integral equation that expresses the projection of $dX(f)$ onto the Fourier transform of the data, $y(f)$ (Strang, 2005, pp. 204–206). If one takes the discrete Fourier transform of the observed data,

$$y(f) = \sum_{t=0}^{N-1} x(t)e^{-i2\pi ft}, \quad (2.41)$$

and uses the spectral representation (2.39) for $x(t)$, one gets the fundamental equation of spectrum estimation,

$$y(f) = \int_{-1/2}^{1/2} \mathbf{K}_N(f - \xi) dX(\xi), \quad (2.42)$$

where the kernel, the Dirichlet kernel multiplied by a phase factor, is defined as

$$\mathbf{K}_N(f) = \frac{\sin(N\pi f)}{\sin(\pi f)} \exp\left(-i2\pi f \frac{N-1}{2}\right). \quad (2.43)$$

There are several key points about (2.43).

1. As one can take the inverse Fourier transform of $y(f)$ and recover $x(t)$ for $0 \leq t \leq N - 1$, $y(f)$ is a trivially sufficient statistic and completely equivalent to the original data. In practice, when an FFT is used, the inverse FFT will return the original data. The inverse can be performed given either: (1) the complete FFT result including any redundant complex conjugate in the case of real data, or (2) the FFT result without the redundant complex conjugate and the original length (Frigo and Johnson, 2005). The latter can only be used in the case of real data input to the FFT.
2. The finite Fourier transform $y(f)$ is *not* equivalent to $dX(f)$, because $dX(f)$ is assumed to generate the entire data sequence for all t , not just the observed samples.
3. $\frac{1}{N} |y(f)|^2$, the periodogram, is not the spectrum; it is biased and inconsistent.
4. (2.42) is a convolution of dX with a Dirichlet kernel. Thomson (1982) regards it as a Fredholm integral of the first kind which does not have a unique solution, but it does have approximate solutions.
5. Multitaper spectral estimators refer to the class of estimators that use any set orthogonal data tapers, and this work focuses on using Slepian sequences as orthogonal data tapers.

2.9.1 Discrete Prolate Spheroidal Sequences

The multitaper estimates used in this thesis use the discrete prolate spheroidal wave functions (dpswfs). These functions provide reasonable solutions to (2.42). The Slepian sequences $v_n^{(k)}(N, W)$ are defined as real, unit-energy sequences on $[0, N - 1]$

having the greatest in bandwidth energy, W , and are the solutions to the symmetric Toeplitz matrix eigenvalue equation,

$$\lambda_k v_n^{(k)} = \sum_{m=0}^{N-1} \frac{\sin(2\pi W(n-m))}{\pi(n-m)} v_m^{(k)}, \quad \text{for } 0 \leq n \leq N-1. \quad (2.44)$$

From this point we will write $v_t^{(k)}$ to indicate $v_t^{(k)}(N, W)$ for the Slepian sequences, and the arguments N and W are implied. To compute the Slepian sequences, we use the tridiagonal form given in Slepian (1978). In practice, the LAPACK functions `dstezbz` and `dstein` (Anderson et al., 1999) are used as described in P&W93. These LAPACK functions are called from the multitaper R package; see page 177 for details on obtaining Slepian sequences using the R package. We used the normalization used in Thomson (1982); Park et al. (1987), and not those used in P&W93 or the signal processing toolbox in Matlab. Thomson (1990a) defines

$$V_k(N, W; f) = \sum_{n=0}^{N-1} v_n^{(k)} e^{-i2\pi n f}, \quad (2.45)$$

the $V_k(N, W; f)$ satisfy the homogeneous integral equation,

$$\lambda_k V_k(f; N, W) = \int_{-W}^W \mathbf{K}_N(f - \nu) V_k(N, W; \nu) d\nu. \quad (2.46)$$

For simplicity, $V_k(f)$ will be written instead of $V_k(f; N, W)$ from here. The eigenvalues are bounded between 0 and 1 with $K \approx \lfloor 2NW \rfloor$, with “large” eigenvalues near one. The dimension of the subspace is $2NW$, and the eigenvalues give the fraction of the energy in the bandwidth $(-W, W)$, such that $1 - \lambda_k$ defines the “leakage” of the k^{th} window.

Because of the energy concentration of the Slepian sequences, the approximation,

$$e^{-i2\pi f t} \approx \sum_{k=0}^{K-1} v_t^{(k)} V_k(f), \quad (2.47)$$

on $L_2(-W, W) \times [0, N - 1]$ has better sidelobe leakage properties than other sets of orthogonal windows (Thomson, 2001, p. 327). The sequences are orthonormal, and the functions are orthonormal on $[-1/2, 1/2)$ and are also orthogonal on $(-W, W)$:

$$\int_{-W}^W V_j(f) V_k^*(f) df = \lambda_j \delta_{jk}. \quad (2.48)$$

2.9.2 Adaptive Multitaper Spectral Estimate

In general, the multitaper estimate (2.19) is not used. Instead, weights are used to replace averaging of the independent eigencoefficients (2.18). The weights are calculated iteratively from

$$d_k(f) \approx \frac{\sqrt{\lambda_k} S(f)}{\lambda_k S(f) + B_k(f)}, \quad (2.49)$$

where $B_k(f)$ is an estimate of the power in terms of broad-band bias. This weighting can be described as a Wiener filter applied to the eigencoefficients with the local terms considered as “signal” and the broad-band parts as “noise” (Thomson et al., 2007). An initial estimate and bound for the bias term is given by

$$B_k(f) \leq \sigma^2(1 - \lambda_k). \quad (2.50)$$

The initial weights, calculated from the initial estimate, are applied to the eigencoefficients creating the weighted eigencoefficients, which are used to obtain a new estimate, and this becomes the initial estimate in the next iteration. In practice, only a handful of iterations are required (Thomson, 1982). The canonical multitaper spectrum estimate is

$$\widehat{S}_x(f) = \frac{\frac{1}{K} \sum_{k=0}^{K-1} |d_k(f) y_k(f)|^2}{\frac{1}{K} \sum_{k=0}^{K-1} |d_k(f)|^2}. \quad (2.51)$$

The equations used to calculate the weights, (2.49), and the canonical multitaper, (2.51), follow the original definitions given in Thomson (1982); Thomson et al. (2007); Park et al. (1987).

The statistical properties of (2.51) are known, and some important points from Thomson et al. (2007) are listed below.

1. Bias from signals at frequencies *outside* the “local” band $(f - W, f + W)$ is bounded using the properties of the Slepian sequences and the Cauchy inequality (Thomson, 1982, 2001).
2. The eigencoefficients (2.18) can be inverted as

$$x(t) = \frac{\sum_{k=0}^{K-1} \nu_t^{(k)} \mathcal{F}^{-1}\{y_k(f)\}}{\sum_{k=0}^{K-1} [\nu_t^{(k)}]^2}, \quad (2.52)$$

where \mathcal{F}^{-1} is the inverse Fourier transform.

3. Under a locally white assumption, where we assume that the data is white within the bandwidth W , multitaper estimates can be jackknifed by deleting one window at a time (Thomson and Chave, 1991b). See Section 2.12.
4. Multitaper estimates are more efficient than conventional spectrum estimates (Thomson, 1982; Lindberg, 1986; Bronez, 1992; Hansson and Salomonsson, 1997).

2.10 Aliasing

Aliasing higher frequency components appearing as lower frequency components occurs in sampled time series and in digital spectral analysis. Suppose a continuous

time series $X(t)$ has the Fourier transform $G_c(f)$. If the series is sampled at discrete time intervals, $t = 0, 1, \dots, N - 1$, with equal spacing Δt , then denote the Fourier transform of the sampled series as $G_d(f)$. It can be shown (Blackman and Tukey, 1959) that

$$G_d(f) = \sum_{k=-\infty}^{\infty} G_c\left(f + \frac{k}{\Delta t}\right). \quad (2.53)$$

This indicates that $G_d(f)$ depends on a countably infinite set of frequencies $f + \frac{k}{\Delta t}$, for $k = \frac{\pm 1}{\Delta t}, \frac{\pm 2}{\Delta t}, \dots$. The Nyquist or folding frequency is defined as $\frac{1}{2\Delta t}$ (Nyquist, 1928).

2.11 Zero Padding Spectral Estimates

Zero padding is generally used in conjunction with FFT algorithms, and the default option using the multitaper R package (introduced in Appendix A) zero pads the data to twice the next power of two given the length of the data. Specifically, if n is the length of the data, then zero padding is performed to a total length, n_{FFT} , of

$$n_{\text{FFT}} = 2 \times 2^{\lceil \log_2(n) \rceil}, \quad (2.54)$$

where $\lceil \cdot \rceil$ represents the ceiling function.

The spectral estimate can be used as an intermediate step in calculating the autocovariance function (see Section 5.5.1 and Appendix A.3.4). In such cases, zero-padding to the length of $n_{\text{FFT}} \geq 2n$, as in (2.54), is required to prevent the calculation of circular autoregressive coefficients (Chatfield, 2004, pp. 138–139).

2.12 Jackknife Estimates

This section follows the development in Thomson and Chave (1991b). In jackknifing, separate estimates are formed by deleting one sample at time. This differs from bootstrapping, in which one resamples from the original sample with replacement (Wu, 1986; Good, 2001). Let $\{x_i\}$, for $i = 1, \dots, N$ be a sample of N independent observations drawn from some distribution characterized by a parameter θ , which we estimate by $\hat{\theta}$. We denote the estimate of θ using all N observations by $\hat{\theta}_{\text{all}}$. If we now create N additional estimates, $\hat{\theta}_{(i)}$, each one leaving out the i^{th} of the original samples. Each new *leave-one-out* estimate is made from $N - 1$ samples

$$\hat{\theta}_{(i)} = \hat{\theta}\{x_1, \dots, x_{i-1}, x_{i+1}, \dots, x_N\}. \quad (2.55)$$

Pseudovalues are defined as

$$p_i = N\hat{\theta}_{\text{all}} - (N - 1)\hat{\theta}_{(i)}. \quad (2.56)$$

The *jackknife estimate*, is the mean of the pseudovalues,

$$\tilde{\theta} = \frac{1}{N} \sum_{i=1}^N p_i. \quad (2.57)$$

This estimate was introduced as a lower bias replacement for $\hat{\theta}$. Let the average of the delete-one estimates be

$$\theta_{(\cdot)} = \frac{1}{N} \sum_{i=0}^N \hat{\theta}_{(i)}. \quad (2.58)$$

The *jackknife variance* is

$$\text{Var}\{\hat{\theta}_{\text{all}}\} = \frac{N - 1}{N} \sum_{i=1}^N \left[\hat{\theta}_{(i)} - \theta_{(\cdot)} \right]^2. \quad (2.59)$$

(2.59) was considered the variance estimate for $\tilde{\theta}$, but simulations with small samples indicate that it is more accurate than the variance estimate of $\hat{\theta}_{all}$ (Hinkley, 1978). This estimate of the variance, (2.59), has been shown to be conservative (Efron and Stein, 1981). In practice, it is informative to further research datasets or frequencies where jackknife variances do not coincide with theoretical variances. The idea of “jackknifing over tapers” was introduced in Blackman and Tukey (1959) and then developed in Thomson (1984).

In the case of the multitaper estimate, the jackknife is computed by leaving out one of the eigenspectra in computing each $\hat{\theta}_{(i)}$. Each $\hat{\theta}_{(i)}$ is calculated using (2.51) such that weights are calculated differently for each $\hat{\theta}_{(i)}$. In this case $\theta_{(\cdot)} \neq \hat{\theta}_{all}$. In practice, a two-standard-deviation jackknife confidence interval can be plotted on a multitaper spectral estimate, multitaper magnitude-squared coherence estimate and phase coherence estimate; these can be used in determining significance and can be used to draw attention to details in estimates in which theoretical and jackknife confidence intervals do not match.

2.13 Coherence

We define the coherence as a measure of the degree to which both variables are jointly influenced by cycles near frequency f (Jenkins and Watts, 1968; Koopmans, 1995).

We begin with the k^{th} autocovariance matrix for a two-dimensional process. Let

$$\mathbf{y} = \begin{bmatrix} X_t \\ Y_t \end{bmatrix}, \quad (2.60)$$

and define the k^{th} autocovariance matrix as

$$\begin{aligned}\mathbf{\Gamma}_k &= E \begin{bmatrix} (X_t - \mu_X)(X_{t-k} - \mu_X) & (X_t - \mu_X)(Y_{t-k} - \mu_Y) \\ (Y_t - \mu_Y)(X_{t-k} - \mu_X) & (Y_t - \mu_Y)(Y_{t-k} - \mu_Y) \end{bmatrix} \\ &= \begin{bmatrix} \hat{S}_{XX}^{(k)} & \hat{S}_{XY}^{(k)} \\ \hat{S}_{YX}^{(k)} & \hat{S}_{YY}^{(k)} \end{bmatrix}.\end{aligned}\quad (2.61)$$

Note that $\mathbf{\Gamma}_k^T = \mathbf{\Gamma}_{-k}$ where superscript T denotes transpose.

Now, given two series sampled at the same time, $x(t)$ and $y(t)$ for $t = 1, 2, \dots, N$, we define the sample cross-spectrum from x to y at frequency f as

$$\hat{S}_{yx}(f) = \sum_{k=-(N+1)}^{N-1} \hat{s}_{yx}^{(k)} e^{-i2\pi fk}.\quad (2.62)$$

The sample cross-spectrum can be written as the sample *cospectrum*, $\hat{c}_{yx}(f)$, and the sample *quadrature spectrum*, $\hat{q}_{yx}(f)$ (Jenkins and Watts, 1968),

$$\begin{aligned}\hat{S}_{yx}(f) &= \hat{c}_{yx}(f) + i\hat{q}_{yx}(f) \\ &= \hat{R}(f)e^{i\hat{\theta}(f)},\end{aligned}\quad (2.63)$$

where

$$\hat{R}(f) = \sqrt{[\hat{c}_{yx}(f)]^2 + [\hat{q}_{yx}(f)]^2}, \text{ and}\quad (2.64)$$

$$\hat{\theta}(f) = \tan^{-1} \left(\frac{\hat{q}_{yx}(f)}{\hat{c}_{yx}(f)} \right).\quad (2.65)$$

Sample magnitude-squared coherence (MSC) is defined as

$$|\hat{\gamma}(f)|^2 = \frac{|\hat{S}_{xy}|^2}{\hat{S}_{xx}(f)\hat{S}_{yy}(f)},\quad (2.66)$$

and the sample phase coherence, $\hat{\phi}(f)$, is defined as

$$\hat{\phi}(f) = \tan^{-1} \left(\frac{\hat{q}_{yx}(f)}{\hat{c}_{yx}(f)} \right). \quad (2.67)$$

In calculating the actual angle, one must consider the quadrant unless a function such as `atan2` in R is used (Bloomfield, 2000).

Using the multitaper method we calculate the cross-spectrum from the eigencoefficients or by the eigencoefficients weighted by the square root of the weights determined in (2.49),

$$S_{xy}(f) = \frac{1}{k} \sum_{k=0}^{K-1} x_k(f) y_k^*(f), \quad (2.68)$$

and then calculate the coherence as above (Kuo et al., 1990; Thomson, 1982).

The magnitude-squared coherence, when calculated using an untapered spectral estimate, such as the periodogram, is known to be biased (Carter et al., 1973a). In this thesis we use the multitaper method to correct for bias; specifically, the bias correction in Thomson and Chave (1991b, p. 87) is used in this thesis.

2.14 Spectrograms

Spectrograms were introduced in analog form in Koenig et al. (1946), and we plot the multitaper spectrogram or the high-resolution spectrogram introduced in Thomson (1998). This estimator is considered similar to the evolutionary periodogram estimator (Kayhan et al., 1994), which was introduced independently a few years earlier (Moghtaderi, 2009, pp. 24–26). A spectrogram is a graphical representation of multiple power spectral estimators in succession, each estimator based on a section, or “block,” of the entire data series. The spectrogram can consist of blocks with both

no overlap and considerable overlap. We construct spectrograms using the multitaper spectral estimator from Section 2.6.3.

Chapter 3

Frequency-domain Methods for Change-point Detection

3.1 Introduction

We introduce a frequency-domain methodology for detecting change-points in time series. The methodology is based on an introduced estimator of *level-of-change* between spectra, and it incorporates existing multitaper spectral estimate tools. We will use methodology on the nonstationary Pinot Noir grape harvest date (GHD) (Chuine et al., 2004) series in Chapter 4, and we will split the series into two locally stationary sections separated by a change-point. The majority of change-point detection techniques focus on detecting a change in mean, variance, slope, autoregressive coefficients or other parameter in a parametric model. The methodology proposed here detects a change-point by estimating the level-of-change in spectral estimates over time and locating a change-point where the level-of-change between adjacent spectra is high. We propose a method that focuses on detecting a change in spectral structure over time.

Assume that a time series with unit spacing is sectioned into m non-overlapping segments of equal length, and a spectral estimate $\hat{S}_j(f)$ is made for each segment $j \in 1, 2, \dots, m$ at frequency $f \in [0, 1/2]$. The proposed methodology uses the following estimate of level-of-change between spectral estimates:

$$\Delta_{j,j+1} \hat{S}(f) = [\ln(\hat{S}_j(f)) - \ln(\hat{S}_{j+1}(f))]^2. \quad (3.1)$$

Section 3.5 provides properties of (3.1), which is related to the Fisher Z -distribution (Fisher, 1924).

This chapter is organized as follows: Section 3.2 introduces the change-point problem; Section 3.3 provides a literature review of change-point techniques; Section 3.4 gives additional preliminary information; Section 3.5 introduces the frequency-domain

change-point estimator and discusses some of its properties; Section 3.6 studies the estimator using simulations on a stationary model and a model where there is change in the frequency component(s); Section 3.7 puts forward a methodology for using the estimator which incorporates existing tools; and Section 3.8 presents a summary and concluding remarks.

3.2 Change-points Problem Overview

A detailed background of change-points is presented in Zacks (1982). A majority of the change-point literature considers independent observations, not serially correlated time series.

Consider a finite sequence of N independent random variables X_1, X_2, \dots, X_N and a sequence of positive integer-valued parameters $2 \leq \tau_1 < \tau_2 < \tau_3 < \tau_n$, where $n \ll N$. The points τ_j for $j = 1, 2, \dots, n$ partition the data into epochs. For example, if there is one change-point, τ_1 there would be two epochs, $\{X_1, X_2, \dots, X_{\tau_1-1}\}$ and $\{X_{\tau_1+1}, X_{\tau_1+2}, \dots, X_N\}$, where the random variables in each epoch have the same identical distribution, either F_1 or F_2 . When this is extended to time series, the assumption of independence is dropped, but the residuals from a fitted parametric model, such as an autoregressive model, are still assumed independent.

Often the distributions F_1, F_2, \dots, F_{n+1} are known or are partially known, but the points of change τ_j are unknown. Many methods look for changes in the parameters of the distributions; for example, the parametric distributions F_1, F_2, \dots, F_{n+1} can be assumed except for a parameter such as mean, variance, location, or scale. The problem is to (1) estimate the parameters τ_j or (2) test hypotheses concerning these

points of change. The methodology presented here estimates change-points by locating structural change in the frequency domain, and then offers a primarily graphical test, as opposed to a statistical hypothesis test.

The literature covers various formulations of the problem and different approaches; there are formulations with one and multiple change-points; there are procedures for fixed and sequential sampling. The inference framework can be either classical or Bayesian.

Andrews (1993) commented that many of the change-point detection methods in the statistical literature are too simple for economic applications, and many consider scalar parameter models that deal with independent observations. We contend that current change-point detection techniques make distributional assumptions that provide statistical power; however, parametric models are restrictive, and we propose a nonparametric frequency-domain change-point approach.

Change-point problems consider the following two cases: (1) where some information is available about a possible change-point, such as the one that occurs at the time of a wide-scale change in agricultural practice following European colonization, and (2) where no prior information is known about a possible change-point. In the context of economics time series Andrews (1993, p. 825) discussed “rolling change-point tests” used in econometric packages. Structural changes indicate a change in model specification. There are partial-sample and full-sample generalized method-of-moments (GMM) estimators. Another popular method is the cusum (cumulative sum) method suggested in Brown et al. (1975), which is used for linear regression models.

The terms change-point and *disorder* (in Eastern Europe) have been used to describe a change in the statistical distribution of samples. We are interested specifically in changes over time that are observed in the spectral (time \times frequency) domain. Chen and Gupta (2012, pp. 1–5) give an introduction to the *change-point problem*. Their discussion, like the methodology presented here, focuses on offline change-point analysis, or retrospective change-point analysis of finite samples, as opposed to on-line¹ change-point detection, which occurs in real-time with sequential data. Online change-point problems are generally presented in statistical quality control, public health surveillance, and signal processing (see Mei (2006) for an overview).

A broad selection of methods are used in change-point detection, including maximum likelihood ratio tests, Bayesian based tests, nonparametric tests, stochastic process analysis and information theory approaches. Generally parametric change-point problems focus on detecting change in the mean, scale and shape parameters of a probability distribution, whereas nonparametric methods focus on rank or order statistics, as in, for example, Pettitt (1979) studies Mann-Whitney-Wilcoxon-like (Mann and Whitney, 1947) tests in change-point detection. In economics, one looks for evidence of “structural change” related to externally influenced changes (Andrews, 1993).

3.3 Literature Review of Change-point Techniques

Change-point problems originated in the statistical quality control context, but have spread into areas such as stationarity of stochastic processes, estimation of the *current position* of a time series, and testing and estimating changes in the patterns

¹Online change-point analysis occurs in statistical control systems where an observed change requires an immediate correction, whereas offline change-point analysis occurs after the data has been collected, and the analysis is to determine whether (and where) a change occurred.

of regression models. Change-point techniques have also been used in comparisons and matching of DNA sequences. Techniques used in estimating change-points are maximum likelihood, isotonic regression, Bayesian techniques, piece-wise regression, nonparametric regression and grid searching (Khodadadi and Asgharian, 2008). In climate science, change in trend is located and assessed using ramp regression, a subset of piecewise regression (Mudelsee, 2010, pp. 150–157).

Statistical analysis of change-point problems depends on the method of data collection. If the data collection is ongoing up to a random time, the appropriate statistical procedure is called *sequential*, and time series fall in this category. If a finite large data set is collected with one change-point, the procedure is referred to as *non-sequential*.

Tsay (1989) introduces threshold models for multiple autoregressive (AR) processes. In this type of model, the type of AR process changes once a particular term, say x_{t-1} , passes a particular threshold. An example using such models on influenza data is presented in Shumway and Stoffer (2010, pp. 290–292). Their analysis indicated that a threshold model was acceptable for a prediction of up to one month ahead, but not beyond. Picard (1985) studied the asymptote of two methods for detecting that a failure (change) has occurred in a time series. Specifically, (1) changes that do not affect the mean of the observations but do affect changes to covariance structure and (2) detecting a failure and estimating the parameters of the failure occurring as a change in mean, covariance or autoregressive parameters were studied. For the first case, a form of Kolmogorov-Smirnov goodness-of-fit test to look at the nonparametric type of covariance structure was presented, and for the second, a likelihood ratio test finding that normalizations are required because of problems at the edges of the observation intervals was investigated.

An overview of recent contributions to the change-point literature is presented. In Seidou and Ouarda (2007), a recursion-based multiple linear regression change-point detection technique used to detect change-points in river streamflows is presented.

One recent theme in change-point detection research has focused on using Bayesian techniques to detect a change in parameter, such as mean or variance, in a parametric distribution. Examples of such techniques can be found in neuroscience (Pillow et al., 2011) and in climate science (Ruggieri, 2012).

Liu et al. (2013) propose using a nonparametric method of using the direct density-ratio to estimate the Pearson divergence (Pearson, 1900), and this method is applied to detecting nonstationarity in human-activity sensing, speech signals and Twitter messages. The methodology and divergence estimate proposed in Liu et al. (2013) is similar to the methodology proposed here; however, the method we propose focuses on changes detected in the frequency domain. Essentially, Liu et al. (2013) propose estimating divergence with direct density-ratio estimation, which has been explored in Kanamori et al. (2009).

In shorter time series, the change-point detection problem has been addressed by (1) fixing the number of discontinuities, and then using both Haar (square) wavelets with brute force minimization (Karl et al., 2000), and (2) creating a matrix of overdetermined linear equations and consecutively solving the system for every possible combination of change-points that satisfies the constraints (Tomé and Miranda, 2004). In longer time series, dynamic programming techniques have been developed to reduce the computational burden when considering all possible change-points (Ruggieri et al., 2009). Additionally, branch and bound techniques have been developed to screen and

eliminating sub-optimal segmentation (Aksoy et al., 2008). Another approach is to first identify the change-points by visual inspection and then refine the location to (1) minimize the number of change-points, (2) be consistent with previous research, and (3) have support from a nonparametric statistical method (Lanzante, 1996). This is an iterative approach that tests for statistical significance in an attempt to minimize *a priori* assumptions on the number and location of change-points. A semi-hierarchical splitting algorithm for placing change-points has been proposed (Menne, 2006). In this case, a splitting step is followed by a merge step to determine whether the change-points chosen earlier are still significant. A normal homogeneity test for one or more artificial discontinuities in climate time series has been proposed (Beaulieu et al., 2010). A Bayesian approach to detecting multiple change-points in any parametric regression model was proposed in Ruggieri (2012).

3.4 Additional Preliminaries

3.4.1 Bartlett M-Test

Thomson (1977b, p. 1994) suggests using the Bartlett M-test for heteroscedasticity (Bartlett, 1937) to test for nonstationarity as a function of frequency. This test evaluates the logarithm of the ratio of the arithmetic mean of spectral estimates, across blocks, over the geometric mean across blocks. If for each block $j \in 1, \dots, n_b$, where n_b is the number of blocks, we have a spectral estimate, $\hat{S}_j(f)$, then at frequency, f , the Bartlett M-test statistic is constructed as:

$$\widehat{M}(f) = n_b \nu \ln \left(\frac{1}{n_b} \sum_{j=1}^{n_b} \hat{S}_j(f) \right) - \nu \sum_{j=1}^{n_b} \ln \hat{S}_j(f), \quad (3.2)$$

where ν represents the chi-squared degrees-of-freedom, which, when the multitaper spectral estimate with K tapers is used without adaptive weighting, is $\nu = 2K$ except at the end points (see Section 2.6.3). When the individual spectra are non-zero, the geometric mean is at most equal to the arithmetic mean, and this test returns a high value when the ratio of arithmetic mean over geometric mean is high. Bartlett's M-test statistic is the likelihood-ratio test for homogeneity of chi-squared variates, but its use in ordinary statistics has been discouraged because of sensitivity to departures from normality; however, in spectrum estimates, the complex-valued Fourier transforms are close to Gaussian. Thomson (1990b, pp. 609–610) used this technique to show nonstationarity in annual growth rings of the Mount Campito (east-central California) bristlecone pine series.

3.4.2 Fisher Z -distribution

The level-of-change estimator introduced in (3.1) is related to the Fisher Z -distribution, the logarithm of the F -distribution, which is also used to compare two independent random variables with chi-squared distributions. It was introduced in Fisher (1924) and is given by

$$f(z; \nu_1, \nu_2) = \frac{2\nu_1^{\nu_1/2} \nu_2^{\nu_2/2}}{B(\nu_1/2, \nu_2/2)} \frac{e^{\nu_1 z}}{(\nu_1 e^{2z} + \nu_2)^{(\nu_1 + \nu_2)/2}} dz, \quad (3.3)$$

where B is the beta function, or Euler integral, and ν_1 and ν_2 are the degrees-of-freedom. We restrict ourselves to the case where $\nu_1 = \nu_2 = \nu$, and the Fisher Z -distribution, which is then symmetric, becomes

$$f(z, \nu, \nu) = \frac{\operatorname{sech}^\nu(z)}{2^{\nu-1} B(\nu/2, \nu/2)} dz. \quad (3.4)$$

A notable difference between the Fisher Z -distribution and the F -distribution is that when $\nu_2 > 2$, the F -distribution has the expected value $\frac{\nu_2}{\nu_2 - 2}$, whereas $E\{Z\} = 0$.

If the spectra are independent, the proposed level-of-change estimator is the square of an estimator with a Fisher- Z distribution, which is approximately Gaussian (Aroian, 1941), and thus the square has approximately a chi-squared distribution. The use of Fisher- Z distribution in this work is currently limited to the expected value in Section 3.5.2; however, the distribution may be of value in future work.

3.5 Level-of-change in Frequency-domain

3.5.1 Proposed Level-of-change

In this thesis, spectral estimates in (3.1) are computed using the multitaper procedure described in Section 2.6.3. It is known that such estimates have approximately a chi-squared distribution with $2K$ degrees-of-freedom, where K is the number of tapers used (P&W93, p. 222). We propose the following level-of-change estimator:

$$\hat{Q}_0(f) = [\ln \hat{S}_j(f) - \ln \hat{S}_{j+1}(f)]^2. \quad (3.5)$$

Section 3.5.2 examines statistical properties of \hat{Q}_0 .

3.5.2 Statistical Properties of the Estimator

We are interested in the mean and variance of the level-of-change estimator (3.5) at one frequency. In this section, we obtain the first two central moments for the level-of-change estimator and express these in terms of polygamma functions. We use simulations to check these values in Section 3.6.

In the following, the subscripts j and $j + 1$ are replaced with 1 and 2 for convenience. Begin assuming that \hat{S}_1 and \hat{S}_2 are independent spectrum estimates at any frequency f . We omit the frequency f in this section for convenience.

Let

$$\hat{Z}_0 = \ln \frac{\hat{S}_1}{\hat{S}_2}. \quad (3.6)$$

For examining the level-of-change in two spectral estimates, consider the two estimates $\hat{l}_1 = \ln \hat{S}_1$, and $\hat{l}_2 = \ln \hat{S}_2$. Let $\bar{l} = E\{\hat{l}_1\} = E\{\hat{l}_2\}$, then $\hat{Z}_0 = \hat{l}_1 - \hat{l}_2$, and $E\{\hat{Z}_0\} = 0$.² Then we can write \hat{Q}_0 as

$$\begin{aligned} \hat{Q}_0 &= \hat{Z}_0^2 \\ &= (\hat{l}_1 - \hat{l}_2)^2 \\ &= [(\hat{l}_1 - \bar{l}) - (\hat{l}_2 - \bar{l})]^2 \\ &= (\hat{l}_1 - \bar{l})^2 - 2(\hat{l}_1 - \bar{l})(\hat{l}_2 - \bar{l}) + (\hat{l}_2 - \bar{l})^2. \end{aligned} \quad (3.7)$$

We take the expected value of \hat{Q}_0 ,

$$\begin{aligned} E\{\hat{Q}_0\} &= E\{(\hat{l}_1 - \bar{l})^2 - 2(\hat{l}_1 - \bar{l})(\hat{l}_2 - \bar{l}) + (\hat{l}_2 - \bar{l})^2\}, \\ &= E\{(\hat{l}_1 - \bar{l})^2\} + E\{(\hat{l}_2 - \bar{l})^2\}. \end{aligned} \quad (3.8)$$

As $E\{(\hat{l}_1 - \bar{l})^2\} = E\{(\hat{l}_2 - \bar{l})^2\}$, we write (3.8) as

$$E\{\hat{Q}_0\} = 2 E\{(\hat{l} - \bar{l})^2\}, \quad (3.9)$$

omitting the subscripts 1 and 2 in \hat{l} . Equation (3.9) will be written in terms of a trigamma function by (1) writing the second central moment of \hat{Z}_0 in terms of

²If we do not require $\bar{l} = E\{\hat{l}_1\} = E\{\hat{l}_2\}$, then $E\{\hat{Z}_0\} = 0$ as \hat{Z}_0 has a Fisher Z distribution.

cumulants of $\ln \hat{S}$ (Stuart and Ord, 2010, pp. 88–89), and (2) equating cumulants of $\ln \hat{S}$, the natural logarithm of an estimate with a chi-squared distribution, to a trigamma function (Bartlett and Kendall, 1946, p. 128).

Next, the variance of the level-of-change estimator, \hat{Q}_0 , is considered:

$$\text{Var}\{\hat{Q}_0\} = \text{E}\{\hat{Q}_0^2\} - \text{E}^2\{\hat{Q}_0\}. \quad (3.10)$$

In order to examine (3.10) more closely, write \hat{Q}_0^2 as a fourth power of \hat{l}_1 and \hat{l}_2 .

$$\begin{aligned} \hat{Q}_0^2 &= ((\hat{l}_1 - \bar{l}) - (\hat{l}_2 - \bar{l}))^4, \\ &= (\hat{l}_1 - \bar{l})^4 - 4(\hat{l}_1 - \bar{l})^3(\hat{l}_2 - \bar{l}) + 6(\hat{l}_1 - \bar{l})^2(\hat{l}_2 - \bar{l})^2 - 4(\hat{l}_1 - \bar{l})(\hat{l}_2 - \bar{l})^3 + (\hat{l}_2 - \bar{l})^4. \end{aligned} \quad (3.11)$$

In order to examine the second and fourth terms in (3.11), we take their expected value, noting $\text{E}\{\hat{l}_1 - \bar{l}\} = \text{E}\{\hat{l}_2 - \bar{l}\} = 0$, so by independence,

$$\text{E}\{(\hat{l}_1 - \bar{l})^3(\hat{l}_2 - \bar{l})\} = \text{E}\{(\hat{l}_1 - \bar{l})^3\} \text{E}\{(\hat{l}_2 - \bar{l})\} = 0, \quad (3.12)$$

and the expectation of the second and fourth terms in (3.11) is zero. For the expected value of the third term, we have

$$\begin{aligned} \text{E}\{(\hat{l}_1 - \bar{l})^2(\hat{l}_2 - \bar{l})^2\} &= \text{E}\{(\hat{l}_1 - \bar{l})^2\} \text{E}\{(\hat{l}_2 - \bar{l})^2\} \\ &= \text{E}^2\{(\hat{l} - \bar{l})^2\}. \end{aligned} \quad (3.13)$$

Taking the expected value and grouping terms of the level-of-change estimator, we have

$$\text{E}\{\hat{Q}_0^2\} = 2 \text{E}\{(\hat{l} - \bar{l})^4\} + 6 \text{E}^2\{(\hat{l} - \bar{l})^2\}. \quad (3.14)$$

From (3.10) and (3.14), we obtain an expression for the variance of the level-of-change estimator:

$$\begin{aligned} \text{Var}\{\hat{Q}_0\} &= 2 \text{E}\{(\hat{l}_1 - \bar{l})^4\} + 6 \text{E}^2\{(\hat{l}_1 - \bar{l})^2\} - \text{E}\{\hat{Q}_0\}^2, \\ &= 2 \text{E}\{(\hat{l}_1 - \bar{l})^4\} + 6 \text{E}^2\{(\hat{l}_1 - \bar{l})^2\} - 4 \text{E}^2\{(\hat{l}_1 - \bar{l})^2\}, \\ &= 2 \text{E}\{(\hat{l}_1 - \bar{l})^4\} + 2 \text{E}^2\{(\hat{l}_1 - \bar{l})^2\}. \end{aligned} \quad (3.15)$$

The expressions $\text{E}\{(\hat{l} - \bar{l})^2\}$ and $\text{E}\{(\hat{l} - \bar{l})^4\}$, the second and fourth central moments, can be expressed in terms of the second- and fourth-order cumulants of the distribution of $\ln \hat{S}$ (Stuart and Ord, 2010, pp. 88–89), represented as κ_2 and κ_4 as follows:

$$\text{E}\{(\hat{l} - \bar{l})^2\} = \kappa_2, \text{ and} \quad (3.16)$$

$$\text{E}\{(\hat{l} - \bar{l})^4\} = \kappa_4 + 3 \kappa_2^2. \quad (3.17)$$

The cumulants of $\ln \hat{S}$, where \hat{S} has chi-squared distribution with $2K$ degrees-of-freedom, in (3.16) and (3.17), can be expressed in terms of polygamma functions (Bartlett and Kendall, 1946, p. 128)—specifically,

$$\kappa_{r+1} = \psi^{(r)}\left(\frac{n}{2}\right), \quad (r > 0). \quad (3.18)$$

In (3.18), $\psi^{(r)}$ is a polygamma function of order r , and n represents the degrees-of-freedom from \hat{S} . A polygamma function when $r = 1$ in (3.18) is called a trigamma function and it is represented as $\psi'(\frac{n}{2})$. This allows the mean and variance of \hat{Q}_0 , (3.9) and (3.15), to be written in terms of cumulants and ultimately of polygamma functions. Specifically,

$$\begin{aligned} \text{E}\{\hat{Q}_0\} &= 2 \kappa_2 \\ &= 2\psi'(K) \end{aligned} \quad (3.19)$$

and

$$\begin{aligned}
 \text{Var}\{\hat{Q}_0\} &= 2(\kappa_4 + 3\kappa_2^2) + 2\kappa_2^2 \\
 &= 2\kappa_4 + 8\kappa_2^2 \\
 &= 2\psi^{(3)}(K) + 8[\psi'(K)]^2.
 \end{aligned} \tag{3.20}$$

The polygamma functions of order one and three (when $r = 1$ and 3 in (3.18)) have the following series expansions, which hold as $\alpha \rightarrow \infty$ (Abramowitz and Stegun, 1965, p. 260):

$$\psi'(\alpha) = \frac{1}{\alpha} + \frac{1}{2\alpha^2} + \frac{1}{6\alpha^3} - \frac{1}{30\alpha^5} + \frac{1}{42\alpha^7} - \frac{1}{30\alpha^9} + \dots, \tag{3.21}$$

and

$$\psi^{(3)}(\alpha) = \frac{2}{\alpha^3} + \frac{3}{\alpha^4} + \frac{2}{\alpha^5} - \frac{1}{\alpha^7} + \frac{4}{3\alpha^9} - \frac{3}{\alpha^{11}} + \frac{10}{\alpha^{13}} - \dots \tag{3.22}$$

In practice, the numerically optimized algorithm presented in Amos (1983), which uses a different series expansion and a recursion step, is used.

3.5.3 Correlation of Spectra from Adjacent Blocks

The discussion in Section 3.5.2 makes the assumption of independent spectrum estimates, which does not exactly hold. It has been shown to hold asymptotically as the distance between the sections on which the two spectral estimates are made grows infinitely large (Brillinger, 2001, p. 130). However, this work deals with adjacent sections or slightly overlapping sections.

As a partial justification of the independence assumption we present the following observations. Correlations between spectrum estimates made on different blocks using

a single tapered estimate including overlapping blocks are discussed in Thomson (1977a, pp. 1790–1791), and essentially the correlation is found to be small even when the blocks have a moderate overlap (Welch, 1967b).

Multitaper spectral estimates are functions of complex-valued eigencoeficients which are approximately normally distributed. Providing these eigencoeficients are uncorrelated, and assuming the complex-valued eigencoeficients are approximately jointly normally distributed, then the eigencoeficients are approximately independent. If the eigencoeficients are independent, then multitaper spectral estimates, which are functions of the eigencoeficients, are also independent. This concludes the partial justification into relaxing the independence assumption, and we move on to studying the level-of-change estimator using simulations.

3.6 Simulation Study of Estimator

3.6.1 Models with No Change-points

3.6.1.1 Normal Distribution

The first simulation studies the case of no change-points in independent data. This example is designed to produce the mean and variance values derived in Section 3.5.2. Random samples from a $N(0, 1)$ distribution are generated, the level-of-change estimator is calculated, and the sample mean and the sample variances are observed. This is done using sample sizes of 2048, 4096, and 8192 and block sizes of 128, 256, and 512, respectively. The number 128 is a power of two that is close to an appropriate block size for analyzing the GHD series, and larger sample sizes will be required to show convergence in some simulation examples. Ten thousand realizations

of each sample size are generated. For each realization, multitaper spectrograms³ with 16 time blocks are constructed, and then the level-of-change estimator between adjacent blocks is formed. The set of 4000 simulations was run with four different time-bandwidth parameters, $NW = 2, 3, 4$ and 5 , each using $2NW - 1$ tapers. Figure 3.1 shows a sample 16-block multitaper spectrogram constructed of blocklength 128 from a $N(0, 1)$ sample of length 2048 using multitaper parameters $NW = 5$ and $K = 9$. This spectrogram represents a matrix that is the first step in obtaining the level-of-change estimator. Figure 3.2 shows the associated level-of-change estimator, and it provides a pictorial representation of the between-block-pair level-of-change. Block pair 1 in Figure 3.2 represents the level-of-change between blocks 1 and 2 in Figure 3.1, block pair 2 represents the level-of-change between blocks 2 and 3, and so forth. The frequencies within W of zeroth and Nyquist (0.5) are dropped from the level-of-change estimator. All multitaper spectrograms are plotted on a logarithmic colour scale. The values on the scale indicate power, and are in units²/frequency, where “units” indicates the units of the original variable.

All simulations presented in this thesis use code in R based on the multitaper software package introduced in Appendix A, which includes optimized Fortran 90 code. The simulations require modifications not included in the package, and the simulations are performed using the R “parallel” software package, which allow multiprocessor use provided that the code is appropriately designed. The set of simulations presented in this Normal Distribution subsection takes approximately 10 hours on an Intel “Core 2” 2.50 GHz quad core processor running a Linux Mint operating system when all four cores are utilized.

³One spectrogram constructed from periodogram estimates is included, and this estimate can be considered a single tapered spectral estimate with a constant taper.

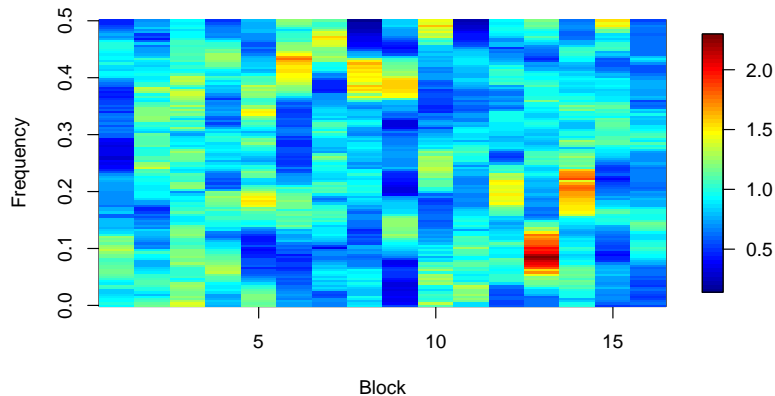


Figure 3.1: Multitaper spectrogram plot with adaptive weighting of white noise data using 16 non-overlapped blocks of length 128—that is, the total length is $N = 16 \times 128 = 2048$. The multitaper spectral estimates use the parameters $NW = 5$ and $K = 9$.

The primary object of the first set of simulations is to check the derived mean and variance of the level-of-change estimator against simulated values, and Figures 3.1 and 3.2 are presented to provide an example of the procedure. Matrices represented by the two figures are from one realization of the simulation, each matrix is generated for each of the 4000 realizations, and the 4000 simulations are run for each sample size, and for each of the five spectral estimators, which includes the four time-bandwidth parameters and the constant taper (periodogram). Figure 3.2 plots the level-of-change estimator. The matrix represents a level-of-change value for each frequency, and for each pair of adjacent block. We initially set a cutoff of 4.16 (level-of-change ≥ 4.16) as an initial way of consider detecting a change-point in this example. Our simulations indicate such a cutoff leads to a $\sim 5\%$ false detect rate (Type I error). Note that 4.16 represents a value above 12σ where σ^2 is the variance based on the

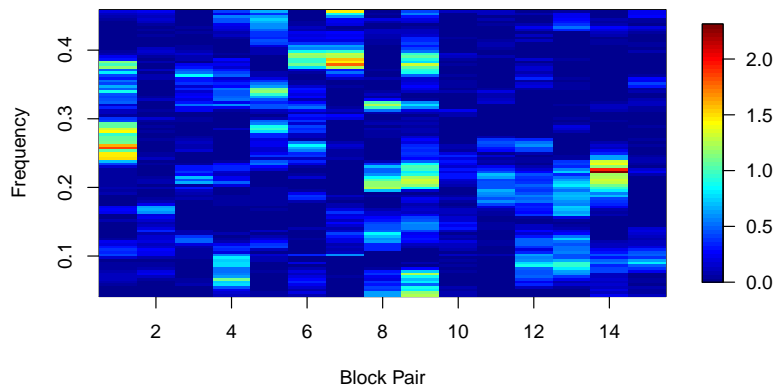


Figure 3.2: Level-of-change estimator between block pairs based on the spectrogram of white noise shown in Figure 3.1, using multitaper parameters $NW = 5$, and $K = 9$. The frequency range is reduced as we omit frequencies within W of the zeroth and Nyquist (0.5). In this example, we use a cutoff value of 4.16, giving a 5% error rate for the complete matrix, and this matrix exhibits no change-points.

chi-squared degrees-of-freedom presented in Table 3.2. This high cutoff is the result of the multiple hypothesis tests in the matrix represented by Figure 3.2, which has over 1600 values that are not independent. A discussion of selecting a cutoff is presented in Section 3.6.3. We do not propose this as a standalone estimator but rather as a tool to be used with other existing tools (see Section 3.6.2).

The simulation means of the level-of-change estimator over all blocks and frequencies for each time-bandwidth pair are shown in Table 3.1. Additionally, the average of the level-of-change estimator obtained from a periodogram is shown. Table 3.2 shows the average simulation variances of the level-of-change estimator. The tables show close agreement between observed simulation means and variances and theoretical

values derived in Section 3.5.2. Table 3.2 also shows how using multitaper spectral estimation considerably reduces the variance associated with the level-of-change estimator as a result of the degrees-of-freedom increase.

Block Size	Level-of-change Simulated Means			
	$NW = 2$	$NW = 3$	$NW = 4$	$NW = 5$
128	0.7985	0.4446	0.3082	0.2357
256	0.7993	0.4447	0.3081	0.2356
512	0.7994	0.4449	0.3083	0.2358
Theory	0.7899	0.4426	0.3071	0.2350

Table 3.1: Random samples of spectral means were generated of size 2048, 4096, 8192, then multitaper adaptively weighted block spectrograms were constructed by using block lengths of 128, 256, and 512 respectively. The table gives sample means found using simulation for the periodogram and nonadaptive weighted multitaper spectral estimates with time-bandwidth parameters, $NW = 2, 3, 4$ and 5.

Block Size	Level-of-change Simulated Variances			
	$NW = 2$	$NW = 3$	$NW = 4$	$NW = 5$
128	1.6156	0.4432	0.2055	0.1181
256	1.6268	0.4429	0.2052	0.1178
512	1.6199	0.4438	0.2055	0.1180
Theory	1.4857	0.4347	0.2030	0.1169

Table 3.2: Variances of random samples were generated and multitaper spectrograms constructed as in Table 3.2. Observed sample variances constructed using adaptive weighting are higher than both theoretical variances and simulated variances constructed from multitaper spectrograms without adaptive weighting.

Note that the level-of-change estimator was plotted with adaptive weighting, which lowers the associated degrees of freedom; however, there is no visible difference in the estimator with and without adaptive weighting in this example. Tables 3.1 and 3.2 were constructed using adaptive weighting and while the associated degrees of freedom is lower without adaptive weighting, in the Gaussian noise example the difference between the estimator with and without adaptive weighting is confined to the third

and fourth decimal place. We do not find adaptive weighting of use in these first simple examples; however, in Section 3.6.1.4, a stationary no change-point example in which adaptive weighting is of visible benefit to the level-of-change estimator will be presented. This section demonstrates that random $N(0, 1)$ simulations using block sizes as small as 128 closely match the values derived in Section 3.5.2.

3.6.1.2 Normal Cubed Distribution

Next we consider another independent no change-point example; however, in this case data are drawn from an $N(0, 1)^3$ distribution, which has a high variance. We expect that it represents an extreme case of independent data on which to test our level-of-change estimator. Note that if σ^2 is the variance for a Gaussian process, $N(0, \sigma^2)$, then $15\sigma^6$ is the variance for the cube of the Gaussian process. The kurtosis is 46.2^4 indicating that the distribution has heavy tails, a lack of shoulders, and high peak. As in the previous example, 4000 samples were generated, but this time the block sizes and sample sizes were greatly increased. Sample sizes of 2048, 4096, 8192, 16384, 32768, 65536, 131072, and 262144, and block sizes of 128, 256, 512, 1024, 2048, 4096, 8192, and 16384 respectively. From this section forward, this work studies the level-of-change estimator constructed only from multitaper spectral estimates in simulation models.

The large sample sizes were required as the simulation sample mean and sample variance converge slowly to the derived values and only with large block sizes, indicating that the spectral estimator converges to an approximate chi-squared distribution with $2K$ degrees of freedom. This is expected because of the heavy tails, and we that the shape of the distribution be determined before attempting the level-of-change

⁴The standardized fourth moment prior to subtracting 3, as is done with relative kurtosis.

estimator. The multitaper spectral estimator assumes that the data is locally smooth in order to justify the χ^2_{2K} distribution of spectral estimate, and the locally smooth condition is increasingly satisfied with larger sample sizes when NW is unchanged. Thomson (1982, pp. 1062–1065) suggests that a non-central chi-squared distribution may be more appropriate when the locally smooth assumption is violated.

Tables 3.3 and 3.4, which are constructed in the same way as the tables in Section 3.6.1.1, present the observed sample means and sample variances averaged over 4000 realizations. From the perspective of the level-of-change estimator, the high variance in the cubed Gaussian distribution requires impractically large block sizes to converge in mean and variance to the derived values. These tables indicate that the proposed level-of-change estimator can be effective with high-variance non-Gaussian data sets, but only with extremely large samples. Next, dependent data samples are considered. The variances are high with sample sizes we expect to see in climate data, and we cannot propose use of the level-of-change estimator on data with a $N(0, 1)^3$ distribution unless the sample sizes are very large, at least 16384 samples, and we expect a 634 samples in a long series.

Block Size	Level-of-change Simulated Means			
	$NW = 2$	$NW = 3$	$NW = 4$	$NW = 5$
128	1.2460	0.9160	0.7884	0.7200
256	1.0342	0.6980	0.5672	0.4979
512	0.9192	0.5787	0.4462	0.3758
1024	0.8560	0.5127	0.3790	0.3081
2048	0.8242	0.4790	0.3444	0.2728
4096	0.8072	0.4610	0.3260	0.2542
8192	0.7986	0.4519	0.3166	0.2542
16384	0.7944	0.4474	0.3120	0.2400
Theory	0.7899	0.4426	0.3071	0.2350

Table 3.3: Sample means of the level-of-change estimator from an $N(0, 1)^3$ distribution. 4000-run simulations were made, each having 16 blocks in length. The bottom row gives the approximations derived in Section 3.5.2.

Block Size	Level-of-change Simulated Variances			
	$NW = 2$	$NW = 3$	$NW = 4$	$NW = 5$
128	3.3518	1.7377	1.2754	1.0621
256	2.3842	1.0253	0.6667	0.5117
512	1.9304	0.7151	0.4153	0.2918
1024	1.7034	0.5692	0.3025	0.1972
2048	1.5964	0.5017	0.2516	0.1555
4096	1.5409	0.4680	0.2270	0.1357
8192	1.5134	0.4512	0.2149	0.1357
16384	1.5000	0.4432	0.2090	0.1216
Theory	1.4857	0.4347	0.2030	0.1169

Table 3.4: Sample variances from simulated level-of-change estimator from an $N(0, 1)^3$ distribution. 4000-run simulations were made, each having 16 blocks in length. The bottom row gives the approximations derived in Section 3.5.2.

3.6.1.3 AR(2) Model

In Section 3.5.3, a discussion for relaxing dependence was made, and we present two dependent data models without change-points. The first is an AR(2) model with coefficients $\phi = (0.75, -0.5)^T$, a model that has been used in the literature as a simple dependent data model (P&W93, p. 45). Four thousand simulations indicate that the sample means and sample variances averaged across blocks and frequencies are close to those derived in Section 3.5.2. Admittedly, in this case the dependence between adjacent blocks is low.

As in Section 3.6.1.1, sample sizes of 2048, 4096, and 8192 and block sizes of 128, 256, and 512 respectively are used in 4000 run simulations. Figure 3.3 shows a realization of the multitaper spectrogram of the process. In a sufficiently long data set, the dependence structure in the data is seen as being of higher power between frequencies 0.1 and 0.2 (P&W93, p. 309); however, in the short blocks, the power in this frequency range varies from high to low, as the sample size is not sufficiently long to capture the signal in each block. Figure 3.4 shows the level-of-change estimator for the same realization, and differences between blocks resulting from the AR(2) structure are hard to observe in a single realization. If, as previously, a cutoff of 4.16 is selected, this realization will not be significant.⁵ A plot of the average level-of-change estimator across all 4000 realizations, not presented here, shows that a higher level-of-change is observed between frequencies of 1.8 and 2.6, representing the down slope of the peak resulting from the AR(2) process, which is not well resolved with smaller sample sizes (Thomson, 2001, p. 349). A similar pattern is observed in a plot of the standard errors over the 4000 realizations, also not presented here.

⁵Note that simulations show this AR(2) model will have a $\sim 7.6\%$ false detect rate with a cutoff of 4.16 which was selected from white noise simulations.

Tables similar to Tables 3.1 and 3.2, not shown, here indicate that the simulation values of sample mean and sample variance are close to those derived in Section 3.5.2, indicating some robustness to dependence. We propose use of a spectral estimate of the entire series along with this level-of-change estimator, as the spectral estimate of the complete series would aid in assessing a reason for any increased level-of-change.

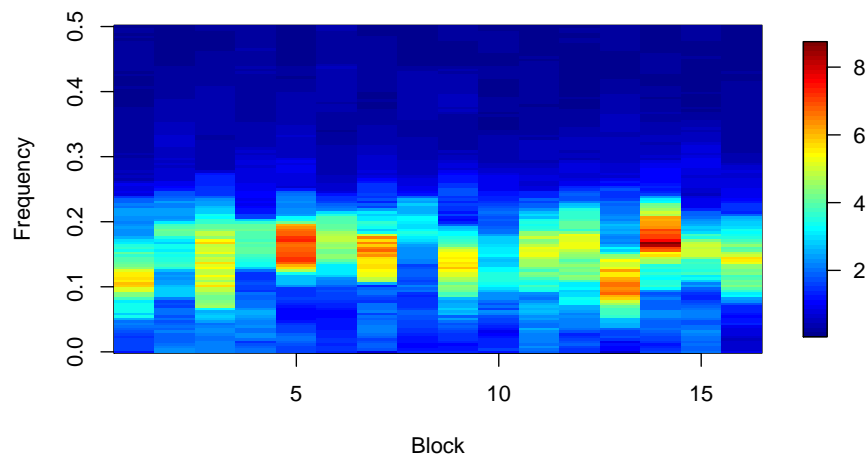


Figure 3.3: Multitaper spectrogram of a realization of the stationary AR(2) process. Each block is 128 samples long, and the multitaper parameters used are $NW = 5$ and $K = 9$. The spectral estimate in each block is not well resolved with 128 sample block sizes.

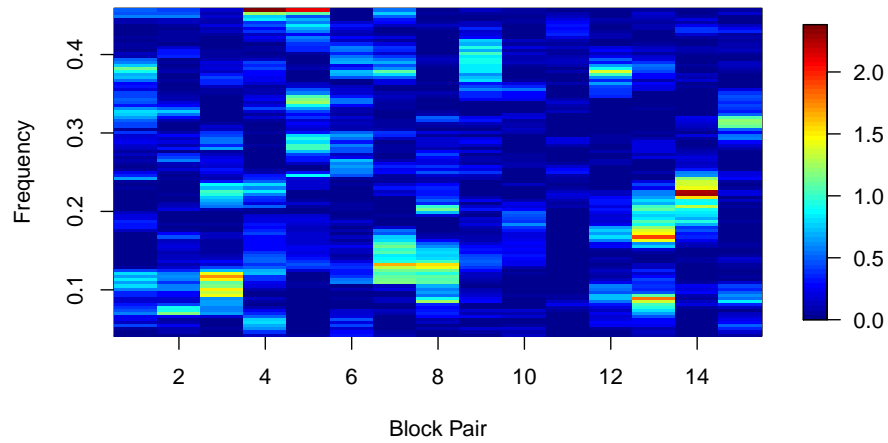


Figure 3.4: Level-of-change estimator for the AR(2) example shown in Figure 3.3. This realization indicates the potential for a false detect. This risk can be reduced by recognizing a higher likelihood of false detect around the unresolved peak in the spectrum.

3.6.1.4 ARMA (4,2) Model

The next no-change-point model we simulate is an ARMA(4,2) process. Such models have been used in previous studies (Kay, 1980), and in our experience it simulates data with a higher dynamic range than is seen in climate time series. As in previous examples, sample sizes of 2048, 4096 and 8192 and block sizes of 128, 256, and 512 respectively are used. A higher power is expected between the frequencies 1.8 and 2.4; however, as a result of the short block sizes, the spectral peaks are not consistently resolved, and we observe variance in power in that frequency range. Figure 3.5 shows a realization of the adaptively weighted multitaper spectrogram of the process. Looking at the complete spectral estimate one sees the two close peaks which are not consistently and fully resolved in the spectrogram. In this case the image of the spectrogram, with the between block variance, is consistent with the autoregressive moving average (ARMA)(4,2) process. Note that there was no visible difference between spectrograms with and without adaptive weights, not shown; however, the level-of-change estimator appears more stable with adaptive weighting. Figure 3.7 plots the level-of-change estimator constructed without adaptive weights, and Figure 3.8 plots the estimator of the same realization constructed using adaptive weights. The image without adaptive weighting is considerably more noisy, and the level-of-change with adaptive weights indicates a higher level-of-change when no change is present. Thomson (1982, p. 1065) introduced adaptive weighting to address the problem of higher-order eigenspectra estimates being unreliable in regions where the spectrum is small; the bias characteristics or higher-order tapers decrease. Adaptive weighting essentially down-weights higher-order eigenspectra, and this has a visible benefit in the level-of-change plot for this process. We propose using the

figure with adaptive weights, as the false detect is associated with the power range where there is a frequency fluctuation in the spectrogram. The power fluctuation is a random pattern resulting from high and lower resolution in the short blocks; however, a simple spectral estimate of the entire length of the series will make it clear whether such a signal is present. Figure 3.6 plots the spectral estimate for all 2048 samples used in the spectrogram, and the two peaks are readily apparent. It is the two peaks which are not well resolved and creating the colour pattern in Figure 3.5.

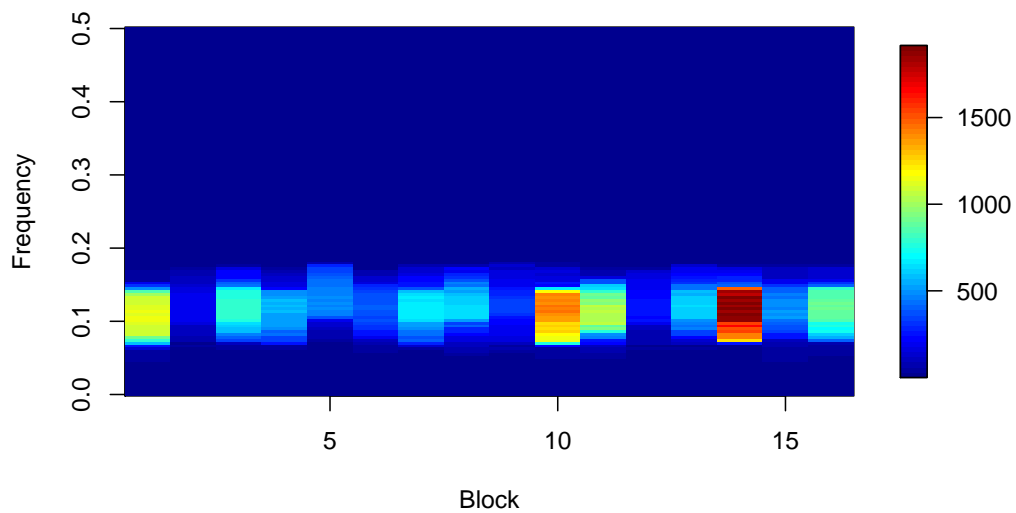


Figure 3.5: Multitaper adaptively weighted spectrogram of a realization of the ARMA(4,2) process. Each block is 128 samples long, and the multitaper parameters used are $NW = 5$ and $K = 9$.

As with previous estimators, we select a default cutoff of 4.16; however, in this case there is high likelihood of false detect around the unresolved spectral peak in each block. To address this, Figure 3.9 displays only values that are above the cutoff.

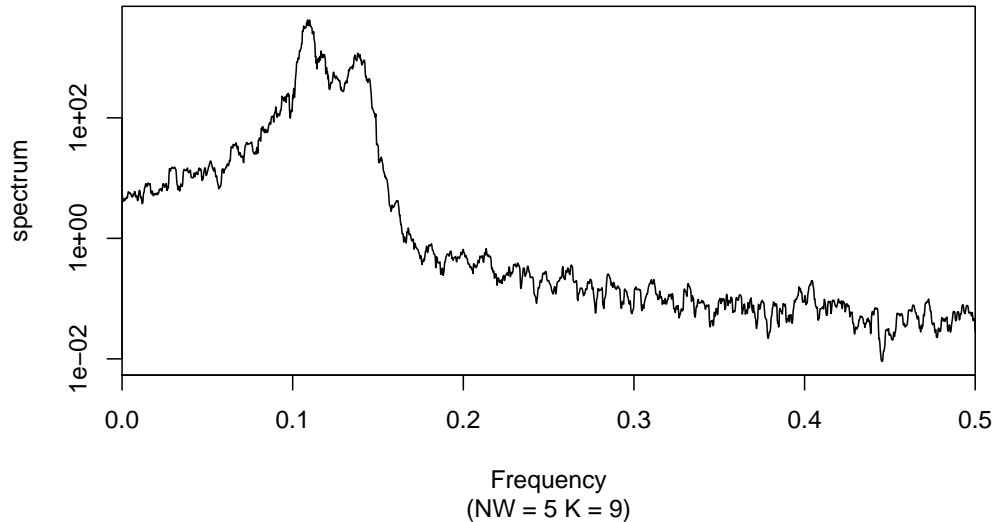


Figure 3.6: Multitaper adaptively weighted spectrum estimate all 2048 samples from the same realization of the ARMA(4,2) process using multitaper parameters $NW = 5$ and $K = 9$.

We see that in this realization, all values above the cutoff are in the vicinity of the unresolved spectral peak and, as such, can be recognized as a false detect. In general, AR and ARMA processes have lower spectral values than white noise, with similar mean and variance, in the higher frequencies. This artificial example can appear as non stationary on the Bartlett M test, and can produce false detects using the proposed level-of-change method for detecting change points; however, we submit, if one plots a complete spectral estimate, and recognizes the spectral peaks are not resolved in short blocks, then inappropriately classifying such a process as nonstationary can be avoided. We also recommend fitting an ARMA model and plotting the residuals as a standard diagnostic. These techniques should help avoid missclassification.

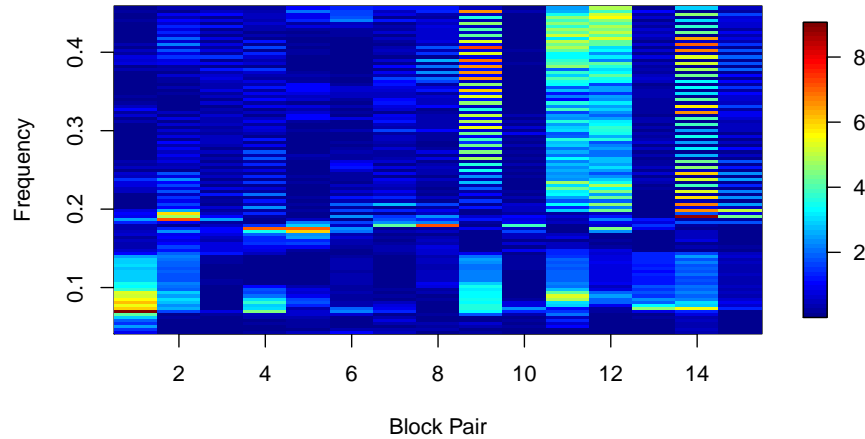


Figure 3.7: Level-of-change estimator, constructed without adaptive weights, $NW = 5$, and $K = 9$, for the ARMA(4,2) example shown in Figure 3.5. This plot has h

As in the previous examples, Tables 3.5 and 3.6 show the average over the sample mean matrix, and the average over the sample variance matrix respectively, for $NW = 2, 3, 4$ and 5 , with $K = 2NW - 1$. The tables show the values of sample mean and standard errors not close to those derived in Section 3.5.2 at a sample size of 128; however, the values approach the derived values as the sample size doubles. This indicates that with smaller block sizes, as are likely to be seen in the GHD series, the mean and variances may not equal the derived values. The values presented in the tables are constructed with adaptive weighting, and when adaptive weighting is not used, the means and variances are closer to theoretical values with fewer tapers; however, when a higher number of tapers are used, the values are actually closer to theoretical values with adaptive weighting. Tables of values without adaptive weighting are not presented.

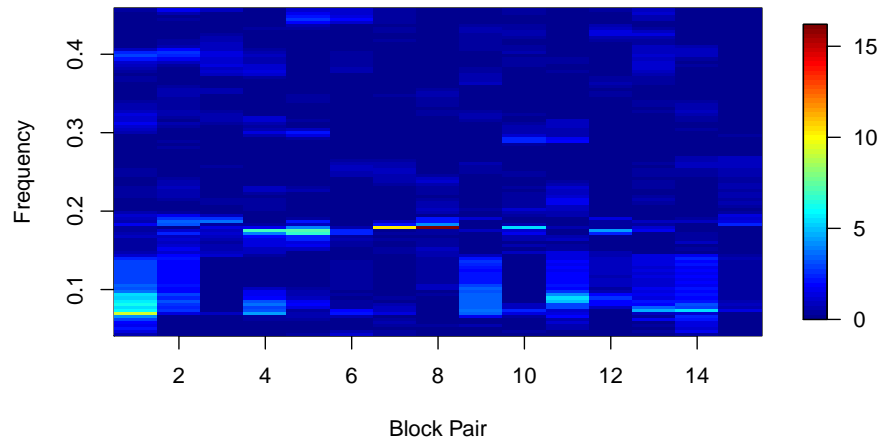


Figure 3.8: Level-of-change estimator, constructed using adaptive weights, with $NW = 5$ and $K = 9$ for the ARMA(4,2) example shown in Figure 3.5. This image is less noisy than the one without adaptive weights. This plot has some high valued false detects which require further examination.

In summary, we have shown the level-of-change estimator in several no-change-point examples, two of which are nontrivial, and we have found that the estimator can be made to work with dependent data, and that adaptive weighting is preferable. We have given an example of a strict cutoff and shown that is feasible in these examples. We do not propose that the level-of-change estimator be used as a stand-alone tool in detecting change-points; in fact, is more likely to be effective when used in conjunction with standard spectral estimation techniques.

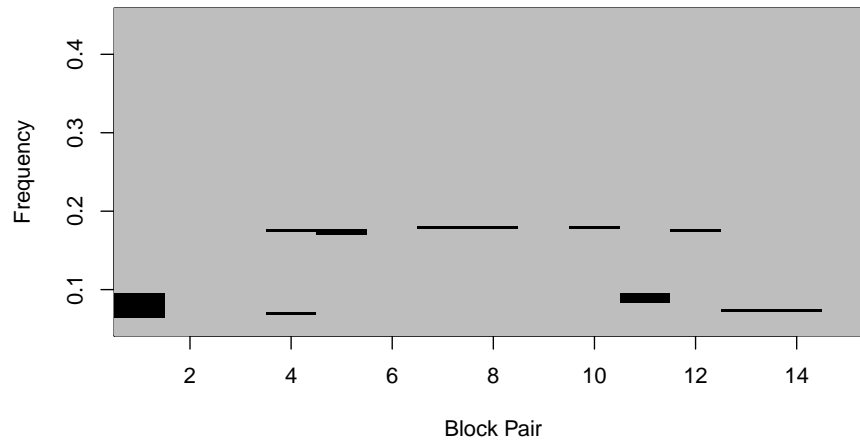


Figure 3.9: Level-of-change estimator plot showing only values above the 4.16 cutoff constructed using adaptive weights, with $NW = 5$, and $K = 9$ for the ARMA(4,2) example shown in Figure 3.5. The only detected values are in a region where false detects are expected due to the low resolution of each block.

Block Size	Level-of-change Simulated Means			
	$NW = 2$	$NW = 3$	$NW = 4$	$NW = 5$
128	2.3585	0.9071	0.5865	0.4789
256	2.3214	0.8430	0.5019	0.3782
512	2.3069	0.8134	0.4644	0.3357
Theory	0.7899	0.4426	0.3071	0.2350

Table 3.5: Average across blocks and frequencies of the standard error matrix of the level-of-change estimator, using adaptive weights, with $NW = 5$, and $K = 9$, from 4000 simulations of the ARMA(4,2) process.

Block Size	Level-of-change Simulated Variances			
	$NW = 2$	$NW = 3$	$NW = 4$	$NW = 5$
128	20.0399	2.7091	0.9036	0.6595
256	19.8702	2.4743	0.6134	0.3482
512	19.8713	2.3702	0.5193	0.2541
Theory	1.4857	0.4347	0.2030	0.1169

Table 3.6: Average across blocks and frequencies of the standard sample mean of the level-of-change estimator, using adaptive weights, $NW = 5$ and $K = 9$, from 4000 simulations of the ARMA(4,2) process.

3.6.2 Change in Frequency Model

We next explore the practicality of this estimator using a simulation model where there is a change-point specifically, we simulate a series where a frequency component changes. This is the type of structural change we focus on detecting in the GHD data. We study a simplified version of the models described in Lees and Park (1995). A single time series is constructed by concatenating the following two models:

$$x_1(n) = \cos(2\pi n/11) + 0.6 \cos(2\pi n/4.8) + w_t, \tag{3.23}$$

and

$$x_2(n) = 0.2 \cos(2\pi n/19) + 0.6 \cos(2\pi n/4.8) + w_t, \tag{3.24}$$

where w_t is random noise drawn from an $N(0, 1)$ distribution. The model is:

$$x(n) = \begin{cases} x_1(n) & \text{if } n \leq n_0 \\ x_2(n) & \text{if } n > n_0. \end{cases} \tag{3.25}$$

We simulate x_1 and x_2 , each of length 1024,⁶ and concatenate the series for $x(n)$ with $x(n)$ having 2048 points and being indexed $n = 1, 2, \dots, 2048$. This is equivalent to the smallest simulation size in the stationary examples, and each realization will have 16 blocks of length 128.

Figure 3.10 plots one realization of the spectrogram from this example. In the figure, the parameters $NW = 5$ and $K = 9$ are used, and changing the time-bandwidth parameter and number of tapers affects the appearance of the spectrogram. In general, decreasing the time-bandwidth can help to resolve signals better, given the resulting shorter time blocks sizes; however, increasing time-bandwidth and the associated number of tapers increases the degrees of freedom of the spectral estimate

⁶We set n_0 in (3.25) to 2048.

and thus lowers the variance in the level-of-change estimator (Thomson, 1982). The vertical bar in the first block, at approximately $f = 0.38$, indicates the bandwidth parameter $2W$ used in the spectral estimate. In this figure, the harmonic component at 0.2 is visible throughout the spectrogram, but it is not well resolved in each block; the sinusoid at frequency 0.09 is much better resolved for the first half of the series where it is present, and the lowest-amplitude sinusoid, at 0.0526, is barely distinguishable from noise in the second half of the series.

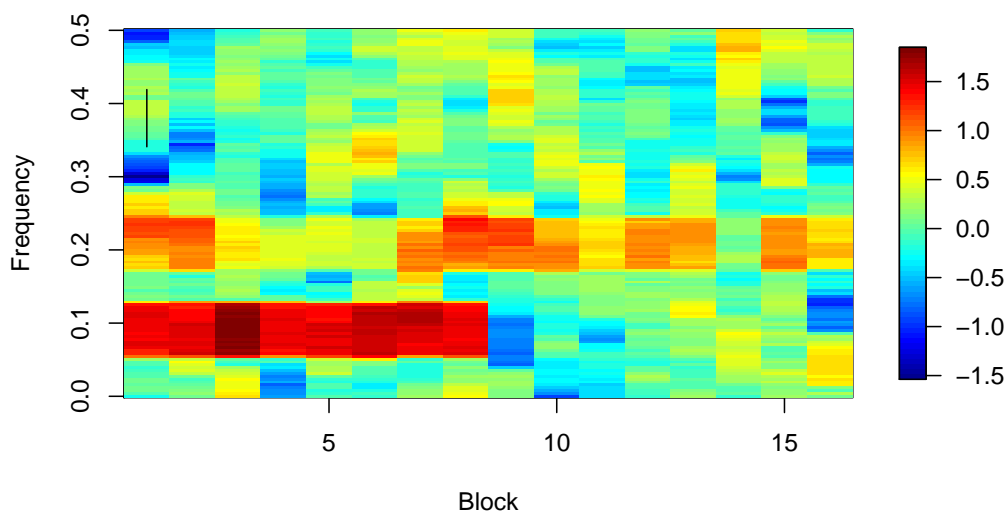


Figure 3.10: Multitaper spectrogram plot of simulated data containing two sinusoidal frequencies, with one that considerably damps down at the halfway point. In this case the nonstationarity is clearly visible in the spectrogram. The black line segment in the upper left indicates the bandwidth, $2W$. The first half of the data has a sinusoid of amplitude $A_{1a} = 1$ at $f_{1a} = .09$, and a sinusoid of amplitude $A_2 = 0.6$ at $f_2 = 0.2$. The second half has a sinusoid of amplitude $A_{1b} = 0.2$ at $f_{1b} \approx 0.0526$. The background noise has constant variance of one. The multitaper parameters used were $NW = 5$, and $K = 9$. The ≈ 0.0526 low-amplitude frequency is not distinguishable at this block length.

Figure 3.11 plots the level-of-change matrix from the spectrogram using adaptive weighting and multitaper parameters $NW = 5$ and $K = 9$. Once again, other multitaper parameters and block lengths were attempted; and this plot demonstrates the level of change in this realization of the process. In this example, if we select a cutoff of 4.15, we can detect the change.

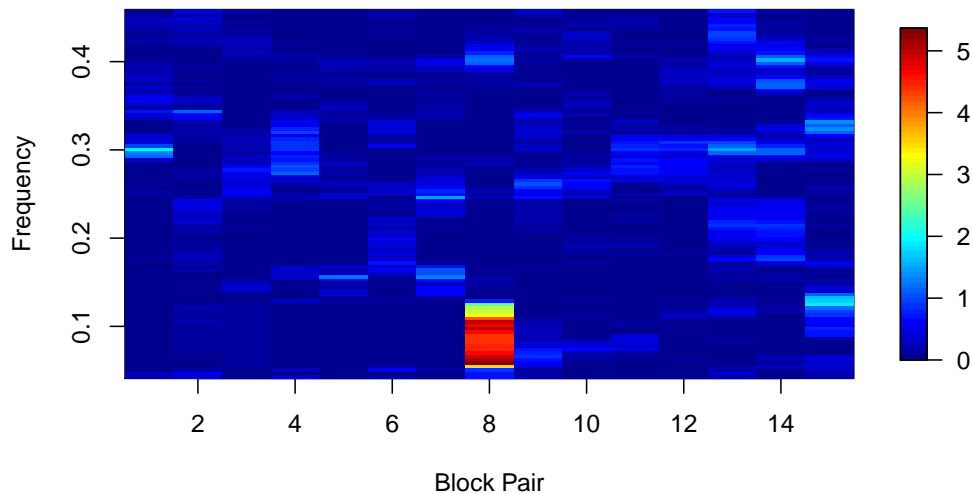


Figure 3.11: We plot the level-of-change estimator between adjacent blocks, trimming the blocks by w at the frequency edges (zero and Nyquist frequencies). Note that we visually detect a level-of-change estimator between blocks 8 and 9 at a frequency of approximately 0.091 ($1/11$).

The above example is not randomly selected from the set of the 4000 simulations containing the change-point; instead, it is randomly selected from the set of simulations with values above the 4.16 cutoff in the correct frequency range and block. We estimate that approximately 25% of the 4000 simulations containing a change-point

as specified in (3.25) have such a value. The selected cutoff gives low statistical power; however, simulations demonstrate that cutoffs selected to control Type I error across the entire frequency range on single hypothesis tests, such as the harmonic F -test, also have low statistical power. This tells us that while we can select such a cutoff to help in reading the level-of-change estimator matrix in simulating simple examples, such a cutoff is not feasible in practice. We do not propose the level-of-change estimator as a stand-alone tool with a strict cutoff set to control Type I error across the whole matrix, but we propose that it to be incorporated with other existing tools to help in detecting change-points.

One tool that this estimator should be used with is the Bartlett M-test; Figure 3.12 plots the Bartlett M-test for this example, and it clearly shows non-stationarity at approximately $f = 0.2$. This is a tool that can aid in identifying which frequencies to pay attention to when attempting to detect a change-point.

Figure 3.13 plots the average values of the level-of-change estimator for the eighth block pair column, the column for the block pair which contains the change point, over the 4000 simulations using $NW = 5$ and $K = 9$. The plot indicates that the level-of-change estimator has a mean value that is considerably higher in the appropriate frequency range, while frequency where there is no change has values close to the expected values. This pattern is similar when other multitaper parameters are used. A plot of the standard errors, not shown, is similar.

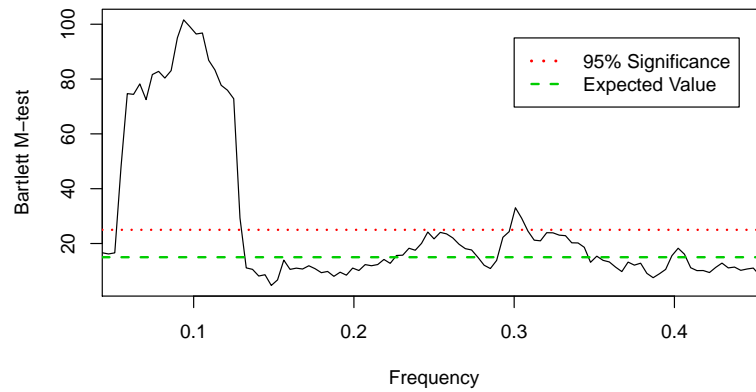


Figure 3.12: Bartlett M-test for this change-point example. This test shows non-stationarity at the frequency where there is a change in amplitude and change in frequency. The line segment in the below the legend indicates the bandwidth, $2W$, and the two dashed lines indicate the chi-squared expected value and the 95% value. The multitaper parameters used were $NW = 5$, and $K = 9$.

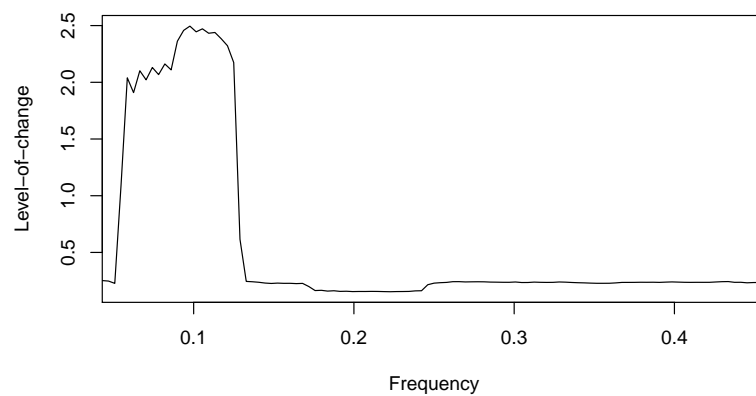


Figure 3.13: Average eighth block pair level-of-change column over 4000 simulations. This figure shows that the average observed level-of-change over the 4000 simulations is considerably higher in the frequency range where the change-point occurs. The multitaper parameters used were $NW = 5$ and $K = 9$.

3.6.3 Cutoff Values for Level-of-Change

The cutoff value of 4.16 was obtained to control Type I error over the entire level-of-change matrix; the value sets Type I error at 5%. The cutoff was selected from simulations of the $N(0,1)$ process described in Section 3.6.1.1 for multitaper parameters $NW = 5$ and $K = 9$, and it has a high Type II error in the frequency change-point model introduced in Section 3.6.2. We do not propose a general use of such cutoff values with this method, and on the basis of simulations not presented here, we observed that a similarly obtained cutoff would lead to low power in tests such as ubiquitous harmonic F -test. Table 3.7 shows the cutoffs obtained for the parameters $NW = 2, 3, 4$ and 5 and $K = 3, 5, 7$ and 9 . One can see that the cutoffs increase as the degrees of freedom decrease.

Level-of-change Cutoffs			
$NW = 2$	$NW = 3$	$NW = 4$	$NW = 5$
24.50	9.46	5.82	4.16

Table 3.7: Cutoffs for controlling Type I error for the level-of-change estimator based on maximum values in each level-of-change matrix for 4000 simulation and a $N(0,1)$ process.

Figure 3.14 shows a density plot comparing maximum values from change points to maximum values of no-change-points. This figure is presented to point out that selecting a cutoff in order to control both Type I and Type II error is not a trivial task for the change-point model discussed in Section 3.6.2. However, such problems are truly difficult and are generally not solved for other spectral analysis tests; for example a similar problem exists for the harmonic F -test. Once again we propose using this estimator as part of a suite of tools to help detect change-points.

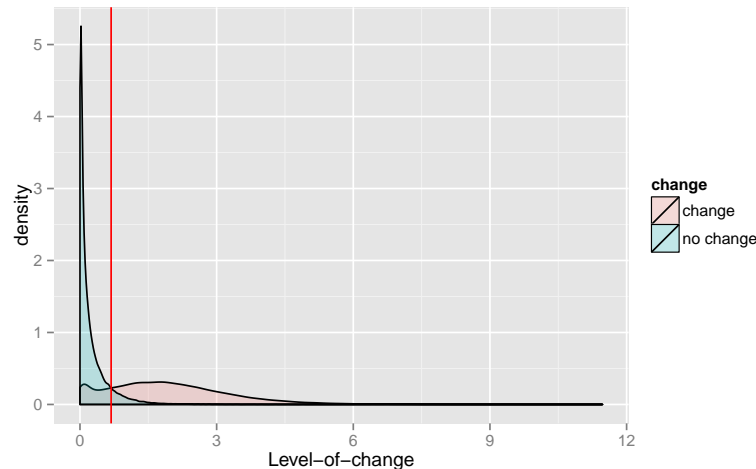


Figure 3.14: Plots of densities of the level-of-change estimator for a model with a change-point and a model without. These are based on 4000 simulations comparing maximum values of a model with a change-point to one without. The intersection point is 0.68.

3.6.4 Computational Issues

The level-of-change estimator requires calculation block spectral estimates. This is implemented using multiple fast Fourier transforms, each of which are order $\mathcal{O}(N \log_2 N)$, where \mathcal{O} is used as “big O” notation for the order of operations. In calculating each multitaper spectral estimate, k Fourier transforms are used, and generation of the Slepians is done using a traditional formulation, which is $\mathcal{O}(N)$. The same set of Slepians can be used over the entire block spectral estimate. In practice, we use both popular optimized Fourier transforms which perform better when N is a power of two, and we propose increasing the sample size to a power of two.

3.6.4.1 Original Example

In practice, we are faced with a fixed sample size. For example, the grape harvest date (GHD) (Chuine et al., 2004) time series, examined in Chapter 4, has 634 annual samples, from 1370 to 2003, and obtaining more samples is not feasible, except perhaps for the most recent decade. While one can look to other proxy measures, such as the ^{14}C Bristlcone Pine Record (Suess and Linick, 1990), we chose to work with the existing data and consider the case of overlapped block sizes. In selecting block size, we have to consider the power of the signal we are trying to resolve, the acceptable amount of overlap, and whether we are willing to discard data at the edges or between blocks. As demonstrated earlier in this chapter, block size selection affects the spectrogram, the Bartlett M-test results, and the level-of-change estimator. In this case we may have to consider various levels of overlap, lost data at the ends of the series, or lost data between blocks. When there are few samples, we prefer to use all the data. The procedure that we adopt is to test all reasonable estimates of overlapped and omitted block lengths in an appropriate range, and try to find an appropriate compromise. Table 3.8 gives a sample of possible block size choices.

Block Length	Block Offset	Number of Blocks	% Overlap
109	105	6	3.7
106	88	7	17.0
81	79	8	2.5
82	69	9	15.9
84	55	11	34.5
106	48	12	54.7
128	46	12	64.1

Table 3.8: A sample of potential block sizes, selected by using the criterion that data at the end points not be discarded. In general, when the offset size is small, the options for block size increase, and the trade-off occurs when block size and offset are close and thus minimizing the overlap.

3.7 Suggested Methodology

The level-of-change (frequency-domain change-point) estimator is not a stand-alone tool, and in practice it should be used with other tools such as graphical aids to help with exploratory data analysis, with the aim of determining whether a change-point specifically characterized as a change in process structure exists. We propose its use in conjunction with the Bartlett M-test, the multitaper spectral estimate, the multitaper spectrogram, and other standard statistical tools. As a general rule, multiple parameters for block length and overlap should be explored for the initial multitaper spectral estimate in addition to multiple bandwidth parameters. We propose a methodology for incorporation and use of this tool; however, we acknowledge that

Exploratory data analysis is still partly [largely] an art so, for a given time series, several approaches are possible. (Thomson, D. J., pers. comm.)

We propose the following general approach when using the proposed level-of-change estimator to locate change-points:

1. Plot the spectrum with and without several prewhitening models and ensure that a simple ARMA model is not sufficient.
2. Plot the spectrogram with various levels of overlap and time-bandwidth parameters, while being careful to recognize what an unresolved signal looks like and ensure that no possible change-point can be explained that way.
3. Use the Bartlett M-test for stationarity to determine whether it is possible to isolate nonstationarity to a specific set of frequencies.

4. Examine a plot of level-of-change estimator, consider several cutoff values and compare this plot with the Bartlett M-test.
5. Plot the spectrum before and after the change-point on the same scale and determine if it is possible to identify what is going on.
6. If there is uncertainty about the selection a change point, assign a probability based on the null white noise model, or another model noting that the significance under this assumption may not always be appropriate.

3.8 Summary and Comments

In this chapter we introduced a level-of-change estimator to examine the dynamic frequency spectrum of a time series to look for change in frequency-domain structure. The estimator focuses specifically on changes in spectral structure. We noted that spectral estimates have a chi-squared distribution and derived the mean and variance for this estimator assuming independence of spectral estimates. We proposed using this estimator with the multitaper spectral estimate, which has higher chi-squared degrees of freedom than other spectral estimates. We demonstrated that the derived mean and variance are correct and that they hold with various underlying dependent and independent spectral models. We demonstrated the estimator on a change-point example and proposed a methodology for using the estimator.

Future applications of this technique could include automated methods for model-specific change-point detection, possibly studied with simulations. The spectral estimator is based on a chi-squared distributed spectrum, which holds for a single spectral estimate; however, we use this estimate on a spectrogram which has many correlated

spectral estimates that are not independent, and it has been suggested that a non-central chi-squared distribution is more appropriate in such matrices. Further work is required to study the properties of the estimator as a part of a matrix containing many (over 1600) non-independent points.

Chapter 4

Burgundy Grape Harvest Dates

4.1 Introduction

This chapter presents an analysis of the Burgundy grape harvest date (GHD) series first assembled by Chuine et al. (2004).¹ This series is of particular interest because, starting in 1370, it is the longest climate time series available that has known dates and continues into the present. Other similar time series, proxy data for climate, have uncertain dates; the time (or date) is estimated. Thus this series can be used both to calibrate dates of proxy series such as ice cores and to compare pre-industrial and industrialized European climate. Natural climate variability and its impact on ecosystems and plant phenology have been discussed in Jones and Goodrich (2007), and this series is considered to track climate variability accurately. We note that there are concerns about production practices and socioeconomic pressures resulting in artificially low within-year variability (Chabin et al., 2007). Burgundy represents 18, regions and the capital, Dijon, regularly mandated the harvest date for the entire region, resulting in artificially low within-year variability. We are specifically interested in the annual climate signal captured by the harvest date.

In addition to the Burgundy GHD series, the following long-term time series have been produced for the region of interest: a Swiss GHD series from 1480 CE to 2006 (Meier et al., 2007) and a 335-year Central England Temperature (CET) series (Manley, 1974). In the Burgundy GHD series, harvest dates correlate negatively with April to August temperatures, September temperatures do not correlate significantly, and the overall relationship is dominated by interannual variability (Chuine et al., 2004; Krieger et al., 2011). Two additional series are used in coherence estimates: (1) an estimate of total solar irradiance (TSI) from Stocker et al. (2013) and

¹The series consists of harvest dates taken from multiple sites, but harvest dates were often selected by the central authority and imposed on all sites.

based on Krivova et al. (2010) and Vieira et al. (2012), and (2) an El Niño—southern oscillation (ENSO) reconstruction based on Wilson et al. (2010). TSI can be thought of as a reconstruction of solar brightness, and ENSO represents a record of anomalous sea surface temperatures that are known to affect climate.

The primary GHD data set provides the longest series. The median, over the 18 regions, of the standardized dates is reported as day after September 1st. A multitaper spectral analysis of the Burgundy GHD series for years 1678 to 2003 is presented in Tourre et al. (2011). This is the most recent published analysis of this series, and they did not consider the entire series.

A four-stage analysis is presented: (1) Compare the Burgundy GHD series with other series to determine the magnitude-squared coherence (MSC) and phase coherence, (2) perform analysis of the complete series, (3) use the methodology discussed in Chapter 3 to locate one change-point, and (4) perform multitaper spectral analysis of the (two) sections. The novel contributions are the coherence study, which indicates that the series captures a climate signal, the location of the change-point using the methodology introduced in Chapter 3, and the multitaper spectral analysis of the complete and sectioned GHD time series.

4.2 Initial Analysis

We begin by plotting the Burgundy GHD series along with five other similar series for comparison. Figure 4.1a plots the original series, and Figure 4.1b plots the Swiss GHD as days after September 1st (Meier et al., 2007). Note that there are several large gaps in the first part of the Swiss series. Figure 4.1c plots the CET annual series (Manley, 1974), Figure 4.1d plots the annual phase of the CET series in (angular) degrees,

which was calculated using the monthly temperature series (see Appendix A.6.1), Figure 4.1e plots the estimated TSI in watts per square metre (Stocker et al., 2013), and Figure 4.1f plots three reconstructions of the ENSO cycle shown in normalized degrees Celsius. Note that the CPR, composite plus regression, which relies on simple averaging of the proxy series, and PCR, principal component regression, reconstructions of the ENSO cycle appear almost identical.

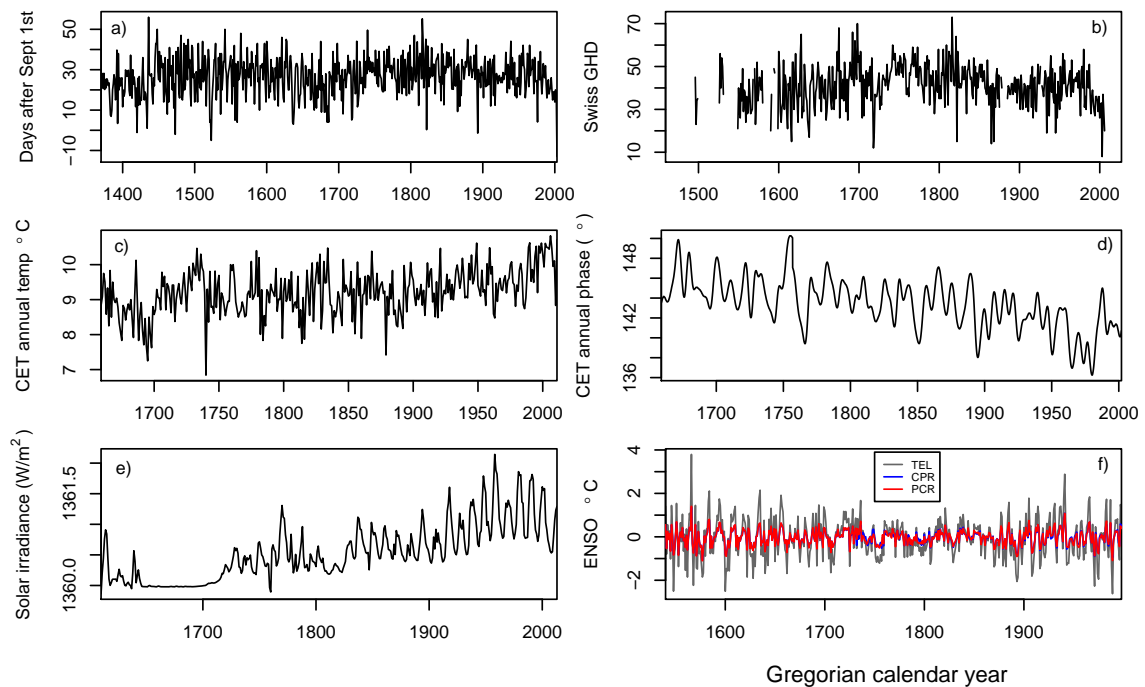


Figure 4.1: (a) Burgundy GHD plotted as number of days after September 1st. Five additional series are also shown: (b) Swiss GHD as days after September 1st. There are several large gaps in the first part of this series. (c) CET annual temperature series. (d) Annual phase of the CET series in (angular) degrees. (e) Estimated TSI in watts per square metre. (f) Three reconstructions of the ENSO cycle shown in normalized degrees Celsius.

Figure 4.2 gives multitaper spectral estimates of the GHD series. The crosses at approximately 0.135 cycles/year indicate the passband bandwidth, $2W$, and height

of the approximate theoretical 95% confidence interval based on the χ_{2k}^2 distribution. The confidence interval uses the median number of degrees-of-freedom from the adaptively weighted spectral estimate, which is generally slightly below $2k$. Mean and trend are also removed from the GHD series prior to taking the multitaper spectral estimate. One can detrend with simple mean and variance or smoothing splines; however, we opt to detrend using an expansion into the Slepian sequences employed in Thomson (1982).² We use different time-bandwidth parameters and observe how the plots change. This gives us a sense of whether we are over- or under-smoothing—we smooth by selecting the bandwidth or time-bandwidth parameter. The spectra are plotted on a logarithmic scale and the units of the y-axis are (Days after Sept 1)²/(cycles/year). The observed pattern is consistent with that seen in the climate literature. Increasing the bandwidth by increasing NW results in increasing the passband, and this can potentially smooth out frequency components of interest. For example, the top left plot has a small peak at approximately 1/11 cycles per year, corresponding to a solar cycle, but this component is not large enough to be statistically significant and is smoothed out when the passband, W , increases. To avoid missing lines, the series was zero-padded to 8192. Figure 4.3 presents a harmonic F -test of the entire 600 year series; we note that the peak at a period of 3.9 years is slightly less than the period 4.14 years found in Tourre et al. (2011), and we attribute this shift to including the entire GHD series.

In this section, we present a multitaper spectral analysis of the complete Burgundy GHD series, and we find a significant harmonic component at 3.9 years which represents a slight shift from 4.14 year period found in the multitaper analysis of

²In our experience, there is little observable difference between different methods of detrending, and this method is specifically adapted to multitaper spectral estimates.

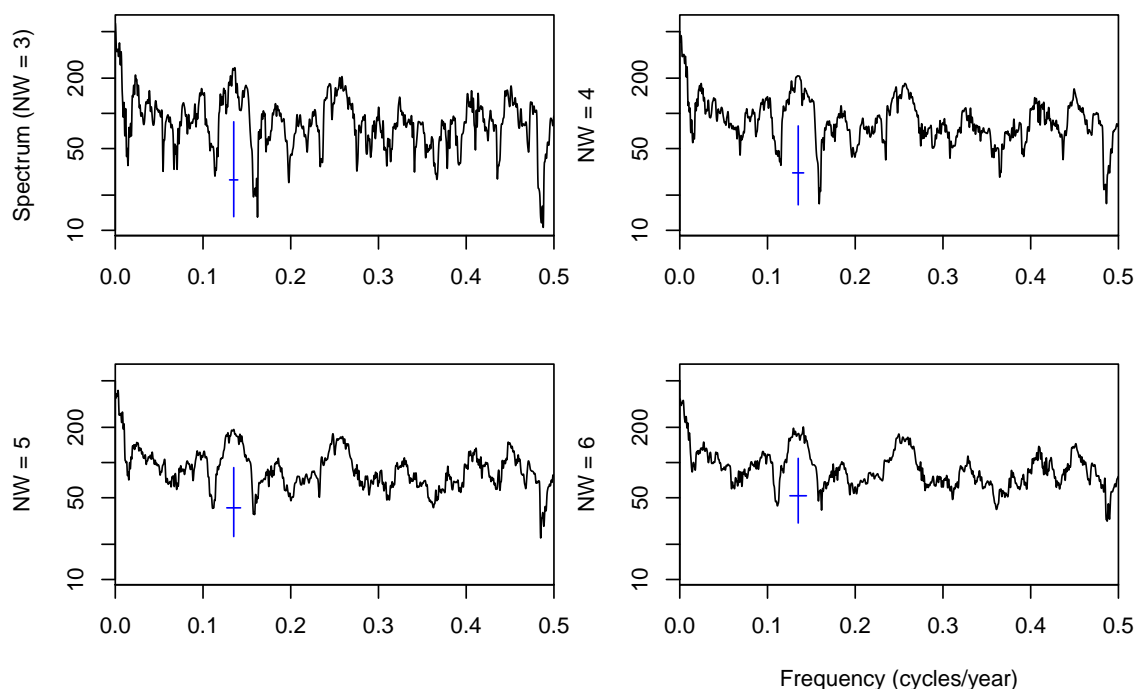


Figure 4.2: Multitaper spectra of GHD series. Multitaper spectral estimates were made with $NW = 3, 4, 5$ and 6 , and with $K = 5, 7, 9$ and 10 , respectively starting at the top left. The crosses at approximately 0.135 cycles/year indicates the passband bandwidth, $2W$, and height of the approximate theoretical 95% confidence interval based on the χ^2_{2k} distribution. Note that the peak at a period of 3.9 years almost agrees with Tourre et al. (2011).

years 1678–2003 in Tourre et al. (2011, p. 247). The plots also demonstrate the effect of increasing parameters NW and K . The change-point detection method discussed in Section 3.7 is suited to a higher NW and K , but a more descriptive plot can be seen with lower values for NW and K .

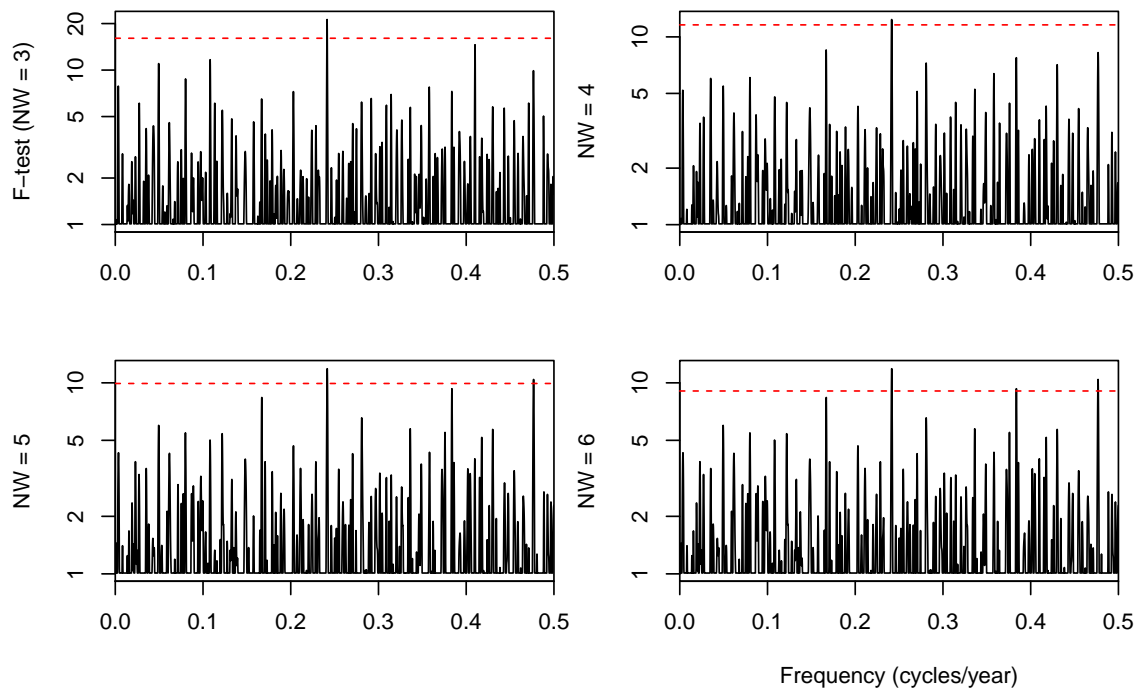


Figure 4.3: This figure shows the harmonic F -test statistic for the harvest dates. The parameter values used are $NW = 3, 4, 5$ and 6 , with $K = 5, 7, 9$ and 10 respectively. The red dashed line indicates a $1 - 1/N$ level of significance where $N = 634$, in keeping with the rule of thumb for the harmonic F -test (see Section 3.6.1). We note that the most significant peak occurs at a period of 4.14 years, which is close to the reported period of 3.9 years reported in Tourre et al. (2011, p. 247).

4.2.1 Data Quality

In this section we present several spectral coherence comparisons obtained using the multitaper method to correct for bias (see Section 2.13). We do this to examine whether the GHD series has properties similar to those of other related data sets, using MSC and phase coherence (see Section 2.13). We do not prewhiten the series with AR filters prior to examining coherence, but we do remove mean and trend, as is customary, prior to working with spectral estimates. We note that MSC has a slight but well-calibrated bias, and we use a multitaper estimate to reduce the bias (Thomson and Chave, 1991a). Expected values for the moments of the MSC are given in equation (5) from Carter et al. (1973b), and from this the expected value of the MSC for independent data is $1/K$ where K is the number of independent tapers used in a multitaper MSC estimate.³

Figure 4.4 plots the Burgundy and Swiss GHD series, presented in Meier et al. (2007), for the overlapping dates. On average, the Swiss harvest date lags the Burgundy harvest date by 14 days. Figure 4.5 indicates the MSC between the Burgundy GHD series and the Swiss series, and the MSC indicates a high coherence. The dashed red line in Figure 4.5 indicates the known bias in the estimate; specifically, in this case it shows that a coherence of 0.14 will be observed for estimated values of uncorrelated samples. The dashed blue lines indicate MSC of 0.393 and 0.534, which correspond to a significance of 95% and 99% respectively.⁴ The observed MSC is considerably above the expected value for uncorrelated examples, and we consider

³Prior to plotting the coherence plots, we plot either a single plot with both series or two adjacent plots with both series. It is customary to plot them on the same page where possible, and we opt for plotting them in sequence.

⁴These significance values are based on the normal transformation from Thomson and Chave (1991b), and normality was assessed prior to assigning significance.

this as evidence of high data quality—i.e., the records are relatively accurate—and both data sets capture similar climate variability. The faint dashed line on the coherence plot represents the lower one standard deviation jackknife confidence interval. Note that the confidence interval is approximately one standard deviation based on the inverse hyperbolic tangent transformation (the scale of the y-axis). Coherence estimates are jackknifed as other multitaper spectral estimators (see Section 2.12, page 27). Many of the details in Figure 4.4 track; however, the low frequencies often depart for decades. This agrees with the coherence plot, Figure 4.5, where the MSC is decreased at low frequencies. We consider this an interesting result; both series have similar signals with periods of 11 to 2 years. Figure 4.6 plots the phase

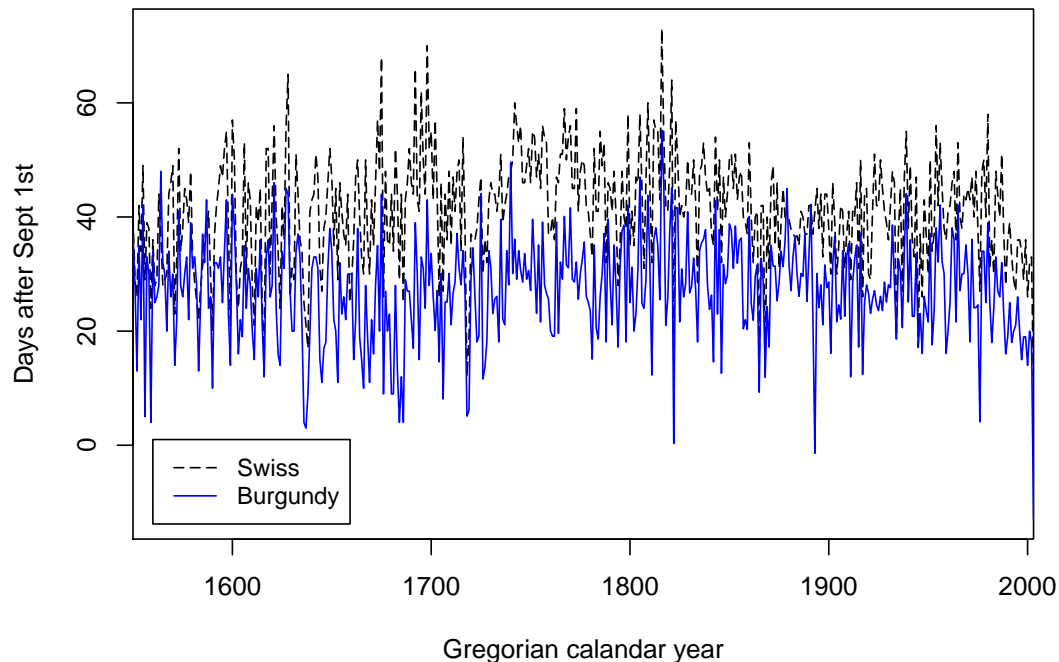


Figure 4.4: Overlapping section of the Swiss and Burgundy GHD series consisting of years 1550 to 2003; no prewhitening has been applied, and MSC is presented in the next plot. We note that the Swiss harvest is on average ~ 14 days after the Burgundy harvest.

coherence—note that the slope is relatively flat, once again indicating both series are affected by similar periodic fluctuations. The nine-day lag is slightly less than that captured in the average series lag, and we subtract the mean prior to computing the spectral estimates. The delay captured here corresponds to change in year-to-year fluctuations.

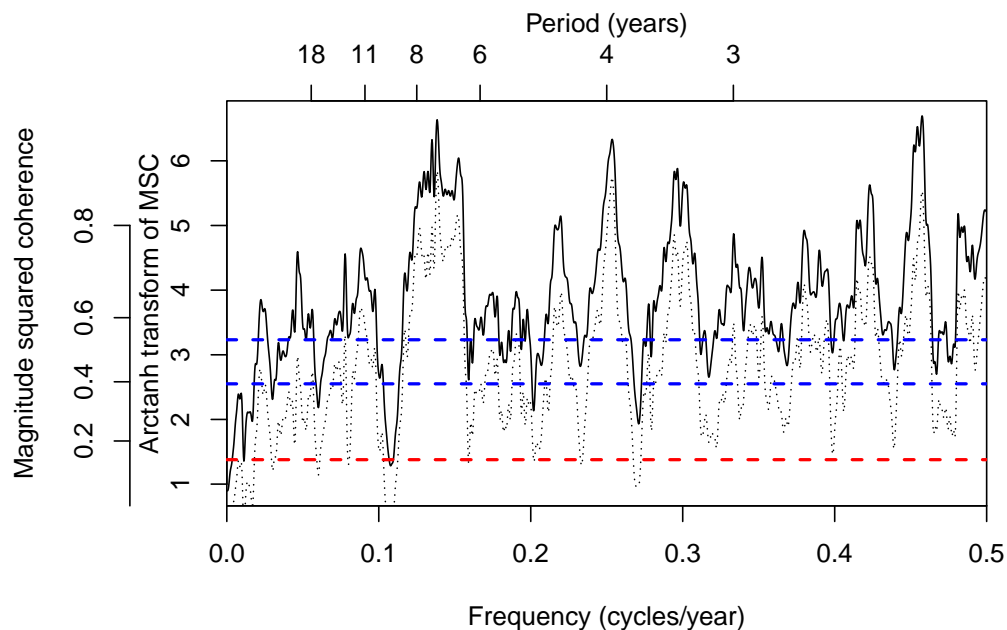


Figure 4.5: MSC between Swiss and Burgundy GHDs. The coherence is constructed from overlapped years 1550 to 2003 and is based on the multitaper spectral estimates with parameters $NW = 4$ and $K = 7$. The y-axis indicates a normalized MSC; a hyperbolic inverse tangent transform is known to transform the MSC to a standard normal distribution (Thomson and Chave, 1991b). The dashed red line indicates the inherent bias in the estimate; specifically, it shows that a coherence of 0.14 will be observed for estimated values of uncorrelated samples. The faint dashed line on the coherence plot represents the lower of a one standard deviation jackknife confidence interval. The two dashed blue lines indicate a significance of 95% and 99%, corresponding to an MSC of 0.39 and 0.54 respectively.

The high coherence between the two series represents an incidental but significant

result of this thesis. This result provides evidence that two separate GHD series in different countries were subject to similar cyclical climate effects between years 1550 and 2003.

The next coherence study compares the Burgundy GHD series with the annual value of the CET series. The CET monthly series is shorter, and we compare the years 1661 to 2003. Figure 4.7 plots the two series for observation. Figure 4.8 plots the MSC of the two series. Once again the dashed red line indicates the bias inherent on the estimate; uncorrelated series are expected to have an MSC of 0.09. The two dashed blue lines indicate an MSC of 0.173 and 0.211 corresponding to a significance of 85% and 95% respectively, and the faint grey line indicates a one standard deviation jackknife confidence interval. Once again the MSC is unusually high, providing evidence of similar cyclical climate effects in both the Burgundy GHD and the CET series.

Next we examine the coherence between the Burgundy GHD and the annual phase of Central England temperature series, this phase was originally presented in Thomson (1995), and is reproduced in Appendix A.6.1. Figure 4.10 plots the two series next to each other, and then Figure 4.11 plots the MSC. It indicates the coherence between the two series is modest, especially at low frequencies.⁵ The MSC is not consistently and considerably above the bias value as the other two are, yet still there is some evidence of coherence in certain frequency ranges. The MSC values of 0.17 and 0.21, represent significance values of 85% and 95% respectively. Finally, Figure 4.12 plots the phase of the coherence. In this case a slope is apparent; the phase is not flat. In this case the slope corresponds to a delay of approximately 305

⁵The GHD series is annual and we compute the coherence estimate directly between the harvest date, in days after Sept. 1st, and the *annual phase* of the Central England temperature series calculated from three years of monthly temperatures.

days, the annual phase of the Central England temperature series lags the Burgundy GHD by approximately 305 days. We cannot explain this last observed delay and further study is required.

In these coherence estimates three phenomena are evident. First, low and high frequencies often show distinct patterns. The “low frequencies” up to ~ 0.16 cycles/year (approximately 6 years in period) often show long term solar phenomena (Suess and Gleissberg cycles, in addition to the ordinary 11-year solar cycle). It may be coincidence, but the average decay time of a sunspot cycle, where most of the large flux occurs, is 6.3 years. Second, at high frequencies a linear phase trend is often visible and has typical slopes corresponding to a few days, about the time required for ordinary weather patterns to drift across the continent. This illustrates one of the advantages of coherences; one may often reliably detect time delays of a few days in series with annual sampling. Third, the MSC often alternates between high values at periodic climate cycles and at low values between them.

The trend in Figure 4.6 is not as obvious with a larger y-range. It is possible that there is a slight trend visible between 0.16 and 0.5 cycle/years. We note that a frequency of 0.16 cycles per year corresponds to a ~ 6 year period, approximately the shortest cycle in the sunspot record. The trend corresponds to a delay of ~ 3 days, possibly the propagation time for ordinary weather patterns between Burgundy and Switzerland.

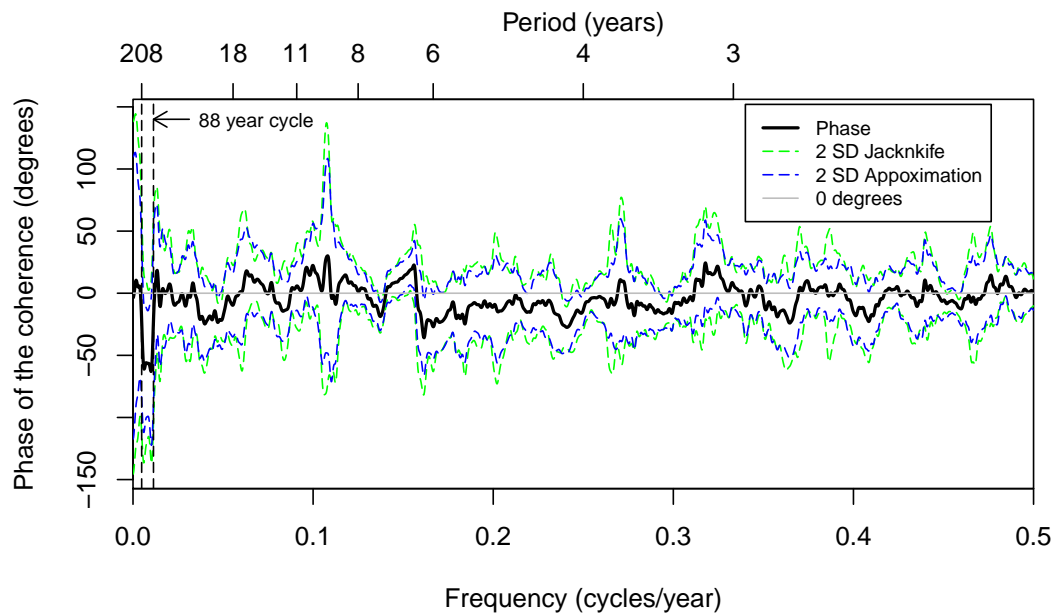


Figure 4.6: Phase coherence between Burgundy and Swiss GHDs. Coherence is defined in (2.67) and based on the multitaper cross-spectrum in (2.68). In these equations the Burgundy series is represented by x and the Swiss series is represented by y . Two standard deviation confidence intervals are indicated on the plots; the green line represents multitaper jackknife confidence intervals, and the blue line represents approximate theoretical confidence intervals (Bendat and Piersol, 2011, p. 306). It may be observed that these agree well. The phase is generally consistent with zero, excluding the low-frequency part, and no phase unwrapping was required. Between periods of ~ 208 and ~ 90 years there is a sharp drop to -69 degrees. Both edge frequencies are well known in the climate literature: 208 years is one of the main “Suess cycles” (Thomson, 1990b), and 90 years is very close to the upper peak, 91.5 years, of the ~ 88 year Gleissberg cycle triplet (Peristyk and Damon, 2003). The linear regression line (in grey) has a negative intercept and a positive slope. This indicates that the Swiss series leads the Burgundy series by ~ 9 days.

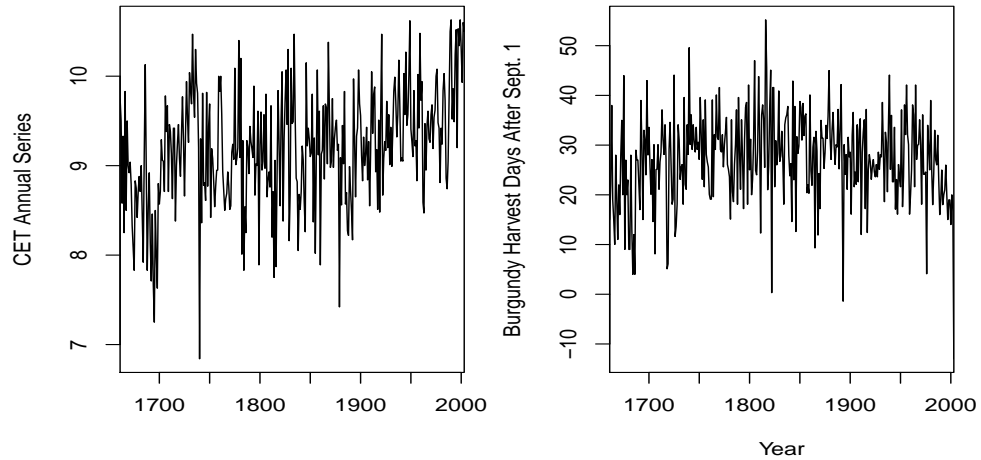


Figure 4.7: Plots of the Burgundy GHD and the CET annual series for overlapping years 1661 to 2003.

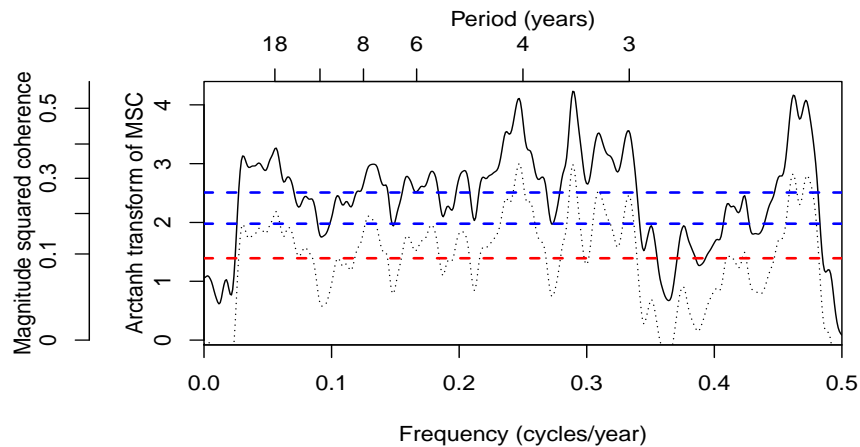


Figure 4.8: MSC between Central England average annual temperature and the Burgundy harvest dates from 1661 to 2003. The parameters used are $NW = 6.5$ $K = 11$. The dashed red line indicates the bias value of 0.09, and the dashed blue lines indicate MSC of 0.173 and 0.201. The coherence is modest, particularly at low frequencies. The association between GHD and April to August temperatures in Burgundy have been established (Chuine et al., 2004; Krieger et al., 2011).

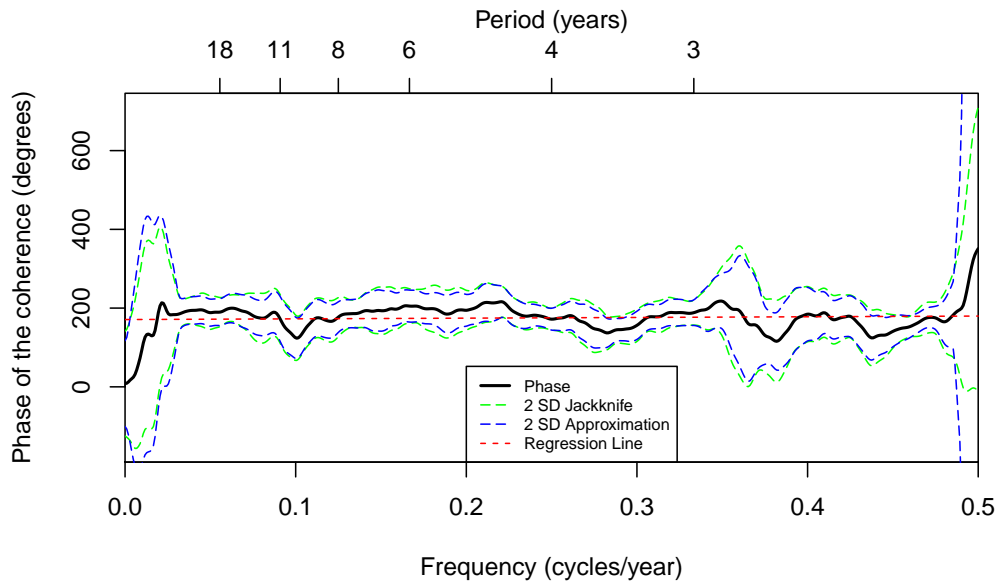


Figure 4.9: Phase coherence between the Burgundy GHD and the average annual temperature of Central England series and for years 1661 to 2003. This figure is based on (2.67) with the Burgundy series is represented by x and the Central England series is represented by y . The multitaper parameters are: $NW = 6.5$, $K = 11$. The linear regression line (in red) has a positive intercept and a positive slope. This indicates that the Burgundy series leads the Central England series by ~ 18 days.

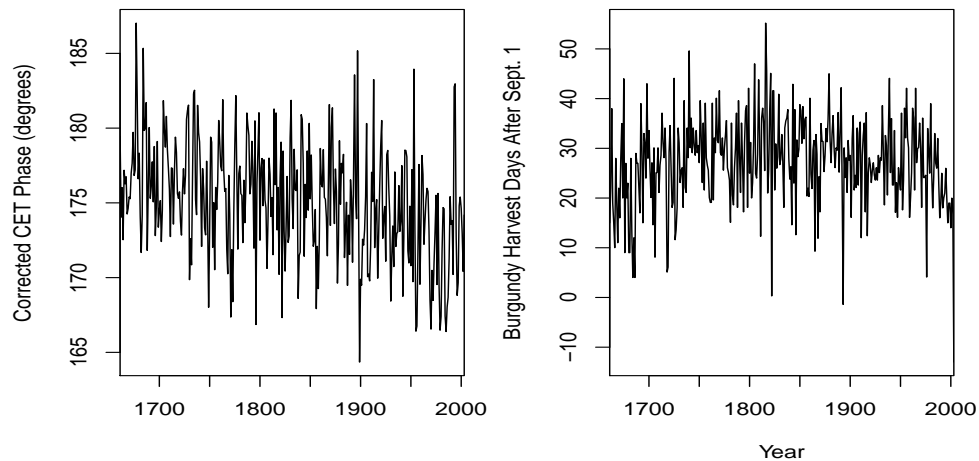


Figure 4.10: Plot of the Burgundy GHD Series and the Central England phase constructed from three years of monthly data. The phase was first corrected for the three day offset. A discussion of obtaining the phase plot is given in Appendix A.6.1.

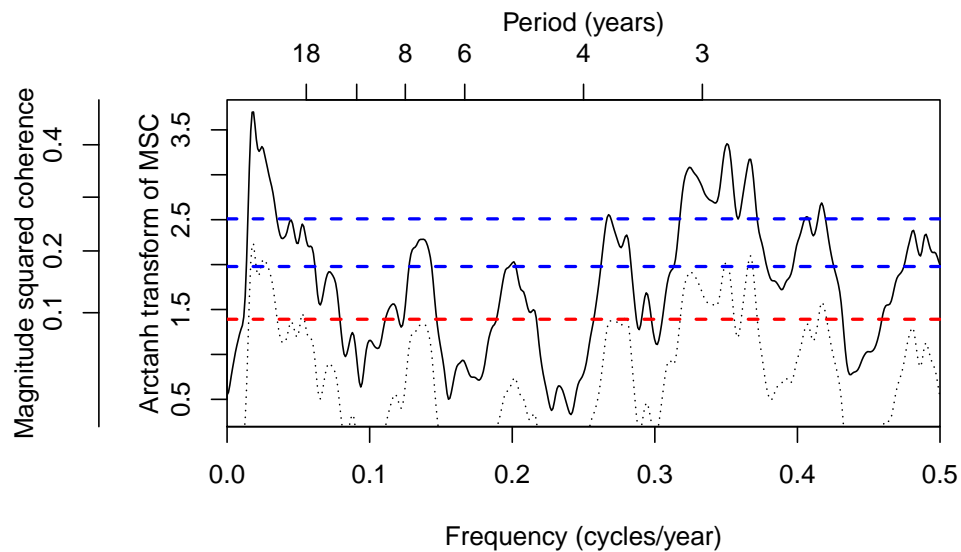


Figure 4.11: MSC between annual phase of the Central England temperature series and Burgundy GHD for years 1661 to 2003. The multitaper parameters are: $NW = 6.5$, $K = 11$. The coherence is modest at low frequencies. The annual phase of the Central England temperature series was calculated with zeroth order Slepian complex demodulation technique with a length of $N = 36$, 3 years of monthly data, with $NW = 4.5$. The three-day offset for years 1661 to 1752, originally reported in Thomson (1995), discussed on page 174, was applied. The dashed red line indicates the bias value of 0.091, and the dashed blue lines indicate a MSC of 0.17 and 0.21.

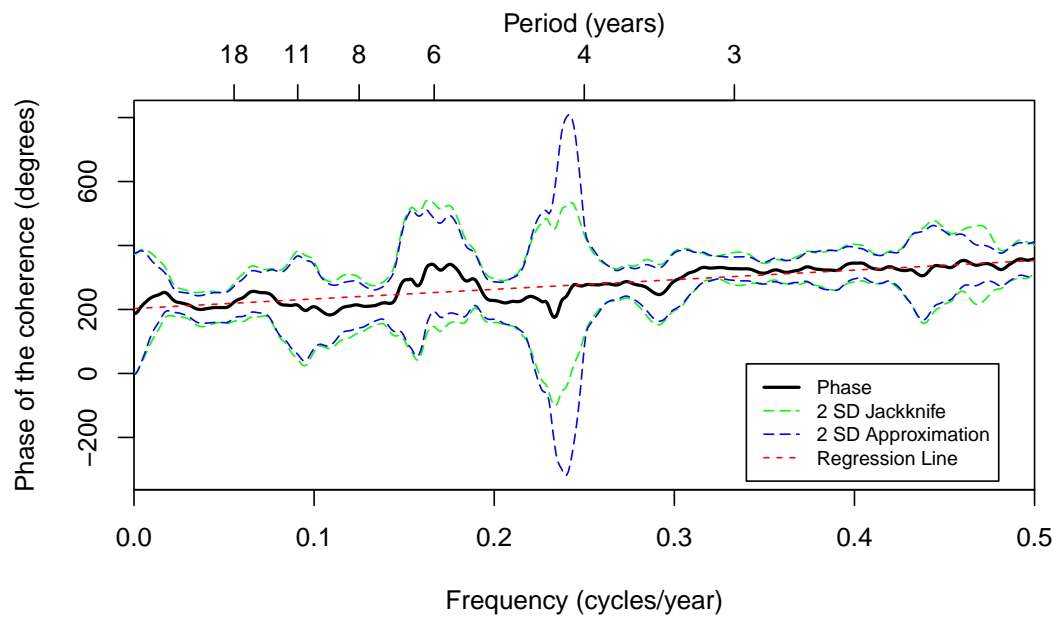


Figure 4.12: Phase coherence between the Burgundy GHD and annual phase of the Central England temperature series calculated over three years. This figure is based on (2.67) with the Burgundy series is represented by x and the annual phase of the Central England temperature series represented by y . The intercept is positive and the slope is ~ 300 degrees per year indicating the Burgundy GHD leads phase of the CET series by ~ 305 days. The multitaper parameters are: $NW = 6.5$, $K = 11$.

4.3 Spectrograms and Level-of-change

In this section, we consider stationarity. Specifically, we determine whether there is a benefit to splitting the GHD series into two series at a change-point and whether there is a particular change-point at which we should split the series. To do this, we use the frequency-domain level-of-change estimator discussed in Chapter 3. We begin by attempting to locate one change-point creating two segments. One can then consider whether it is feasible to search for more change-points. Figure 4.13 presents a spectrogram of the GHD series. The spectrogram has considerable overlap but is presented as a pictorial description of the series. Each column represents a multitaper spectral estimate, and one can see the evolving multitaper spectral estimates. As we mentioned earlier, a harmonic component that is present for only part of the series is captured in spectral estimates. This in essence is the manner in which spectral estimates are robust to nonstationarity. One may ask why we need longer series or why not select only short blocks. The answer is that longer series provide increased resolution in frequency and increased statistical power. This is essentially the Heisenberg uncertainty principal applied to spectral analysis. A visual inspection of the spectrogram in Figure 4.13 leads one to suspect nonstationarity and to ask whether we can detect structural change in the spectra.

Figure 4.14 uses the Bartlett M-test to examine stationarity as a function of frequency (see Section 3.4.1). The red dotted line indicates a 95% significance level, and the green dashed lines indicates the expected value if the series were stationary. It does appear that there is nonstationarity in the frequency band between 0.1 and 0.18 cycles/year and between 0.2 and 0.24 cycles/year. This is consistent with the spectrogram presented in Figure 4.13, and this helps to give a understanding of how

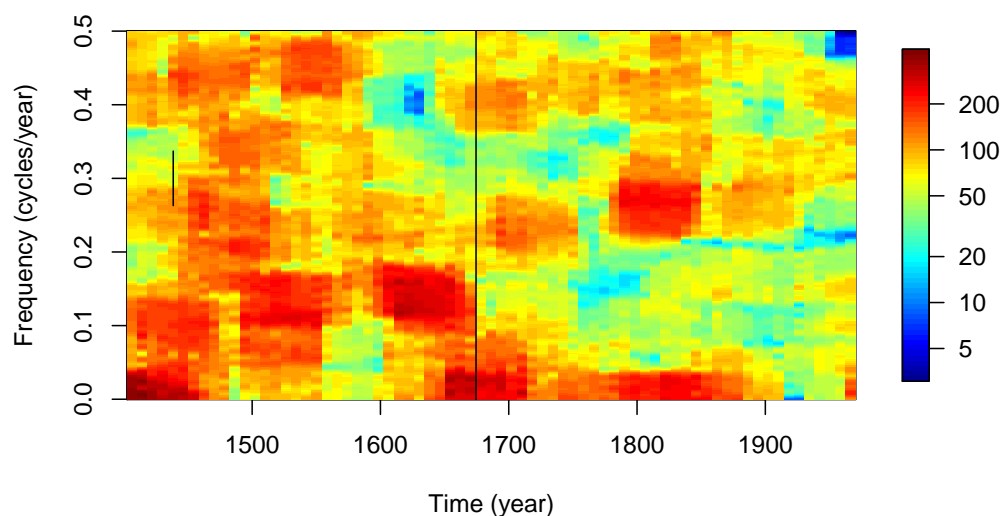


Figure 4.13: Multitaper spectrogram with considerable overlap. In this case the block length is 74, there are 71 blocks, and the offset is 8 years. This indicates an overlap of about 89%, but it allows for higher-frequency resolution. The vertical line segment on the left indicates the bandwidth, $2W$, and one can see the spectral estimates evolve over time. The centre line indicates where are analysis selects to section the series.

the frequency-domain changes over time.

We next construct the level-of-change estimator. To do so, a spectrogram with almost no overlap (only 6%) is constructed. This spectrogram is in essence a rough copy of Figure 4.13 and is not presented. Figure 4.15 shows the level-of-change for this series, and if we restrict ourselves to the frequencies considered nonstationary by the Bartlett M-test, it appears that a change-point exists in the centre of the series.

Having selected a candidate change point at 1675.5, we examine before and after spectra. Figure 4.16 shows the individual multitaper spectra before and after the selected change-point, 1675.5, which is within the Maunder minimum, from 1645 to 1715 (Eddy, 1976). The before and after spectra are considerably different and the

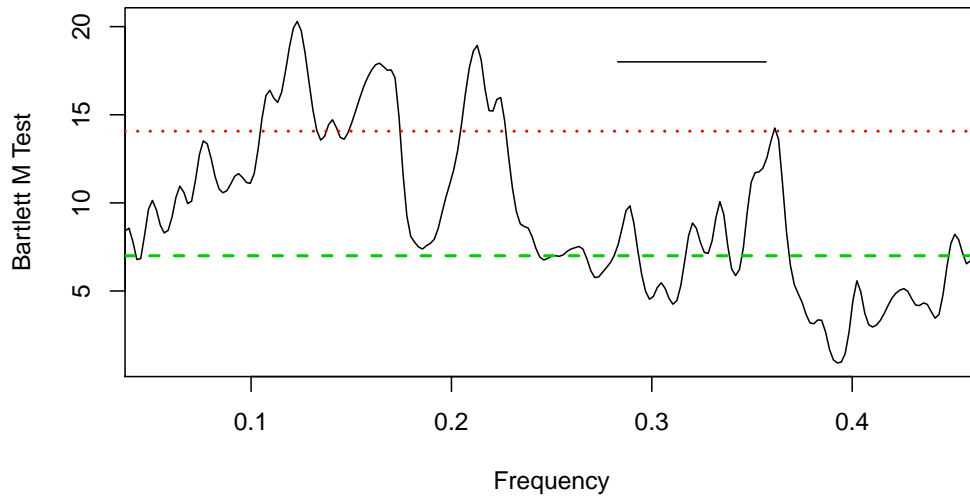


Figure 4.14: Bartlett M-test for stationarity using block sizes with 2.5% (little) overlap. The expected value (green dashed line) and the 95% significance level (red dotted line) are on the graph. The multitaper parameters used were $NW = 3$ $K = 5$, with 8 blocks, each of length 81 with an offset of 79. The line segment in the top right of the plot indicates the bandwidth. Nonstationary components are approximately between the frequencies of 0.1 and 0.18 cycles/year, and between 0.2 and 0.24 cycles/year

harmonic components are different in each half.

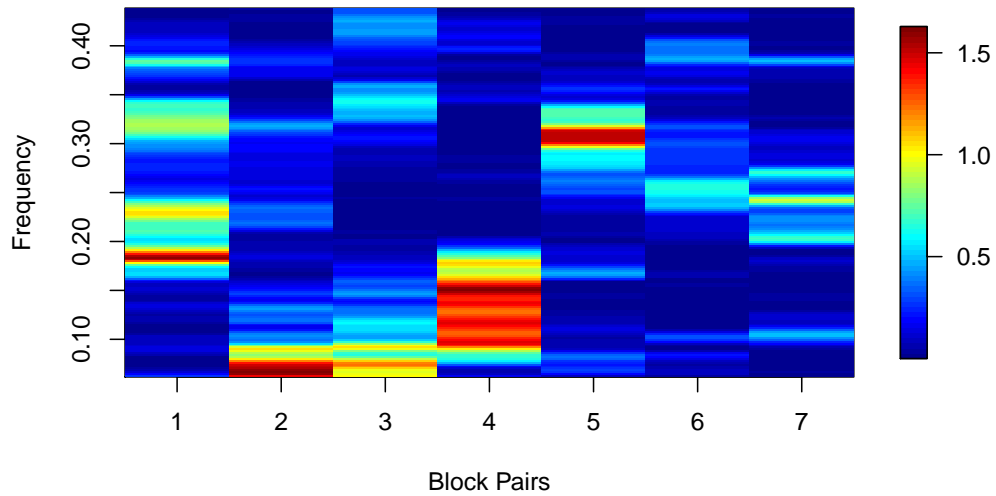


Figure 4.15: We plot the level-of-change between blocks in the spectrogram for the GHD. If we restrict ourselves to the frequency of interest, 0.10 to 0.18, based on the Bartlett M-test, we see that considerable change occurs at approximately the centre of the series.

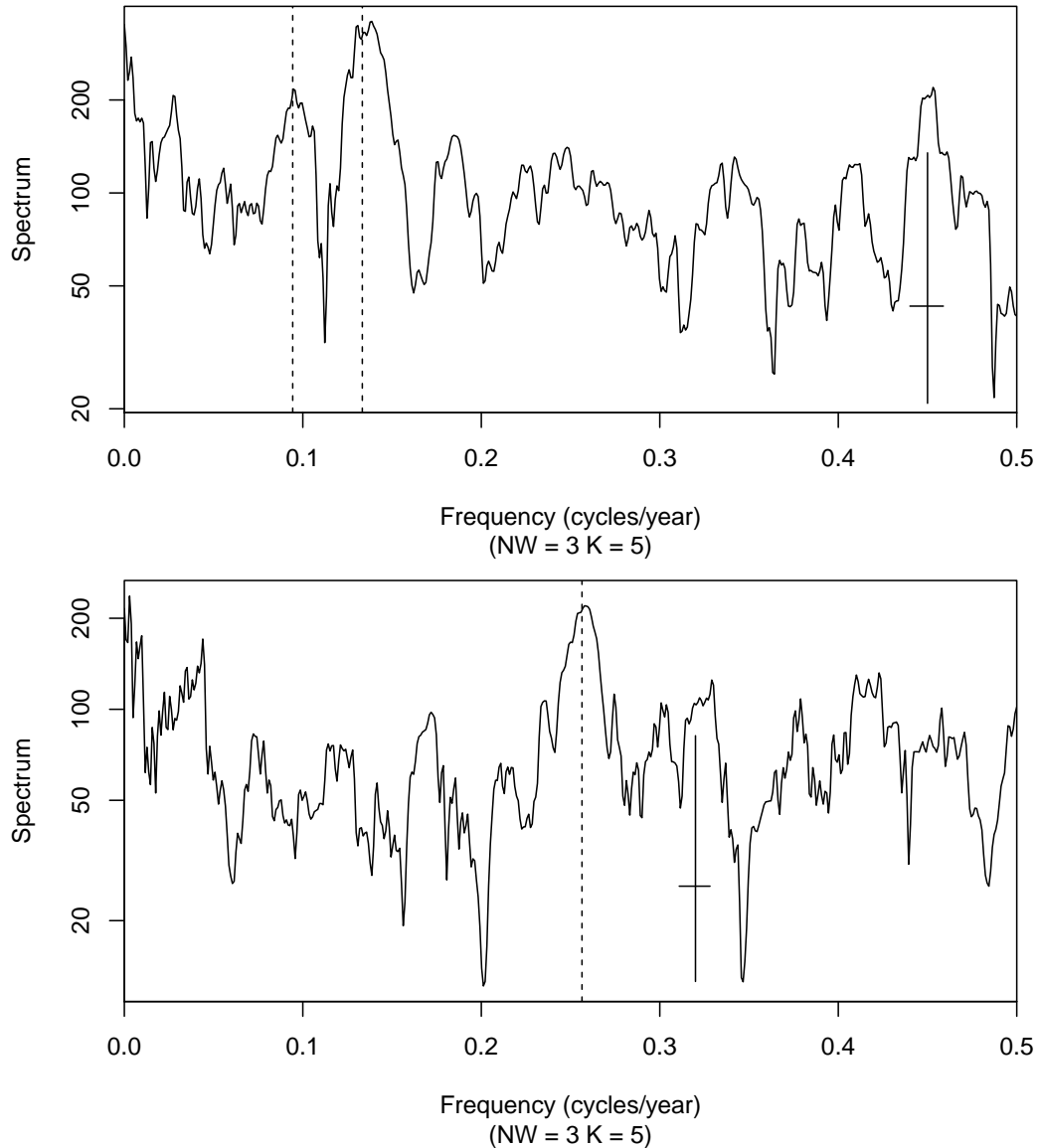


Figure 4.16: Multitaper spectra of the GHD before (top) and after (bottom) the year 1675.5. The crosses indicate 95% confidence levels and the width of bandwidth parameter, $2W$. On the upper plot, the dashed lines indicate a period of 10.6 years (0.94 cycles/year), and 7.5 years (0.133 cycles/year) for the date up to the year 1675. On the lower plot, the dashed line indicates a period of 3.9 years (0.278 cycles/year). It appears that a change in the spectral properties of the GHD series occurs when the data is sectioned at the year 1675.

4.4 Summary and Concluding Remarks

We have presented a coherence study comparing the Burgundy GHD series with similar series and found considerable coherence between it and the Swiss GHD series and between it and the CET series. We consider this as evidence of data quality, and evidence that a similar climate signal is captured in these series. The phase delays observed in Figures 4.6 and 4.9 indicate that the Burgundy GHD series lags the Swiss GHD series by 9 days, and that the Burgundy GHD series lags average annual temperature of Central England series by 18 days.

We applied our spectral analysis change-point detection tool, discussed in Chapter 3, and we found a change-point within the Maunder minimum, which was from 1645 to 1715, a period of cooler climate coincident with decreased solar irradiation (Eddy, 1976)

The analysis indicates changing spectra across the time series, with the largest change-point located at 1676. The sectioned series (before and after the change-point) exhibit different spectra, as seen in Figure 4.16. In the future, a similar analysis can, and should, be carried out on the adjacent series, and the tool and methodology presented here can be used in analysis, specifically to section the data into “quasi” stationary sections, and more in-depth study into the potential causes of coherence should be considered.

Chapter 5

Goodness-of-fit in Autoregressive Processes

5.1 Introduction

This chapter contains a paper submitted to the proceedings of the at the 2013 Joint Statistical Meetings conference (Rahim and Thomson, 2013); the paper is presented here with improvements and changes.

In this paper we (a) use simulation to show that multitaper spectral estimation in conjunction with the Levinson-Durbin recursions provides accurate selection of autoregressive (AR) coefficients, (b) propose a practical test for assessing the goodness-of-fit of AR coefficients, (c) use simulation to examine the proposed goodness-of-fit test, and (d) fit several AR models to the Burgundy grape harvest date (GHD) time series. Our tests find that several models are acceptable for the GHD series. This paper does not consider the problem of AR coefficient order selection, and in practical applications we use the Akaike information criterion (AIC).

AR models are used in many applications. For example, they are used to prewhiten data in engineering applications (Thomson, 1977a) and to prewhiten data in climate science prior to spectrum estimation (Mann and Lees, 1996). The choice of AR model and the estimated AR coefficients used in prewhitening can affect the residuals and the subsequent spectral analysis, masking or enhancing features of the spectrum. Using simulation, we show that AR coefficients obtained by different methods are not equivalently distributed, and we find in our simulations that the Levinson-Durbin recursions with multitaper spectral estimates and Burg's algorithm produce unbiased low-variance estimates. We fit and compare the goodness-of-fit of several AR models to the Burgundy GHD series (Chuine et al., 2004; Tourre et al., 2011).

Goodness-of-fit for AR models has been proposed and discussed in the literature (Priestly, 1981, pp. 475–494; Anderson, 1997). Our test is based on the maximum absolute deviation of the integrated spectrum, originally proposed in Bartlett (1937), and as a practical point, our test uses simulation to determine approximate p-values.

There are multiple methods for fitting AR coefficients, and we review two popular methods: (a) solving the Yule-Walker equations with Levinson-Durbin recursions, and (b) using Burg’s recursions with forward and backward estimators.

Some authors argue that Yule-Walker equations should not be used (De Hoon et al., 1996); however, we dispute this and demonstrate that the Yule-Walker equations can be effective in finding coefficients for an AR process with roots close to the unit circle when used with an appropriate spectral estimate, such as the multitaper, and solved with the Levinson-Durbin recursions. The Levinson-Durbin recursions avoid matrix inversion in straight forward Yule-Walker equations, and have been found more effective on digital computers (P&W93, p. 403). We note that the Burg method is known to split spectral lines (Ulrych and Bishop, 1975),¹ and there are concerns about the Burg method producing unstable models (Burg et al., 1982).

This paper is organized in the following manner: Section 5.2 discusses basic theory of AR coefficients and reviews two of the procedures used in calculation of the coefficients, Section 5.3 reviews some general cautionary notes regarding the use of AR models, Section 5.4 compares methods used in obtaining AR coefficients, Section 5.5 discusses goodness-of-fit tests for AR coefficients, Section 5.6 presents a simulation analysis of goodness-of-fit tests, Section 5.7 compares various fitted AR models for the Burgundy GHD series, and Section 5.8 gives concluding remarks and suggests

¹Adjacent peaks in the power spectrum can appear as one peak when the Burg method is used.

future work.

5.2 Calculation of AR Coefficients

5.2.1 Preliminaries

If $\{Z_t\}$ is a purely random process with zero mean and variance σ_Z^2 , indexed by $t = 1, 2, \dots, N$, then the process $\{X_t\}$ is an AR process of order p (denoted as an AR(p) process), and we have

$$X_t = \phi_1 X_{t-1} + \dots + \phi_p X_{t-p} + Z_t. \quad (5.1)$$

Subject to certain constraints, AR(p) processes are often considered second-order stationary, meaning that the first and second moments are time-invariant. See Chatfield (2004, pp. 43–44) for the constraints. The coefficients $\phi_1, \phi_2, \dots, \phi_p$ are called *autoregressive* coefficients.

Remark 1. Equation (5.1) notes an analogy between an AR(p) model and a regression problem; however, instead of independent variables, the right side of equation (5.1) has lagged copies of the dependent variable.

Remark 2. The p^{th} coefficient of an AR(p) process is called a *reflection* or *partial autocorrelation* coefficient. One can use the notation $\hat{\phi}_{1,p}$ to indicate the first sample autocorrelation coefficient in an order p model, and thus indicate that it differs from $\hat{\phi}_{1,1}$, which would be the sole autocorrelation coefficient in an AR(1) process. We will use the two-subscript notation if the order, p , is unclear or changing.

The partial autocorrelation coefficient $\phi_{j,j}$ represents the correlation between X_t , and X_{t-h} with the linear dependence of the interceding terms, $X_{t-1}, X_{t-2}, \dots, X_{t-h+1}$

removed, or *partialled* out. The idea of partial correlation was introduced in Yule (1897). This was groundbreaking work because, as is common with modern work, it relies on regression. Yule did not invoke normality as much of the data he was using was anything but normal.

Remark 3. An alternative notation for AR(p) often used in engineering and in Priestley (1981) processes is:

$$Y_t + \alpha_1 Y_{t-1} + \alpha_2 Y_{t-2} + \dots + \alpha_p Y_{t-p} = Z_t, \quad (5.2)$$

where $\alpha_j = -\phi_j$.

Remark 4. A nonstationary process is considered unstable and not all AR processes are stationary (Box et al., 1994, p. 10). If all the roots of the polynomial equation $1 - \sum_{j=1}^p \phi_j z^{-j}$ lie *inside* the unit circle, then the AR process is stationary. This chapter focuses on obtaining AR coefficients and testing goodness-of-fit without assuming stability which should be assessed prior to making any forecast. We refer the reader to Chatfield (2004, pp. 262–264) for a discussion of testing unit roots.

The autocovariance sequence (acvs) for lag τ is defined as:

$$\gamma_\tau = E\{[X_t - \mu][X_{t-\tau} - \mu]\}, \quad (5.3)$$

where μ is the mean of the process X_t , and $\tau = 0, 1, \dots, N - 1$. We see that if $\tau = 0$, then $\gamma(0)$ is simply the variance. The autocorrelation sequence is defined as:

$$\rho_\tau = \frac{\gamma_\tau}{\gamma_0}, \quad (5.4)$$

and $\rho_0 = 1$.

Remark 5. A plot of the sample autocorrelation coefficients over increasing lag, τ , is known as a *correlogram*.

The typical *biased* estimator of the acvs is

$$\hat{\gamma}_\tau^B = \frac{1}{N} \sum_{\tau=0}^{N-|\tau|} [X_t - \bar{X}][X_{t+|\tau|} - \bar{X}]. \quad (5.5)$$

Remark 6. If one replaces \bar{X} with μ , and multiplies by $\frac{N}{N-|\tau|}$ one would have an unbiased estimate; however, simply making the second substitution while using an estimator of μ will not produce an unbiased estimator (Bartlett, 1978).

Theorem 1. *The sequence formed by (5.5) is positive definite if and only if the realizations of X_1, X_2, \dots, X_N are not all identical.*

Remark 7. Generally, fast Fourier transforms (FFTs) are used instead of calculating the acvs in equation (5.5) directly. (See remark 9.)

We will often examine and estimate the spectral density function (SDF), which is the Fourier transform of the acvs,

$$S(f) = \sum_{\tau=-\infty}^{\infty} \gamma_\tau e^{-i2\pi f\tau}. \quad (5.6)$$

In equation (5.6), we allow $\tau \in \mathbb{Z}$, thus we are considering a process with infinite past and future. Frequency, f , takes values in $[0, 1/2]$. The above equality is true only in a mean-square sense, but it can be considered pointwise in all practical applications² The fact that autocovariances and the spectral density function are a Fourier transform pair was first discovered in 1914 (Einstein, 1987) but was overlooked (see Yaglom, 1987b) It was rediscovered independently by Wiener and Khintchine in the 1930s.

²A sequence of functions $\{f_n\}$ converges pointwise to the function f if and only if $\lim_{n \rightarrow \infty} f_n(x) = f(x)$. In practice, one uses the acvs and spectrum as Fourier transform pairs.

Remark 8. The SDF for a stationary AR(p) process is

$$S_{AR}(f) = \frac{\sigma_z^2}{\left|1 - \sum_{j=1}^p \phi_j e^{-i2\pi f j}\right|^2}, \quad \text{for } |f| \leq 1/2, \quad (5.7)$$

when $\Delta t = 1$. In practice this equation is calculated using the FFT.

The customary estimator of the SDF is the direct spectral estimator:

$$\hat{S}_D(f) = \Delta t \left| \sum_{t=1}^N h_t x_t e^{-i2\pi f t \Delta t} \right|^2. \quad (5.8)$$

In the above estimator, Δt is the change in time step, t , and h_t is a data taper. If we allow $h_t = \sqrt{1/n}$, the direct spectral estimator becomes the so-called *periodogram*, which we denote as $\hat{S}(f)$. The raw periodogram is asymptotically unbiased, but the bias can exist even with large sample sizes in practical applications (Thomson, 1982, p. 1058). Additionally, the periodogram is an *inconsistent* statistical estimator—that is, the variance does not decrease as the sample size increases (Rayleigh, 1903). It can be shown that, for real data, this estimator has a χ_2^2 distribution when $h_t = \sqrt{1/N}$ for all frequencies except $f = 0$ and $f = 1/2$, which contain only real values and thus have a χ_1^2 distribution (Blackman and Tukey, 1959).

Remark 9. The periodogram and the biased estimator of the acvs, equation (5.5), are Fourier transform pairs,

$$\{\hat{\gamma}_\tau^B\} \longleftrightarrow \{\hat{S}(f)\},$$

We will use a set of orthonormal discrete prolate spheroidal sequence (DPSS), also known as *Slepian sequences*, as data tapers. These sequences are defined as solutions to the system of equations (Slepian, 1978):

$$\sum_{t'=0}^{N-1} \frac{\sin[2\pi W(t-t')]}{\pi(t-t')} v_{t',k}(N, W) = \lambda_k(N, W) v_{t,k}(N, W) \quad (5.9)$$

for $t, t' = 0, 1, \dots, N - 1$. These sequences are discrete time analogs of real functions that are optimally concentrated in time and frequency (Slepian, 1978). The parameter W represents the effective bandwidth, which is often included in the time-bandwidth parameter NW , and k represents the current taper. Typically there are $k = 0, 1, \dots, K - 1$ tapers where $K = 2NW$. We use a set of orthonormal Slepian sequences in constructing multitaper spectral estimate. If we let $\hat{S}_k(f)$ represent the direct spectral estimator in equation (5.8), formed using the Slepian sequence of order k , then the simplest form of the multitaper spectral estimate becomes

$$\hat{S}^{(MT)} \equiv \frac{1}{k} \sum_{k=0}^{K-1} \hat{S}_k(f). \quad (5.10)$$

The individual $\hat{S}_k(f)$ estimates are known as *eigenspectra*, and the averaged estimator, equation (5.10), is distributed as χ_{2K}^2 for $f \neq 0$ and $f \neq 1/2$.

Remark 10. When using Slepian sequences, one selects the time-bandwidth parameter NW which in turn specifies W . Typically one sets a bandwidth parameter between 2 and 6, and noninteger values can be used (Thomson, 1982, p. 1086). Judicious selection of the bandwidth parameter can allow, for example, for the resolution of a lower-power harmonic that would otherwise be masked by an adjacent higher-power harmonic.

Remark 11. In practice, we will use the adaptive weighted multitaper spectral estimate, $\hat{S}^{(AMT)}(f)$, which uses a sophisticated weighted averaging scheme. This weighting scheme generally down-weights higher-order eigenspectra, which have a higher bias. See Thomson (1982, pp. 1065–1066) for more details. This weighted averaging scheme provides a non-integer degree-of-freedom estimate at each frequency that is typically slightly below $2K$, but can also be significantly lower.

The Toeplitz symmetric matrix Γ [equation (5.14)] must be positive definite for the procedure to make sense. If it is not, one obtains nonsensical results such as negative prediction variances. This is the reason that one uses the biased form of $\hat{\phi}_Z^B$ in (5.5). If one replaces the $1/N$ with $\frac{1}{N - \tau}$, the unbiased estimates are not positive definite. However, if the correlations are the Fourier transform of a positive spectrum, they are guaranteed to be positive definite by Bochner's theorem.

(e) Thus we have an estimate:

$$\boldsymbol{\phi}_p = \Gamma_p^{-1} \boldsymbol{\gamma}_p. \quad (5.15)$$

This gives us the variance of white noise term in equation (5.1) estimate as:

$$\sigma_Z^2 = \gamma_0 - \boldsymbol{\phi}_p^T \boldsymbol{\gamma}_p. \quad (5.16)$$

(f) Finally, the *method of moments* is used and the estimator of $\hat{\gamma}^B$ from equation (5.5) is used to form $\hat{\boldsymbol{\gamma}}_p$ in equations (5.15) and (5.16), and these equations become

$$\boldsymbol{\phi}_p = \hat{\Gamma}_p^{-1} \hat{\boldsymbol{\gamma}}_p, \quad \text{and} \quad \hat{\sigma}_Z^2 = \hat{\gamma}_0 - \hat{\boldsymbol{\phi}}_p^T \hat{\boldsymbol{\gamma}}_p. \quad (5.17)$$

Note that the vector $\boldsymbol{\gamma}_p$ represents the autocorrelation coefficients ϕ_j for an AR(p) process.

The Yule-Walker equations can be solved by matrix inversion but are generally solved using the Levinson-Durbin recursions, which are related to a modified *Cholesky* [lower triangular matrix] decomposition. The code must be written carefully, and 64-bit floating point arithmetic is required to avoid loss of precision.

5.2.3 Levinson-Durbin Recursions

5.2.3.1 Preliminaries

Let $\mathcal{W}_{i,j}$ be a subvector extraction operator. If $\mathbf{v} = (v_1, v_2, \dots, v_N)^T$, then

$$\mathcal{W}_{i,j} \mathbf{v} = (v_i, v_{i+1}, \dots, v_j)^T, \quad (5.18)$$

where $1 \leq i \leq j \leq N$.

5.2.3.2 One-step-ahead Prediction

We write the one-step-ahead AR(p) best linear predictor as

$$\vec{X}_{N+1}(p) = \phi_1 X_N + \phi_2 X_{N-1} + \dots + \phi_p X_{N-(p-1)}. \quad (5.19)$$

In matrix notation this can be written as

$$\vec{X}_{N+1}(p) = \boldsymbol{\phi}_p^T \mathbf{X}_p,$$

where $\mathbf{X}_p = \mathcal{W}_{N-(p-1),N} \mathbf{X}$ that is, \mathbf{X}_p is the vector of the last p elements from $\mathbf{X} = (X_1, X_2, \dots, X_N)^T$. The mean-squared one-step-ahead prediction error is given by

$$\begin{aligned} P_{N+1} &= E\{(\vec{X}_{N+1}(p) - X_{N+1})^2\} \\ &= \gamma(0) - \boldsymbol{\gamma}_p^T \Gamma_p^{-1} \boldsymbol{\gamma}_p. \end{aligned} \quad (5.20)$$

See Shumway and Stoffer (2006, p. 112) for details.

5.2.3.3 Levinson-Durbin Algorithm

The recursions begin by setting

$$\phi_{0,0} = 0, \quad \text{and} \quad P_1 = \gamma(0), \quad (5.21)$$

then for $n \geq 1$, the partial autocorrelation coefficients are updated by

$$\phi_{n,n} = \frac{\rho_n - \sum_{k=1}^{n-1} \phi_{n-1,k} \rho_{n-k}}{1 - \sum_{k=1}^{n-1} \phi_{n-1,k} \rho_k}, \quad (5.22)$$

and the mean-squared one-step-ahead prediction error is given by

$$P_{n+1} = P_n(1 - \phi_{n,n}^2). \quad (5.23)$$

The update of the mean-squared one-step-ahead prediction error in equation (5.23) was first used in this context by Burg in 1961 (Burg, 1975, p. 14); however, it has earlier origins (Yule, 1907). The autocorrelation coefficients are obtained when $n \geq 2$ using

$$\phi_{n,k} = \phi_{n-1,k} - \phi_{n,n} \phi_{n-1,n-k}, \quad \text{for } k = 1, 2, \dots, n-1. \quad (5.24)$$

5.2.3.4 Using Tapered Spectra to Estimate the ACVS

It has been noted that there is no reason to restrict oneself to the acvs computed from untapered spectral estimates (Thomson, 1977a, p. 1773) one can use direct spectral estimators and multitaper estimates. It has been shown, (e.g., P&W93, pp. 405–406) that a direct spectral estimator using Slepian sequences with $NW = 2$ accurately depict, the theoretical spectra of a known AR(4) process. We present simulations comparing different estimates for a known AR(4) sequence.

5.2.4 Burg's Method

5.2.4.1 Overview

Burg's method is also a solution of the Yule-Walker equations; however, it focuses on directly estimating the partial autocorrelation coefficients without using an estimate

of $\hat{\gamma}_\tau$. It does this by focusing on minimizing the error in the one-step-ahead and one-step-back prediction estimates (Burg, 1968). In practice, the Burg algorithm has been shown to be more efficient than use of the Levinson-Durbin recursions using the standard biased estimator, $\hat{\gamma}_\tau^B$, for smaller sample size. Our simulations indicate that the Burg method is considerably more effective than the Levinson-Durbin recursions when using the standard biased estimator, $\hat{\gamma}_\tau^B$, but it is not significantly more effective, when a multitaper spectral estimate version of $\hat{\gamma}_\tau$ is used. The Burg estimator is not without its own drawbacks such as splitting lines (Ulrych and Bishop, 1975) (see Section 5.3).

Some authors consider Burg's method more effective when roots of the characteristic polynomial are close to the unit circle (De Hoon et al., 1996). However, in Section 5.4, we show that the Yule-Walker equations when used with the multitaper method is as effective as the Burg method when applied to an AR(4) example with roots close to the unit circle. It was recognized in the 1970's that the Burg method both split lines and gave spectrum estimates with very high variance. A good example is given in Figure 2, Section VII of Burg et al. (1982). Various patches and corrections have been suggested, for example in Kaveh and Lippert (1983), but these destroy the elegance of Burg's original proposal and are usually not include in code. Further, because the algorithm minimizes the sum of the forward and reverse prediction variances it is very sensitive to the stationarity assumption.

One other practical concern about the Burg method due to missing lines is based on the following quote:

At an IEEE conference on underwater sound, N. Owsley of the Naval

Undersea Warfare Centre (NUWC), in Rode Island, noted that, in comparing various spectrum estimation on real data the Burg method had the distinction of missing a supertanker. (Thomson, D. J., pers. comm.)

5.2.4.2 Preliminaries

We write the prediction error associated with the one-step-ahead AR(p) predictor, equation (5.19), as

$$\vec{\epsilon}_t(p) = X_t - \vec{X}_t(p). \quad (5.25)$$

As in (5.19), we write the one-step-back AR(p) best linear predictor as

$$\overleftarrow{X}_t(p) = \phi_1 X_{t+1} + \phi_2 X_{t+2} + \cdots + \phi_p X_{t+p}. \quad (5.26)$$

We write the prediction error associated with the one-step-back AR(p) predictor, equation (5.26), as

$$\overleftarrow{\epsilon}_t(p) = X_t - \overleftarrow{X}_t(p). \quad (5.27)$$

If we define \mathcal{L} as a circular shift operator. If $\mathbf{v} = (v_1, v_2, \dots, v_N)^T$, then

$$\mathcal{L}\mathbf{v} = (v_N, v_1, v_2, \dots, v_{N-1})^T,$$

and next we define $\mathcal{M}_{j,k}$ as a subvector extraction operator, then

$$\mathcal{M}_{j,k}\mathbf{v} = (v_j, v_{j+1}, \dots, v_{k-1}, v_k)^T.$$

We plan to fit an AR(p) model to $\mathbf{X} = (X_1, X_2, \dots, X_N)^T$ and we define the following vector of length $N + p$

$$\vec{\mathbf{e}}(0) = (X_1, X_2, \dots, X_N, 0, 0, \dots, 0)^T.$$

The vector is \mathbf{X} concatenated with p zeros. We also define

$$\overleftarrow{\mathbf{e}}(0) = \mathcal{L}\vec{\mathbf{e}}(0) = (0, X_1, X_2, \dots, X_N, 0, \dots, 0)^T.$$

5.2.4.3 The Burg Estimator

We define the variance $\tilde{\sigma}_0^2 = \hat{\gamma}_0^B$, then for $k = 1, 2, \dots, p$ we recursively compute the following:

$$\tilde{\phi}_{k,k} = \frac{2\langle \mathcal{M}_{k+1,N} \vec{\mathbf{e}}(k-1), \mathcal{M}_{k+1,N} \overleftarrow{\mathbf{e}}(k-1) \rangle}{\|\mathcal{M}_{k+1,N} \vec{\mathbf{e}}(k-1)\|^2 + \|\mathcal{M}_{k+1,N} \overleftarrow{\mathbf{e}}(k-1)\|^2} \quad (5.28)$$

$$\begin{aligned} \tilde{\sigma}_k^2 &= \tilde{\sigma}_{k-1}^2 (1 - \tilde{\phi}_{k,k}^2) \\ \vec{\mathbf{e}}(k) &= \vec{\mathbf{e}}(k-1) - \tilde{\phi}_{k,k} \overleftarrow{\mathbf{e}}(k-1) \\ \overleftarrow{\mathbf{e}}(k) &= \mathcal{L}(\vec{\mathbf{e}}(k-1) - \tilde{\phi}_{k,k} \overleftarrow{\mathbf{e}}(k-1)), \end{aligned}$$

where we use $\langle \cdot, \cdot \rangle$ to denote vector inner product, and $\|\cdot\|^2$ to denote the squared norm.

The Burg estimator $\tilde{\phi}_{k,k}$ differs from the Yule-Walker estimator $\hat{\phi}_{k,k}$. The key point of the Burg estimator is that an estimator of acvs, typically $\hat{\gamma}_\tau^B$ for $\tau > 0$, is no longer required, whereas for the Yule-Walker equations, an estimator of the acvs is required for integer values of $\tau \leq p$.

5.3 Cautionary Notes on Using AR Spectral Estimates

We generally consider AR models useful for prewhitening data, but we caution against its use in general spectral estimation in the physical sciences. Kaveh and Lippert (1983) believe that AR spectral estimation can be patched for use, but Tukey (1984) condemned the general use of parametric spectral estimation. For a general overview of the problems of parametric spectral estimation, see Kay and Marple (1981).

Two specific problems to note when using AR spectral estimates for a sinusoid in additive noise are that (a) the location of the peak in the spectrum is found to depend on the phase of the sinusoid, and (b) two adjacent peaks in the spectrum can appear as one peak (Ulrich, 1970). The second problem is known in the literature as *spectral line splitting*.

Two proposed solutions are (a) replacing the real-valued signal with an analytic signal Kay and Marple (1981, p. 1396) and (b) using improved estimates of the autocorrelation function, equation (5.1). The first solution must consider taking an appropriate Hilbert transform that does not have the same bias properties as estimates based on the raw (biased) periodogram, and the process becomes complicated in the presence of multiple lines. We take the latter approach.

5.4 Comparison of Methods for Finding AR Coefficients

We compare selected AR coefficient estimation techniques on simulated data from a high signal-to-noise ratio AR(4) process that has been used in the literature, $\phi = (2.7607, -3.8106, 2.6535, -0.9238)^T$ (Ulrych and Bishop, 1975; Box et al., 1994). This process has complex roots $0.62 + 0.76i, 0.76 + 0.62i, 0.62 - 0.76i, 0.76 - 0.62i$, which are close to the unit circle. In this section, we show that the Yule-Walker equations used with the multitaper method are as effective as the Burg method on this AR process.

Table 5.1 compares the mean-squared-error (MSE), mean, median, and sample standard deviation, from estimates of the partial autocorrelation coefficient $\phi_{4,4} =$

-0.9238 using the Levinson-Durbin recursions with the biased autocovariance estimator, to an autocovariance estimator based on a single Slepian taper $NW = 5$, and to an autocovariance estimator constructed using the adaptive multitaper method with $NW = 5$ and $k = 5$. The acvs estimators were calculated from the estimated spectrum using the property in Remark 9. Figure 5.1 is a comparison of partial autocorrelation coefficients. In this simulation, the multitaper spectral estimate and the Burg estimate are preferred, and the use of a single Slepian taper is preferred to the standard biased acvs estimator. This table gives a non-trivial example where the Yule-Walker equations based the multitaper spectral estimate and solved with the Levinson-Durbin recursions is as effective as the Burg method.

	No Taper	Single Taper	Multitaper	Burg
MSE	0.30903	0.00062	0.00017	0.00017
Mean	-0.4252	-0.9135	-0.9204	-0.9204
Median	-0.4114	-0.9159	-0.9214	-0.9213
Sample SD	0.2457	0.0227	0.0128	0.0125

Table 5.1: Comparisons of estimates of $\phi_{4,4}$ from 100,000 run simulations using the Yule-Walker equations with the biased autocovariance estimator, an autocovariance estimator using one Slepian taper with $NW = 5$, an adaptive weighted multitaper spectral estimate with $NW = 5$, and $k = 8$, and the partial autocovariance estimator made using Burg's method.

5.5 Goodness-of-fit Test for Autoregressive Processes

The approach for testing the goodness-of-fit of an AR process is based on comparing the observed standardized integrated spectrum to the theoretical standardized

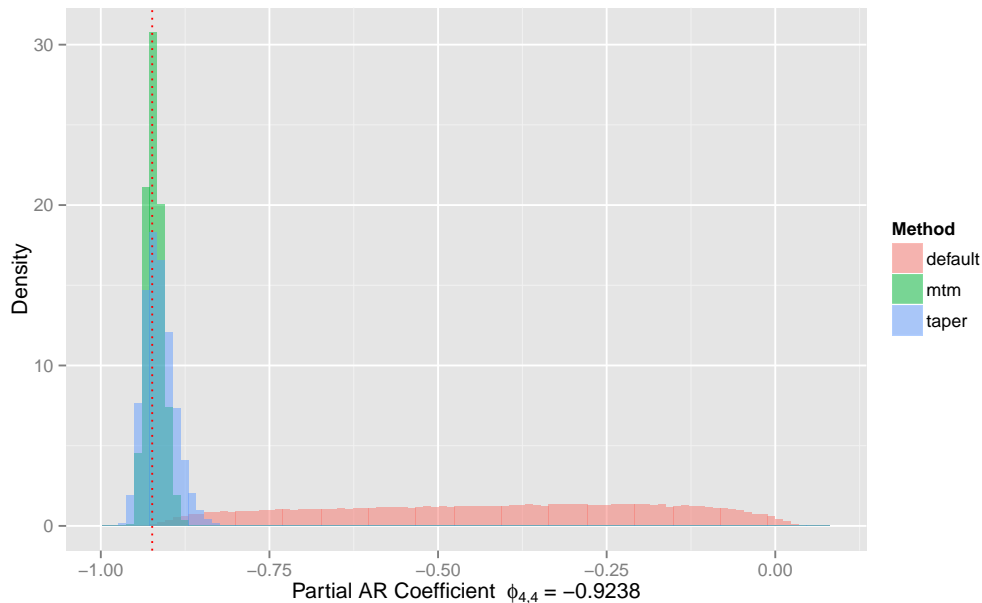


Figure 5.1: Estimated fourth-reflection coefficient based on a 100000-run simulation of an AR(4) process with coefficients 2.7607, -3.8106, 2.6535, -0.9238. Levinson-Durbin estimate using: (a) the default estimate—i.e., using the acvs from unwindowed Fourier transforms; (b) one DPSS taper with $NW = 5$; and (c) the use of an adaptive multitaper estimate with $k = 8$. The dashed line indicates -0.9238, the true value. Mean estimates were -0.425, -0.914, and -0.920 respectively. The distribution of the Burg estimator is very similar to the multitaper spectral estimator and is not shown.

integrated spectrum of the selected autoregressive model (Anderson, 1997).

5.5.1 Preliminaries

The empirical distribution, $\hat{F}(x)$, for a random sample of observations of X is generally $\hat{F}(x) =$ the proportion of samples observations $\leq x$. The integrated spectra $H(f_0) = \int_{-1/2}^{f_0} S(\xi) d\xi$ can be estimated well by

$$\hat{H}(f_0) = \int_{-1/2}^{f_0} \hat{S}_D(\xi) d\xi. \quad (5.29)$$

Note that tests based on the integrated spectrum, standardized or not, are generally not considered to be affected by the bias properties of using the raw periodogram (Priestley, 1981, p. 471). In the case of real-valued data, equation (5.29) can be adjusted to only consider positive frequencies (see: Priestley, 1981, p. 473).

The standardized integrated spectrum can be written as

$$F(f) = \frac{\int_{-1/2}^{f_0} S(\xi) d\xi}{\int_{-1/2}^{1/2} S(\xi) d\xi}, \quad (5.30)$$

which we estimate using the standard spectral estimator in equation (5.8). The goodness-of-fit tests draw on the correspondence between the standardized integrated spectrum and the empirical distribution function, and standardization provides the advantage that asymptotic distributions are valid under more general conditions than those without standardization (Anderson, 1997). As with the integrated spectrum, equation (5.29), this estimator can be constructed from only positive frequencies when restricted to real-valued data. Tests using the integrated spectrum are generally poor because they are insensitive to lower-power parts of the spectrum.

5.5.2 Goodness-of-fit Tests for AR Processes

An overview of goodness-of-fit tests for AR and moving average (MA) models are presented in Priestley (1981, pp. 475–494). We will be using the maximum absolute deviation of the integrated spectrum as a measure of goodness-of-fit, and we will use simulations to estimate p-values for the observed maximum absolute deviation. We note that Anderson (1997) proposes the same test statistic, the maximum absolute deviation of the integrated spectrum, to test the null hypothesis that the observations are on an AR process of an order not greater than the specified one. In place of

asymptotic results linking the Cramér-von Mises, or Kolmogorov-Smirnov statistic, we propose the practical measure of relying on simulations to generate approximate p-values.

Bartlett related the asymptotic distribution of the mean absolute deviation between the estimated normalized spectrum and the theoretical spectrum,

$$\max_{0 \leq f \leq 1/2} \sqrt{N} |\hat{F}_+(f) - F_+(f)|, \quad (5.31)$$

to the Kolmogorov-Smirnov statistic, which has been used in testing the goodness-of-fit in empirical distributions (Priestley, 1981, p. 480). We use the subscript positive sign, +, to indicate that we are constructing the estimate solely on positive frequencies, (see Section 5.5.1).

5.5.3 Proposed Methodology

Limiting distributions for the goodness-of-fit tests have been studied Anderson (1997), but practical software solutions are not readily available, and we propose a simple simulation-based statistical test. In addition, simulations do not constrain us to a one-size-fits-all approach. We propose (a) careful fitting of AR coefficients, (b) plotting the estimated spectra against the theoretical spectra—see equation (5.7), for the selected AR model, and (c) comparing the estimated standardized integrated spectrum to the theoretical spectra for the AR using the maximum absolute deviation as a test. We then use simulations to assess the significance of the observed distance. In constructing the theoretical AR spectrum used in the standardized integrated spectrum, we estimate σ_Z^2 in equation (5.7), from the data.

5.6 Simulations of Goodness-of-fit

We assess the goodness-of-fit for AR models for two AR models used in the literature, the AR(4) model discussed in Figure 5.1 and the AR(2) model $\phi = (0.75, -0.5)^T$ (P&W93, p. 45). Figure 5.2 compares empirical distributions of the distance, showing four comparisons in our simulations, two cases where the simulated AR process matches the theoretical, and two cases where we simulate mismatches (that is, the AR process simulated does not match the theoretical.) The top two plots show the distributions of the distances where the models accurately fit, and the lower two show distributions of misfit models. Comparing the top two plots in Figure 5.2 to the bottom two, one can see considerable change in the x -axis values. The misfit models generate larger distances. The red line indicates fitted Gamma distributions, and Table 5.2 indicates the shape and rate parameters of the fitted Gamma distributions.

The probability density of the gamma function is

$$f(x; k; \theta) = \frac{x^{k-1} e^{-\frac{x}{\theta}}}{\theta^k \Gamma(k)} \text{ for } x > 0 \text{ and } k, \theta > 0. \quad (5.32)$$

In (5.32), θ is the scale parameter, and k is the shape parameter, and the inverse scale parameter, $\beta = 1/\theta$, is called a rate parameter.

In order to get a sense of how the simulated integrated spectra compare to the theoretical, the top left plot in Figure 5.3 shows the observed integrated spectrum from that AR(4) simulation run that had the extreme (largest) value for maximum absolute deviation of the 40000 simulations against the theoretical integrated spectrum for the AR(4) process. The top-right plot shows the AR(2) simulation run that had the extreme (largest) value for maximum absolute deviation of the 40000 simulations

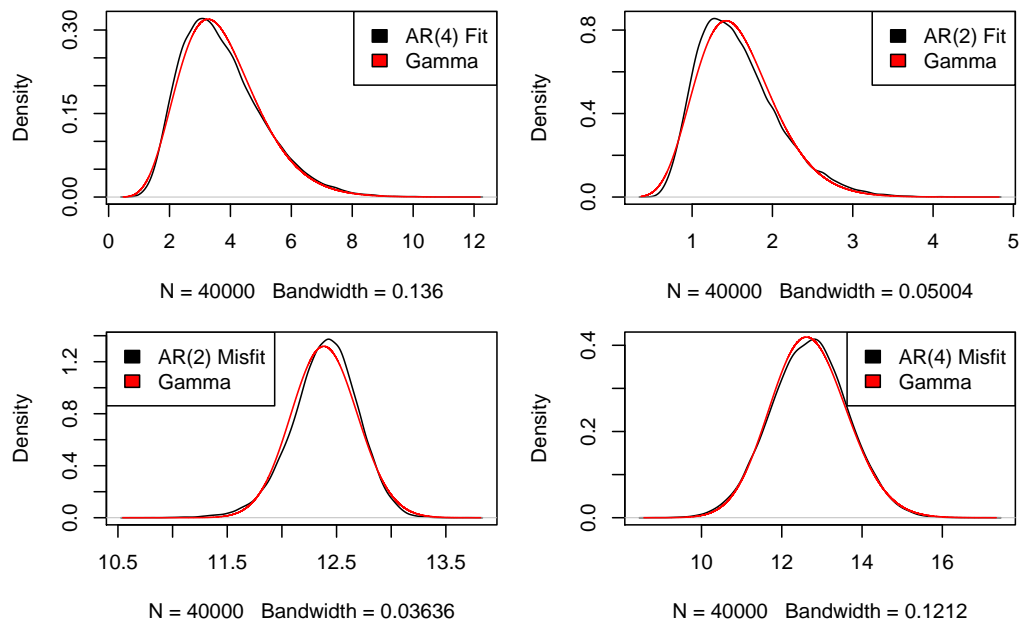


Figure 5.2: This figure shows the observed maximum absolute distance observed from 40000 simulations. The top left plot compares a simulated AR(4) to the theoretical AR(4), the top right plot compares a simulated AR(2) to the theoretical AR(2), the bottom left plot compares a simulated AR(4) to the theoretical AR(2), and the bottom right plot compares a simulated AR(2) to a theoretical AR(4). Note the changing y -axis scales.

against the theoretical integrated spectrum for the AR(2) process. Comparing the bottom left plot and the bottom right plot, one wonders whether a misfit to an AR(4) process is easier to detect than the misfit to the AR(2) process.

5.7 Burgundy Grape Harvest Dates

A link between Burgundy GHD and European climate fluctuations has been proposed (Tourre et al., 2011), and the Burgundy Pinot Noir grape is considered to be highly sensitive to climate variations; specifically, earlier harvest dates correspond to higher April to August temperatures (Chuine et al., 2004; Krieger et al., 2011).

Model	Shape	SE (Shape)	Rate	SE (Rate)
AR(4)	7.9970	0.0554	2.1444	0.0153
AR(2)	10.2543	0.07136	6.5002	0.0464
Misfit AR(2)	1676.2664	11.8357	135.2800	0.9553
Misfit AR(4)	176.7686	1.2486	13.9327	0.0986

Table 5.2: Shape and rate parameters with their respective standard errors, abbreviated SE, for the fitted Gamma distributions shown in Figure 5.2. Both the shape and rate parameters are considerably higher for the case where the simulated AR model did not match the theoretical model.

Prewhitening data reduces bias in analysis (Thomson, 1990b), and AR models are an efficient prewhitening tool (Thomson, 1990a). Mann and Lees (1996) propose removing spectral lines, fitting an AR(1) process, and then assessing significance of harmonic components in the spectrum using confidence intervals from the fitted AR(1) model. Figure 5.4 presents the raw spectrum of the GHD series, and the associated spectra of several AR models, including models of the same order where different techniques were used to obtain the AR coefficients. It does appear from the plot that the selection of AR model prewhitener can affect harmonic analysis of residuals. We proceed to fit several AR models to the Burgundy GHD series and test them for goodness-of-fit.

Using our method for comparing AR goodness-of-fit, Table 5.3 shows the observed maximum absolute deviation of the sample integrated spectrum, and it shows that the simulated p-values. Based on this goodness-of-fit criterion, we see little difference in the choice of models, and certainly no statistically significant difference. We conclude that each of the four models fits reasonably well. In looking at the spectrum in Figure 5.4, it appears that the choice of prewhitner can affect the significance of the harmonic components; however, this test does not enable us to distinguish between

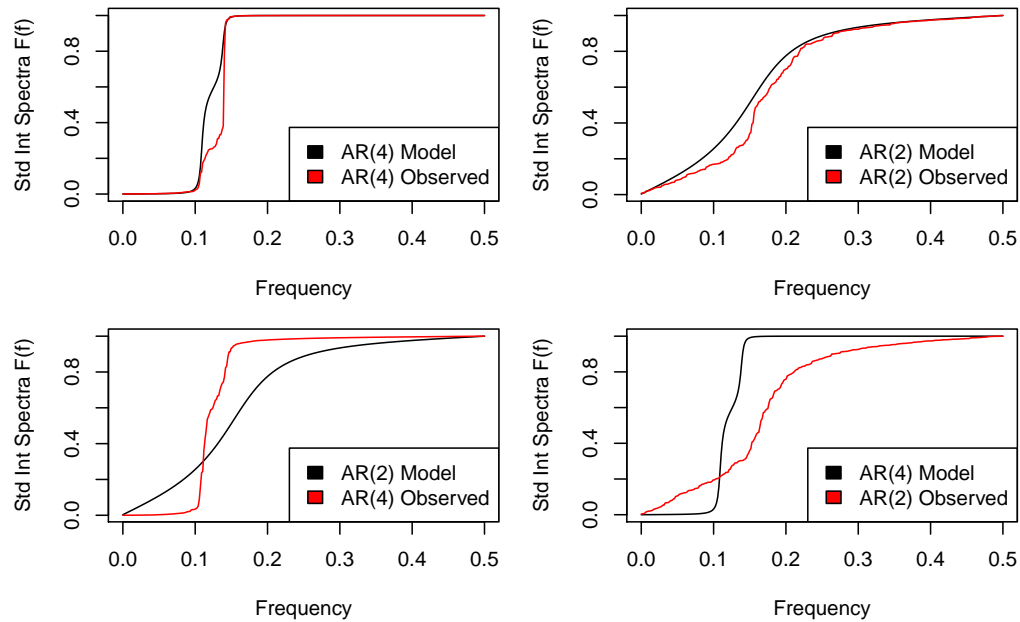


Figure 5.3: We ran 40000 simulations each comparing a simulated AR(4) to the theoretical AR(4), top left, a simulated AR(2) to the theoretical AR(2), top right, a simulated AR(2) to the theoretical AR(2), bottom left, and a simulated AR(4) to the theoretical AR(2), bottom right. The top two plots indicate the *worst fit* of the 40000 runs when the simulations were from the same model as the theoretical AR, and the bottom two plots indicate the *best fit* of the 40000 runs when the simulations are from a model than different from the theoretical AR.

models for this data set.

5.8 Conclusions and Future Work

This chapter demonstrates that multitaper spectral estimation of the acvs used with Levinson-Durbin recursions is more effective than an untapered spectral estimate used with the Levinson-Durbin recursions for a high signal-to-noise ratio AR(4) process with roots close to the unit circle. It also demonstrates that the multitaper Levinson-Durbin versions are as effective as Burg's method in this example. We propose a

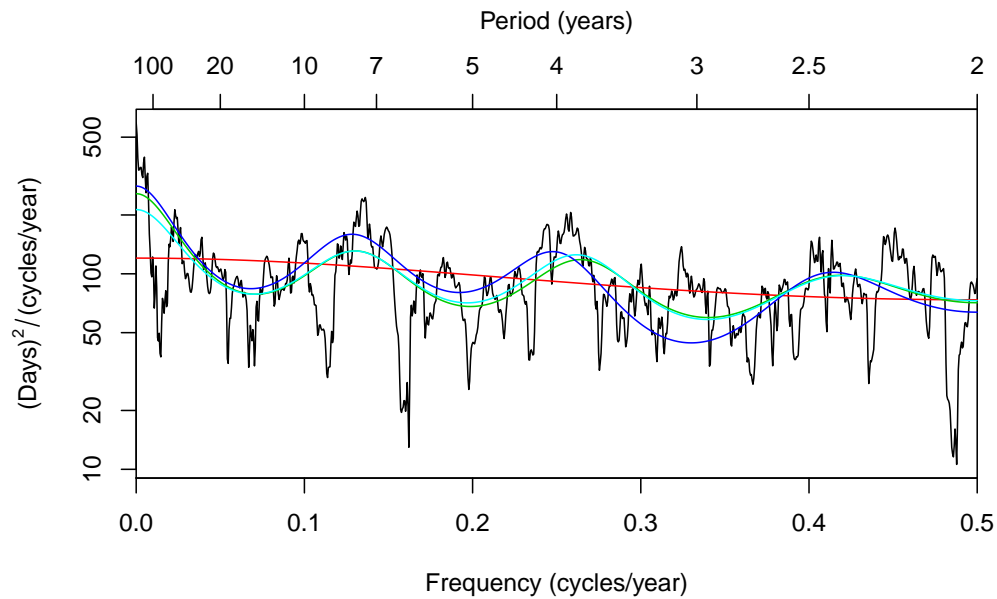


Figure 5.4: Adaptive multitaper spectrum of the GHD series. The parameters used are: $NW = 3$ and $k = 5$. Plotted over the spectrum, we have the standard AR(1) spectrum in red, the standard AR(8) spectrum in green, the DPSS tapered AR(8) spectrum in blue, and the multitaper AR(8) spectrum in cyan. The multitaper AR(8) in cyan and the standard AR(8) follow closely except between the frequencies 0.2 and 0.3 (cycles/year), where the multitaper estimate has slightly higher power and appears to follow the spectral estimate more closely.

AR(p) Model	Max Abs Dist	Simulated P-value
AR(1) no taper	0.9335	0.2634
AR(8) no taper	0.9291	0.2073
AR(8) 1 DPSS	0.9286	0.1839
Misfit AR(4)	0.9305	0.2256

Table 5.3: Maximum absolute deviation (max abs dist) of the observed GHD standardized integrated spectrum to the theoretical standardized integrated spectrum for the various models and approximate p-values based on simulations testing the null hypothesis that the maximum absolute deviation is small enough for the model to be appropriate.

practical method of testing the goodness-of-fit of AR estimators using the maximum absolute deviation between the standardized integrated spectrum from the estimated data and the theoretical standardized integrated spectra from the theoretical AR model, and we tested this method on two AR processes with simulations. We selected different AR models for the Burgundy GHD data set and used our goodness-of-fit tests to determine whether any are appropriate. We concluded that the four selected models fit reasonably well to the GHD series.

There are several areas for future work:

- (1) Test, with simulations, different but closely related AR models in order to determine how the test works, and test simulations with mixed spectra that include discrete line components.
- (2) Consider other ways of comparing two spectra. For example, the L^2 distance would be more sensitive to overall differences, whereas maximum absolute deviation may be more sensitive to high-power line components.
- (3) Explore questions of stationarity and change-point that exist in climate series

such as the Burgundy GHD series. One could section the series at a change-point and make multiple comparisons of spectra before and after the change-point to each other and to AR models for the entire series.

- (4) Study the effect of AR model selection on hypothesis tests, such as Mann and Lees (1996), for spectra exceeding the AR confidence intervals.

Chapter 6

Concluding Remarks and Future Work

We have presented a methodology for detecting change in spectra observed in Late Holocene climate data by presenting an analysis of the Burgundy grape harvest date (GHD) series which includes a coherence study finding underlying similarity between the Burgundy GHD series, the Swiss GHD series and the Central England Temperature (CET) series. Using our change-point detection methodology, we located a change in frequency structure of the Burgundy GHD series within the Maunder minimum. We proposed a method for detecting goodness-of-fit of autoregressive (AR) coefficients, and we demonstrated that the Yule-Walker method, when used with the multitaper spectral estimate and the Levinson-Durbin recursions, is as effective as the Burg method in finding AR coefficients in an example process with roots near the unit circle. The techniques introduced in this thesis rely heavily on multitaper spectral estimation, which is discussed in Chapter 2.

In Chapter 3 we introduced a change-point estimator to detect a change in spectral

structure over time. We derived the mean and variance values for the estimator under an independence assumption, confirmed these using simulations, and then used simulations to study relaxing of the independence assumption. We then tested our procedure on a change-point model and found that the estimator was graphically effective; however, it was not found to be statistically powerful. We presented it as a graphical tool as part of a methodology incorporating existing spectral multitaper tools. In Chapter 4 we presented spectral and statistical analysis of the Burgundy GHD time series. The analysis included a coherence study in which we found the Burgundy GHD series coherent with both the Swiss GHD and the CET series. This provides new evidence that the three series, Burgundy GHD, Swiss GHD and CET, capture similar climate signals. We then used the level-of-change estimator as part of a graphical technique to detect a change-point of 1675 in the Burgundy GHD, a date that was consistent with the Maunder minimum. Finally presented a spectral analysis of the sectioned series. In Chapter 5 we studied methods of calculating AR coefficients and presented a method for estimating the goodness-of-fit for AR estimators. We gave an example in which the coefficients estimated using the Yule-Walker equations with the multitaper spectral estimate, and solved with the Levinson-Durbin recursions provide results similar to those provided by the Burg method. The advantage of the multitaper method combined with the Yule-Walker equations is that the latter has a tunable parameter. The Burg method is known to split lines, and, while we did not find a statistical difference between the two in our examples using our goodness-of-fit test, the Yule-Walker equations with the multitaper method provide a tunable alternative, which, we have shown, can be accurate with a process with roots close to the unit circle.

Finally, we have included as Appendix A a description of software package developed to apply multitaper spectral estimation techniques in R. The material in the appendix constitutes the current update of a paper submitted to the *Journal of Statistical Software*. This software package constitutes a practical contribution that has already found use in the R software community.

The work discussed in Chapter 3 can be advanced with a more complete study of the level-of-change estimator, as we derived the mean and variance under an independence assumption that is violated in practice. Considering a non-central chi-squared distribution is an approach to improving this. The estimator did not have high statistical power in the presented example, and further examples can be studied. Chapter 4 is of interest to the climate community as it adds a coherence study, and a change-point location to the analysis published in Tourre et al. (2011). This work can be advanced by a more in-depth analysis of the sectioned series. Chapter 5 can be improved by including the L^2 distance in comparison, as the goodness-of-fit test was not found to be effective, and further tests can be studied. Additionally, one can consider a graphical use of the goodness-of-fit test. We also propose further study comparing the Burg method and Yule-Walker equations with the multitaper method for finding AR coefficients for processes with roots near the unit circle. We will continue to improve the software discussed in Appendix A. At this time the deadline for submission with corrections to the Journal of Statistical Software has passed. The editors suggested improvements to the code and documentation, added examples in the text, and clarification of the theory. We have worked to address many of their concerns; however, it is worth considering submitting the paper to the journal *Computers &*

Geosciences, as we receive email comments and questions about the multitaper package in R from those the geophysics and climate science community; the multitaper technique is not a standard statistical tool.

Bibliography

- M. Abramowitz and I. A. Stegun. *Handbook of Mathematical Functions*. National Bureau of Standards, Washington, DC, 1965. Applied Mathematics Series 55.
- H. Aksoy, A. Gedikli, N. E. Unal, and A. Kehagias. Fast segmentation algorithms for long hydrometeorological time series. *Hydrological Processes*, 22(23):4600–4608, 2008.
- D. E. Amos. Algorithm 610: A portable fortran subroutine for derivatives of the psi function. *ACM Transactions on Mathematical Software (TOMS)*, 9(4):494–502, 1983.
- E. Anderson, Z. Bai, C. Bischof, S. Blackford, J. Demmel, J. Dongarra, J. Du Croz, A. Greenbaum, S. Hammarling, and A. McKenney. LAPACK user’s guide third edition. *The Society for Industrial and Applied Mathematics*, 1999.
- T. W. Anderson. *The Statistical Analysis of Time Series*. John Wiley & Sons, 1971.
- T. W. Anderson. Goodness-of-fit tests for autoregressive processes. *Journal of Time Series Analysis*, 18(4):321–339, 1997.

- D. W. K. Andrews. Tests for parameter instability and structural change with unknown change point. *Econometrica: Journal of the Econometric Society*, 61:821–856, 1993.
- L. A. Aroian. A study of R. A. Fisher’s Z distribution and the related F distribution. *The Annals of Mathematical Statistics*, 12(4):429–448, 1941.
- M. S. Bartlett. Properties of sufficiency and statistical tests. *Proceedings of the Royal Society of London. Series A—Mathematical and Physical Sciences*, 160(901):268–282, 1937.
- M. S. Bartlett. *An Introduction to Stochastic Processes: With Special Reference to Methods and Applications*. Cambridge University Press, 1978.
- M. S. Bartlett and D. G. Kendall. The statistical analysis of variance-heterogeneity and the logarithmic transformation. *Supplement to the Journal of the Royal Statistical Society*, 8(1):128–138, 1946.
- C. Beaulieu, T. B. M. J. Ouarda, and O. Seidou. A Bayesian normal homogeneity test for the detection of artificial discontinuities in climatic series. *International Journal of Climatology*, 30(15):2342–2357, 2010.
- B. Bell, D. B. Percival, and A. T. Walden. Calculating Thomson’s spectral multitapers by inverse iteration. *Journal of Computational and Graphical Statistics*, 2(1):119–130, 1993.
- J. S. Bendat and A. G. Piersol. *Random Data: Analysis and Measurement Procedures*. John Wiley & Sons, 4th edition, 2011.

- R. B. Blackman and J. W. Tukey. *The Measurement of Power Spectra: From the Point of View of Communications Engineering*. Dover Publications New York, 1959.
- P. Bloomfield. *Fourier Analysis of Time Series*. John Wiley & Sons, 2nd edition, 2000.
- G. E. P. Box, G. M. Jenkins, and G. C. Reinsel. *Time Series Analysis: Forecasting and Control*. John Wiley & Sons, 3rd edition, 1994.
- D. R. Brillinger. *Time series: data analysis and theory*, volume 36. Siam, 2001.
- P. J. Brockwell and R. A. Davis. *Time series: theory and methods*. Springer-Verlag, 2nd edition, 1991.
- T. P. Bronez. On the performance advantage of multitaper spectral analysis. *Signal Processing, IEEE Transactions on [see also Acoustics, Speech, and Signal Processing, IEEE Transactions on]*, 40(12):2941–2946, 1992.
- E. N. Brown, R. E. Kass, and P. P. Mitra. Multiple neural spike train data analysis: State-of-the-art and future challenges. *Nature Neuroscience*, 7(5):456–461, 2004.
- R. L. Brown, J. Durbin, and J. M. Evans. Techniques for testing the constancy of regression relationships over time. *Journal of the Royal Statistical Society. Series B (Methodological)*, pages 149–192, 1975.
- J. P. Burg. A new analysis technique for time series data. *NATO Advanced Study Institute on Signal Processing with Emphasis on Underwater Acoustics (reprinted in Childers, 1978)*, 1, 1968.

- J. P. Burg. *Maximum Entropy Spectral Analysis*. PhD thesis, Stanford University, 1975.
- J. P. Burg, D. G. Luenberger, and D. L. Wenger. Estimation of structured covariance matrices. *Proceedings of the IEEE*, 70:963–974, 1982.
- G. C. Carter, C. Knapp, and A. H. Nuttall. Estimation of the magnitude-squared coherence function via overlapped fast Fourier transform processing. *IEEE Trans. on Audio and Acoustics*, 21(4):337–344, 1973a.
- G. C. Carter, C. H. Knapp, and A. H. Nuttall. Statistics of the estimate of the magnitude-coherence function. *IEEE Trans. on Audio and Acoustics*, 21:388–389, 1973b.
- F. E. Cave-Browne-Cave. On the influence of the time factor on the correlation between the barometric heights at stations more than 1000 miles apart. *Proc. Royal Soc. London*, 74:403–413, 1905.
- F. E. Cave-Browne-Cave and K. Pearson. On the correlation between the barometric heights on eastern side of the atlantic. *Proc. Royal Soc. London*, 70:465–470, 1902.
- J. P. Chabin, M. Madelin, and C. Bonnefoy. Les vignobles beaunois face au réchauffement climatique. In *Colloque Réchauffement climatique, quels impacts probables sur les vignobles*, 2007.
- C. Chatfield. *The Analysis of Time Series: An Introduction*. CRC Press, 2004.
- J. Chen and A. K. Gupta. *Parametric Statistical Change Point Analysis: With Applications to Genetics, Medicine, and Finance*. Springer, 2012.

- I. Chuine, P. Yiou, N. Viovy, B. Seguin, V. Daux, and E. L. R. Ladurie. Historical phenology: Grape ripening as a past climate indicator. *Nature*, 432:289–290, 2004.
- D. R. Cox and H. D. Miller. *The Theory of Stochastic Processes*. John Wiley & Sons, 1965.
- H. Cramér. On the theory of stationary random processes. *Ann. of Math.*, 41:215–230, 1940.
- H. Cramér and M. R. Leadbetter. *Stationary and Related Stochastic Processes: Sample Function Properties and Their Applications*. John Wiley & Sons, 1967.
- H. F. Davis. *Fourier Series and Orthogonal Functions*. Boston Allyn and Bacon, 1963.
- M. J. L. De Hoon, T. H. J. J. Van der Hagen, H. Schoonewelle, and H. Van Dam. Why Yule-Walker should not be used for autoregressive modelling. *Annals of Nuclear Energy*, 23(15):1219–1228, 1996.
- P. Diggle. *Time Series: a Biostatistical Introduction*. Oxford University Press, 1990.
- J. L. Doob. *Stochastic Processes*. John Wiley & Sons, 1952.
- J. A. Eddy. The Maunder minimum. *Science*, 192(4245):1189–1202, 1976.
- B. Efron and G. Gong. A leisurely look at the bootstrap, the jackknife, and cross-validation. *The American Statistician*, 37(1):36–48, 1983.
- B. Efron and C. Stein. The jackknife estimate of variance. *The Annals of Statistics*, 9(3):586–596, 1981.

- A. Einstein. Method for the determination of the statistical values of observations concerning quantities subject to irregular fluctuations. *ASSP Magazine, IEEE*, 4(4):6–6, 1987.
- R. A. Fisher. On a distribution yielding the error functions of several well known statistics. In *Proceedings of the International Congress of Mathematics*, volume 2, pages 805–813, Toronto, Canada, 1924.
- R. A. Fisher, J. H. Bennett, and F. Yates. *Statistical Methods, Experimental Design, and Scientific Inference*, volume 1. Oxford University Press New York, 1990.
- M. Frigo and S. G. Johnson. The design and implementation of FFTW3. *Proceedings of the IEEE*, 93(2):216–231, 2005. Special issue on Program Generation, Optimization, and Platform Adaptation.
- P. I. Good. *Resampling Methods: A Practical Guide to Data Analysis*. Birkhäuser, 2001.
- U. Grenander and M. Rosenblatt. Statistical spectral analysis of time series arising from stationary stochastic processes. *Annals of Math. Stat.*, 24:537–558, 1953.
- U. Grenander and M. Rosenblatt. *Statistical Analysis of Stationary Time Series*. John Wiley & Sons, New York, 1957.
- U. Grenander and M. Rosenblatt. *Statistical Analysis of Stationary Time Series*. Chelsea Publishing Company, New York, 2nd edition, 1984.
- A. Hanssen and L. L. Scharf. Polyspectra for harmonizable stochastic processes. *Signals, Systems and Computers, 2002. Conference Record of the Thirty-Sixth Asilomar Conference on*, 2, 2002.

- M. Hansson and G. Salomonsson. A multiple window method for estimation of peaked spectra. *Signal Processing, IEEE Transactions on [see also Acoustics, Speech, and Signal Processing, IEEE Transactions on]*, 45(3):778–781, 1997.
- F. J. Harris. On the use of windows for harmonic analysis with the discrete Fourier transform. *Proceedings of the IEEE*, 66:51–83, 1978.
- H. X. He and D. J. Thomson. The canonical bicoherence—part II: QPC test and its application in geomagnetic data. *Signal Processing, IEEE Transactions on*, 57(4):1285–1292, 2009.
- D. V. Hinkley. Improving the jackknife with special reference to correlation estimation. *Biometrika*, 65(1):13, 1978.
- G. M. Jenkins and D. G. Watts. *Spectral Analysis*. Holden-day, 1968.
- Gregory V. Jones and Gregory B. Goodrich. Influence of climate variability on wine regions in the western usa and on wine quality in the napa valley. *Climate Research*, 35(3):241, 2007.
- T. Kanamori, S. Hido, and M. Sugiyama. A least-squares approach to direct importance estimation. *The Journal of Machine Learning Research*, 10:1391–1445, 2009.
- T. R. Karl, R. W. Knight, and B. Baker. The record breaking global temperatures of 1997 and 1998: Evidence for an increase in the rate of global warming? *Geophysical Research Letters*, 27(5):719–722, 2000.

- M. Kaveh and G. A. Lippert. An optimum tapered Burg algorithm for linear prediction and spectral analysis. *IEEE Trans. on Acoustics, Speech, and Signal Processing*, ASSP-31:438–444, 1983.
- S. M. Kay. Noise compensation for autoregressive spectral estimates. *IEEE Trans. on Acoustics, Speech, and Signal Processing*, 28(3):292–303, 1980.
- S. M. Kay and S. L. Marple, Jr. Spectrum Analysis—A Modern Perspective. *Proceedings of the IEEE*, 69(11):1380–1419, 1981.
- A. S. Kayhan, A. El-Jaroudi, and L. F. Chaparro. Evolutionary periodogram for nonstationary signals. *Signal Processing, IEEE Transactions on*, 42(6):1527–1536, 1994.
- A. Khintchine. Korrelationstheorie der stationären stochastischen Prozesse. *Mathematische Annalen*, 109(1):604–615, 1934.
- A. Khodadadi and M. Asgharian. Change-point problem and regression: an annotated bibliography. *COBRA Preprint Series, Paper*, 44, 2008.
- W. Koenig, H. K. Dunn, and L. Y. Lacy. The sound spectrograph. *J. Acoustical Soc. Amer.*, 18:19–49, 1946.
- L. H. Koopmans. *The Spectral Analysis of Time Series*, volume 22. Academic Press, 1995.
- M. Krieger, G. Lohmann, and T. Laepple. Seasonal climate impacts on the grape harvest date in Burgundy (France). *Climate of the Past*, 7(2):425–435, 2011.

- N. A. Krivova, L. E. A. Vieira, and S. K. Solanki. Reconstruction of solar spectral irradiance since the Maunder minimum. *Journal of Geophysical Research: Space Physics*, 115(A12), 2010.
- C. Kuo, C. Lindberg, and D. J. Thomson. Coherence established between atmospheric carbon dioxide and global temperature. *Nature*, 343(6260):709–713, 1990.
- J. R. Lanzante. Resistant, robust and non-parametric techniques for the analysis of climate data: Theory and examples, including applications to historical radiosonde station data. *International Journal of Climatology*, 16(11):1197–1226, 1996.
- J. M. Lees. *RSEIS: Seismic Time Series Analysis Tools*, 2013. R package version 3.2-1.
- J. M. Lees and J. Park. Multiple-taper spectral analysis: A stand-alone c-subroutine. *Computers & Geosciences*, 21(2):199–236, 1995.
- K. Q. Lepage and D. J. Thomson. Spectral analysis of cyclostationary time-series: a robust method. *Geophysical Journal International*, 179(2):1199–1212, 2009.
- C. R. Lindberg. *Multiple taper spectral analysis of terrestrial free oscillations*. PhD thesis, Univ. Calif., San Diego, 1986.
- C. R. Lindberg and J. Park. Multiple-taper spectral analysis of terrestrial free oscillations. II. *Geophysical Journal of the Royal Astronomical Society*, 91(3):795–836, 1987.
- S. Liu, M. Yamada, N. Collier, and M. Sugiyama. Change-point detection in time-series data by relative density-ratio estimation. *Neural Networks*, 43:72–83, 2013.

- M. M. Loève. Fonctions aléatoires de second ordre. *Les Comptes Rendus de l'Académie des Sciences, Paris*, 222:942–944, 1946.
- G. Manley. Central england temperatures: monthly means 1659 to 1973. *Quarterly Journal of the Royal Meteorological Society*, 100(425):389–405, 1974.
- H. B. Mann and D. R. Whitney. On a test of whether one of two random variables is stochastically larger than the other. *The annals of mathematical statistics*, 18(1): 50–60, 1947.
- M. E. Mann and J. M. Lees. Robust estimation of background noise and signal detection in climatic time series. *Climatic Change*, 33(3):409–445, 1996.
- Hadley Climate Research Unit. HadCRUT anomaly series, 2011.
- NOAA. ESRL global monitoring division: CO₂ data, 2011.
- United Kingdom Meteorological Office. Hadley Centre Central England Temperature (HADCET) dataset, 2011.
- Y. Mei. Sequential change-point detection when unknown parameters are present in the pre-change distribution. *The Annals of Statistics*, pages 92–122, 2006.
- N. Meier, T. Rutishauser, C. Pfister, H. Wanner, and J. Luterbacher. Grape harvest dates as a proxy for Swiss April to August temperature reconstructions back to AD 1480. *Geophys. Res. Lett*, 34, 2007.
- M. J. Menne. Abrupt global temperature change and the instrumental record. In *18th Conference on Climate Variability and Change*, Atlanta, GA, 2006. American Meteorological Society.

- K. S. Miller. *Complex Stochastic Processes: an Introduction to Theory and Application*. Addison-Wesley Publishing Company, Advanced Book Program, 1974.
- A. Moghtaderi. *Multitaper Methods for Time-Frequency Spectrum Estimation and Unaliasing of Harmonic Frequencies*. PhD thesis, Queen's University, 2009.
- I. C. Moore, D. P. Strum, L. G. Vargas, and D. J. Thomson. Observations on surgical demand time series: Detection and resolution of holiday variance. *Anesthesiology*, 109(3):408–416, 2008.
- M. Mudelsee. *Climate Time Series Analysis: Classical Statistical and Bootstrap Methods*, volume 42. Springer, 2010.
- C. T. Mullis and L. L. Scharf. Quadratic estimators of the power spectrum. In S Haykin, editor, *Advances in Spectrum Estimation*, volume 1, chapter 1, pages 1–57. Prentice Hall, 1991.
- H. Nyquist. Certain topics in telegraph transmission theory. *Transactions of the American Institute of Electrical Engineers*, 47(2):617–644, 1928.
- A. Papoulis and S. U. Pillai. *Probability, random variables and stochastic processes*. McGraw-Hill, 4th edition, 2001.
- E. Pardo-Igúzquiza, M. Chica-Olmo, and F. J. Rodríguez-Tovar. CYSTRATI: a computer program for spectral analysis of stratigraphic successions. *Computers & Geosciences*, 20(4):511–584, 1994.
- J. Park, C. R. Lindberg, and D. J. Thomson. Multiple-taper spectral analysis of terrestrial free oscillations. I. *Geophysical Journal of the Royal Astronomical Society*, 91(3):755–794, 1987.

- D. Parker and B. Horton. Uncertainties in Central England temperature 1878–2003 and some improvements to the maximum and minimum series. *International Journal of Climatology*, 25(9):1173–1188, 2005.
- D. E. Parker, T. P. Legg, and C. K. Folland. A new daily Central England temperature series, 1772–1991. *International Journal of Climatology*, 12(4):317–342, 1992.
- K. Pearson. On the criterion that a given system of deviations from the probable in the case of a correlated system of variables is such that it can be reasonably supposed to have arisen from random sampling. *Philosophical Magazine Series 5*, 50(302):157–175, 1900.
- D. B. Percival and A. T. Walden. *Spectral Analysis for Physical Applications*. Cambridge University Press New York, NY, USA, 1993.
- A. N. Peristyky and P. E. Damon. Persistence of the Gleissberg 88-year solar cycle over the last $\sim 12,000$ years: Evidence from cosmogenic isotopes. *J Geophys. Res.*, 108:A1, 2003. doi: 10.1029/2002JA009390.
- A. N. Pettitt. A non-parametric approach to the change-point problem. *Applied Statistics*, 28:126–135, 1979.
- D. Picard. Testing and estimating change-points in time series. *Advances in Applied Probability*, 17(4):841–867, 1985.
- J. W. Pillow, Y. Ahmadian, and L. Paninski. Model-based decoding, information estimation, and change-point detection techniques for multineuron spike trains. *Neural Computation*, 23(1):1–45, 2011.

- W. H. Press, S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery. *Numerical Recipes in C++: The Art of Scientific Computing Third Edition*. Cambridge University Press, 2007.
- M. B. Priestley. *Spectral Analysis and Time Series. Volume 1: Univariate Series. Volume 2: Multivariate Series, Prediction and Control*. Probability and Mathematical Statistics, 1981.
- G. A. Prieto, R. L. Parker, and F. L. Vernon, III. A FORTRAN 90 library for multitaper spectrum analysis. *Computers & Geosciences*, 35(8):1701–1710, 2009.
- R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2013.
- K. J. Rahim and D. J. Thomson. Practical test for goodness of fit of low-order ar models. In *JSM Proceedings, Section on Statistics and the Environment*, pages 3821–3834, Montréal, Canada, 2013. American Statistical Society.
- Lord Rayleigh. On the spectrum of an irregular disturbance. *Philosophical Magazine*, 41:238–243, 1903. (in *Scientific Papers by Lord Rayleigh*, Volume V, Article 285, pages 98-102, Dover Publications, New York, 1964).
- K. S. Riedel and A. Sidorenko. Minimum bias multiple taper spectral estimation. *Signal Processing, IEEE Transactions*, 43(1):188–195, 1995.
- E. Ruggieri. A Bayesian approach to detecting change points in climatic records. *International Journal of Climatology*, 33(2):520–528, 2012.
- E. Ruggieri, T. Herbert, K. T. Lawrence, and C. E. Lawrence. Change point method

- for detecting regime shifts in paleoclimatic time series: Application to $\delta^{18}\text{O}$ time series of the Plio-Pleistocene. *Paleoceanography*, 24(1), 2009.
- A. Schuster. On the investigation of hidden periodicities with application to a supposed 26 day period of meteorological phenonema. *Terrestrial Magnetism*, 3:13–41, 1898.
- O. Seidou and T. B. M. J. Ouarda. Recursion-based multiple changepoint detection in multiple linear regression and application to river streamflows. *Water Resources Research*, 43(7), 2007.
- R. H. Shumway and D. S. Stoffer. *Time Series Analysis and its Applications: With R Examples*. Springer, 2nd edition, 2006.
- R. H. Shumway and D. S. Stoffer. *Time Series Analysis and its Applications: With R examples*. Springer, 3rd edition, 2010.
- D. Slepian. Prolate spheroidal wave functions, Fourier analysis and uncertainty–IV. *Bell Syst. Tech. J*, 43:3009–3057, 1964.
- D. Slepian. On bandwidth. *Proceedings of the IEEE*, 64(3):292–300, 1976.
- D. Slepian. Prolate spheroidal wave functions, Fourier analysis, and uncertainty. V—the discrete case. *Bell Syst. Tech. J*, 57:1371–1430, 1978.
- D. Slepian. Some comments on Fourier analysis, uncertainty and modeling. *SIAM Review*, 25(3):379–393, 1983.
- D. Slepian and H. O. Pollak. Prolate spheroidal wave functions, Fourier analysis and uncertainty–I. *Bell Syst. Tech. J*, 40:43–64, 1961.

- T. F. Stocker, Q. Dahe, and G. K. Plattner. Climate change 2013: The physical science basis. *Working Group I Contribution to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change. Summary for Policymakers (IPCC, 2013)*, 2013.
- P. Stoica and T. Sundin. On nonparametric spectral estimation. *Circuits, Systems, and Signal Processing*, 18(2):169–181, 1999.
- G. G. Stokes. On a method of detecting inequalities of unknown periods in a series of observations. *Proc. R. Soc. London*, 29:122–123, 1879. Comment on the Preliminary Report to the Committee on Solar Physics, In *Mathematical and Physical Papers, V* Cambridge University Press, 1905, pages 52–53.
- G. Strang. *Linear Algebra and Its Applications Academic*. Cengage Learning, 4th edition, 2005.
- A. Stuart and J. K. Ord. *Kendall's Advanced Theory of Statistics Vol. 1: Distribution theory*. Wiley, 6th edition, 2010.
- H. E. Suess and T. W. Linick. The ^{14}C record in Bristlecone pine wood of the past 8000 years based on the dendrochronology of the late C. W. Ferguson. *Philosophical Transactions of the Royal Society of London. Series A, Mathematical and Physical Sciences*, 330(1615):403–412, 1990.
- D. J. Thomson. Spectrum estimation techniques for characterization and development of WT4 waveguide. *Bell Syst. Tech. J*, 56:1769–1815, 1977a.
- D. J. Thomson. Spectrum estimation techniques for characterization and development of WT4 waveguide–II. *Bell Syst. Tech. J*, 56:1983–2005, 1977b.

- D. J. Thomson. Spectrum estimation and harmonic analysis. *Proceedings of the IEEE*, 70(9):1055–1096, 1982.
- D. J. Thomson. Jackknifed estimates of line spectrum parameters. In *ISIT '84 International Symposium on Information Theory*, Brighton, 1984. IEEE. Abstract 292.
- D. J. Thomson. Quadratic-inverse spectrum estimates: Applications to palaeoclimatology. *Philosophical Transactions: Physical Sciences and Engineering*, 332(1627):539–597, 1990a.
- D. J. Thomson. Time series analysis of Holocene climate data. *Philosophical Transactions of the Royal Society of London. Series A, Mathematical and Physical Sciences*, 330(1615):601–616, 1990b.
- D. J. Thomson. The seasons, global temperature, and precession. *Science*, 268(5207):59, 1995.
- D. J. Thomson. Multiple-window spectrum estimates for non-stationary data. In *Proc. Ninth IEEE SP Workshop on Statistical Signal and Array Processing*, pages 344–347, Portland, Oregon, 1998. IEEE.
- D. J. Thomson. Multitaper analysis of nonstationary and nonlinear time series data. In W. Fitzgerald, R Smith, A Walden, and Young P, editors, *Nonlinear and Nonstationary Signal Processing*, pages 317–394. Cambridge University Press, 2001.
- D. J. Thomson. Jackknifing multitaper spectrum estimates. *Signal Processing Magazine, IEEE*, 24(4):20–30, 2007.

- D. J. Thomson and A. D. Chave. Jackknifed error estimates for spectra, coherences, and transfer functions. In S. Haykin, editor, *Advances in Spectrum Analysis and Array Processing*, volume 1, chapter 2, pages 58–113. Prentice-Hall, Upper Saddle River, NJ, 1991a.
- D. J. Thomson and A. D. Chave. Jackknifed error estimates for spectra, coherences, and transfer functions. In S. Haykin, editor, *Advances in Spectrum Estimation*, volume 1, chapter 2, pages 58–113. Prentice Hall, 1991b.
- D. J. Thomson, L. J. Lanzerotti, F. L. Vernon, III, M. R. Lessard, and L. T. P. Smith. Solar Modal Structure of the Engineering Environment. *Proceedings of the IEEE*, 95(5):1085–1132, 2007.
- A. R. Tomé and P. M. A. Miranda. Piecewise linear fitting and trend changing points of climate parameters. *Geophysical Research Letters*, 31(2), 2004.
- Y. M. Tourre, D. Rousseau, L. Jarlan, E. Le Roy Ladurie, and V. Daux. Western European climate, and Pinot noir grape harvest dates in Burgundy, France, since the 17th century. *Climate Research*, 46(3):243, 2011.
- R. S. Tsay. Testing and modeling threshold autoregressive processes. *Journal of the American Statistical Association*, 84(405):231–240, 1989.
- J. W. Tukey. Discussion, emphasizing the connection between analysis of variance and spectrum analysis. *Technometrics*, 3(2):191–219, 1961.
- J. W. Tukey. Styles of spectrum analysis. In *A Celebration in Geophysics and Oceanography – 1982 In Honor of Walter Munk*, Reference Series 84-5, March, 1984, pages 100–103. Scripps Institution of Oceanography, La Jolla, CA, 1984.

- Pages 1143-1153 of *The Collected Works of J. W. Tukey* Vol II, D. R. Brillinger, Ed., Wadsworth, Monterey, Ca., 1984.
- R. K. Ulrich. The five-minute oscillations on the solar surface. *The Astrophysical Journal*, 162:993–1002, 1970.
- T. J. Ulrych and Thomas N. Bishop. Maximum entropy spectral analysis and autoregressive decomposition. *Reviews of Geophysics*, 13(1):183–200, 1975.
- L. E. A. Vieira, A. Norton, T. Dudok de Wit, M. Kretzschmar, G. A. Schmidt, and M. C. M. Cheung. How the inclination of Earth’s orbit affects incoming solar irradiance. *Geophysical Research Letters*, 39(16), 2012.
- P. D. Welch. The use of the fast fourier transform for estimation of spectra: A method based on time averaging over short, modified periodogram. *IEEE Trans. on Audio and Acoustics*, 15:70–74, 1967a.
- P. D. Welch. The use of fast Fourier transform for the estimation of power spectra: a method based on time averaging over short, modified periodograms. *IEEE Trans. on Audio and Acoustics*, 15(2):70–73, 1967b.
- B. Whitcher. *waveslim: Basic Wavelet Routines for One-, Two- and Three-dimensional Signal Processing*, 2012. R package version 1.7.1.
- R. Wilson, E. Cook, R. D’Arrigo, N. Riedwyl, M. N. Evans, A. Tudhope, and R. Allan. Reconstructing ENSO: The influence of method, proxy data, climate forcing and teleconnections. *Journal of Quaternary Science*, 25(1):62–78, 2010.
- C. F. J. Wu. Jackknife, bootstrap and other resampling methods in regression analysis. *The Annals of Statistics*, 14(4):1261–1295, 1986.

- A. Yaglom. Einstein's 1914 paper on the theory of irregularly fluctuating series of observations. *ASSP Magazine, IEEE*, 4(4):7–11, 1987a.
- A. M. Yaglom. *Correlation Theory of Stationary and Related Random Functions: Vol.: 1: Basic Results*. Springer-Verlag, 1987b.
- G. U. Yule. On the theory of correlation. *J Roy Statist Soc*, 60:249–295, 1897.
- G. U. Yule. On the theory of correlation for any number of variables, treated by a new system of notation. *Proceedings of the Royal Society of London. Series A*, 79 (529):182–193, 1907.
- S. Zacks. Classical and Bayesian approaches to the change-point problem: Fixed sample and sequential procedures. *Statistique et analyse des données*, 7(1):48–81, 1982.

Appendix A

Multitaper R Package

A.1 Appendix Overview

This appendix represents a paper submitted to *Journal of Statistical Software* in July of 2012, describing a R package called “Multitaper,” available on (CRAN). There have been contributions to the code by Wesley Burr and David Thomson, who are co-authors of this paper. At that time it was recommended for publication after major revisions. The revisions were required to the paper, software package and code documentation. This appendix constitutes the current state of the paper, which has undergone the majority of revisions, since submission. Additionally, the code and documentation have been revised since submission. We have not yet resubmitted the paper. We note there may be overlap between the theory sections of this paper and Chapter 2 of the thesis.

A.2 Introduction

Spectral analysis is used by statisticians and researchers to analyze sequential data referred to as time series. Examples of a time series include digitized recorded speech, a sequential record of stock prices, and an electrocardiogram, which is a record of electrical signals from a patient’s heart. The term time series implies successive observations in time, creating a serial correlation, but time series analysis techniques apply to observations related sequentially, even if the sequential relationship is not time. Spectral analysis refers to techniques involving analysis of a representation of the time series in terms of sinusoidal components. One can imagine projecting a time series onto the space spanned by sinusoids of discrete frequencies from zero to a cutoff frequency, and then analyzing the coefficients, or squared coefficients, at each

frequency to determine which frequencies contribute more to the variance of the original series. This projection image is accurate because in spectral analysis, as in linear regression, we consider equality in a mean-square sense. Multitaper spectral analysis is a form of spectral analysis that exploits certain optimal orthogonal sequences, discrete prolate spheroidal sequence (Slepian sequences), to produce consistent, in the statistical sense, spectral estimates with lower bias and variance than the naïve estimator, which is called the periodogram.

We present a package for the R (R Core Team, 2013) statistical programming language that performs multitaper spectral estimation. In addition, the package implements techniques that exploit properties of multitaper spectral estimators using Slepian sequences to provide: a jackknife (non-parametric) variance; a harmonic F -test, a statistical technique for detecting single-frequency line components; a magnitude-squared coherence (MSC) estimate, an improved technique for analyzing a linear dependence in frequency of bivariate time series which includes a jackknifed variance estimate; and a complex demodulate estimate, a technique for observing phase drift, a slow change in frequency over time.

While this paper is self-contained, the authors recommend some familiarity with time series analysis and spectral estimation. Two comprehensible reference texts that include introductory discussions of spectral analysis are Chatfield (2004) and Diggle (1990). Percival and Walden (1993), hereinafter abbreviated as P&W93, present a thorough overview of multitaper spectral estimation theory with many examples. Shumway and Stoffer (2010) present a comprehensive book covering spectral and time series analysis using the R programming language.

Thomson (1982) introduced multitaper spectral estimates using Slepian sequences,

and in the interim this technique has been used in fields such as anesthesiology (Moore et al., 2008), climate science (Tourre et al., 2011), geophysics (He and Thomson, 2009; Lepage and Thomson, 2009), and neuroscience (Brown et al., 2004).

The multitaper spectral estimate is similar to direct spectrum estimates (Blackman and Tukey, 1959), which reduce bias by applying a data taper. It improves over direct spectral estimates in two ways: (1) it makes use of Slepian sequences, which are maximally concentrated in time and frequency (P&W93, pp. 75–81), and (2) it uses several orthogonal Slepian sequences averaging estimates. Typically, one uses an adaptive weighted average to reduce variance while controlling bias. The cost of using this method is (1) a reduction in frequency resolution and (2) increased computational cost, as multiple Fourier transforms are required in place of one. The computational burden can be measured in fractions of a second and should not be a primary concern. The direct spectral estimator controls bias with one taper (Blackman and Tukey, 1959), thus decreasing frequency resolution, and additionally requires smoothing or frequency averaging to increase variance, again decreasing bandwidth. The direct spectral estimator, without frequency averaging, is not statistically consistent, as the variance does not decrease as the sample size increases. The periodogram can be shown to be asymptotically unbiased, but examples exist where considerable bias is observed with a high number of data points (Thomson, 1982, p. 1058)

There are several software packages and programs that implement the multitaper method, and we present a brief review. The programming environment **MATLAB** implements Thomson’s multitaper method, using the adaptive weights, with the “signal processing toolbox.” Code written in C++, available in Press et al. (2007, pp. 662–667), can be used to obtain a multitaper spectral estimate; however, adaptive weights

are not implemented. Pardo-Igúzquiza et al. (1994) introduce a Fortran program that implements the multitaper method using adaptive weights. Lees and Park (1995) present C code implementing the multitaper method with adaptive weighting and the harmonic F -test; however, there is no option to zero-pad to increase the frequency grid. Fortran 90 code implementing the adaptive weighted multitaper spectral estimate and the harmonic F -test is provided in Prieto et al. (2009). LISP code implementing the adaptive weighted multitaper spectral estimate and the harmonic F -test. Some functionality in our multitaper package is based on the LISP code accompanying P&W93. The following packages are available in the programming language R. The package `waveslim` (Whitcher, 2012) obtains the Slepian sequences using the accurate inverse iteration method (Bell et al., 1993). The package `sapa` calculates the multitaper spectral estimate, but without using adaptive weights. The package `RSEIS` implements the multitaper method using adaptive weights and it computes the harmonic F -test (Lees, 2013). We present the `multitaper` package, which implements the multitaper method, allows for adaptive weights, and implements the harmonic F -test. This package adds the ability to obtain the nonparametric, jackknife variance of the spectral estimate, the bivariate MSC with a jackknife estimate, and complex demodulation using the Slepian sequences. The programming environment `S-Plus` provided a native function to calculate complex demodulation based on Bloomfield (2000, pp. 97–130); however, R (as of the development version 3.1.0) does not provide a similar function. To accommodate R users, function calls in the `multitaper` package are *designed to be similar to* existing R spectral estimate calls (see (Shumway and Stoffer, 2010), and the functions return similar objects. We note that this package makes use of well tested Fortran routines developed in Thomson (1982, pp. 219–220).

The `multitaper` package is available from the Comprehensive R Archive Network at <http://cran.r-project.org/web/packages/multitaper>. In this appendix, we briefly detail the theory behind multiple-taper estimation in Sections A.3.1, A.3.2 and A.3.3. In Section A.4.1, we explore the application of jackknifing over tapers for estimating the variance of the estimator. Sections A.4.2 and A.5 give background theory for the two most commonly used extensions of multitaper theory, the harmonic F -test statistic and the magnitude-squared coherence, while Section A.6 explores the derivation for the complex demodulate. Each theory section is followed by worked examples demonstrating the functionality of the package. Section A.3.4 presents a classic AR(4) simulation spectrum, Section A.4.3 shows the application of the harmonic F -test to the Central England Temperature time series, Section A.5.1 reproduces a magnitude-squared coherence example and Section A.6.1 continues the example by examining the phase characteristic of the yearly periodicity. Finally, Section A.7.1 reviews several tools included in the package not included in the other sections, and Section A.7.2 gives some tips for extending the package with additional functionality.

A.3 The Theory of Multitaper Spectral Estimation

A.3.1 Overview

Direct spectral estimates are estimates of the spectrum computed via a direct Fourier transform, on tapered data, as compared to indirect estimates, which are obtained by

taking the Fourier transform of the autocovariance function via the Einstein-Wiener-Khintchine theorem, also known as the Wiener-Khintchine theorem. Multitaper spectral estimation is a technique that uses the weighted average of several direct spectral estimates, each computed using a different member of a family of orthogonal tapers. By default, we will assume that multitaper spectral estimates are computed using the Slepian sequences, which have been shown to be maximally concentrated in both time and frequency (Slepian and Pollak, 1961; Slepian, 1964, 1976, 1978, 1983), as tapers. That is, they define the classical uncertainty principles when time, frequency or both are limited. There are several other taper options available, including the sine tapers, that are also implemented in the `multitaper` package.

The key advantages of multitaper spectral estimation are as follows: first the availability of the Slepian tapers, as they are maximally concentrated in both time and frequency; second, the higher degrees of freedom obtained by use of multiple orthogonal tapers; and third, an optimal weighting scheme for combining the approximately orthogonal spectrum estimates into an approximately maximum-likelihood estimate of the spectrum. Two further advantages of using the Slepian tapers over other choices are the existence of the harmonic F -test statistic and the jackknife estimation of variance, also implemented in this package.

A.3.2 Parameters

If we begin with a time series $\{x_t\}_{t=0}^{N-1}$, in order to compute the multitaper spectral estimate we must select two initial parameters: the time-bandwidth parameter, denoted as NW (where W is the bandwidth over which the Slepian tapers have been concentrated), and the number of tapers, denoted as K . Typically, one selects

$K \in [2NW - 3], \dots, [2NW]$, with $K = [2NW - 1]$ being a reasonable first choice. In making the choice for bandwidth W , the analyst assumes that for each frequency f_0 , the signal is concentrated within a band $(f_0 - W, f_0 + W)$. W is typically chosen in half-integer steps between 2.0 and 6.0, with 4.5 or 5.0 being reasonable first choices for large data-sets. For more information, the reader is directed to Slepian (1978) and Percival and Walden (1993). In the case that no parameters are specified, the `spec.mtm` routine¹ will automatically default to $NW = 4.0$ and $K = 7$. The basic trade-off, however, is that larger W s allow larger dynamic range² and, in more common data, give higher ($\sim 4NW$) degrees-of-freedom.

A.3.3 Multitaper Spectral Estimates

We begin with N discrete measurements $\{x_t\}_{t=0}^{N-1}$ of a realization of a stationary process. Take the classic Cramér representation (Grenander and Rosenblatt, 1953; Doob, 1952) for a discrete stationary stochastic process,

$$x_t = \int_{-1/2}^{1/2} e^{i2\pi n f} dX(f), \quad (\text{A.1})$$

where dX is an orthogonal increment process. The spectrum (or power spectrum) is defined as $S_{xx}(f)df = \mathbf{E}[|dX(f)|^2]$. This gives the inverse problem—namely estimating $S(f)$ given $\{x_t\}$. As has been previously shown (Grenander and Rosenblatt, 1957;

¹This routine in the `multitaper` package is used to compute multitaper spectral estimates.

²Dynamic range (in decibels) is defined as

$$10 \log_{10} \left(\frac{\max_f S(f)}{\min_f S(f)} \right).$$

Mullis and Scharf, 1991; Bronez, 1992; Stoica and Sundin, 1999), every quadratic estimator of the power spectrum must have the form

$$\hat{S}(f) = \sum_{j,k=0}^{N-1} q_{j,k} e^{i2\pi(j-k)f} x_j x_k \quad (\text{A.2})$$

where the $q_{j,k}$ form a symmetric, frequency-independent matrix \mathbf{Q} of order N . If we then approximate $\mathbf{Q} = [q_{j,k}]$ by keeping the K largest eigenvalues,

$$\mathbf{Q} = \sum_{k=1}^K \mu_k v^{(k)} (v^{(k)})^T \quad (\text{A.3})$$

where $\{v^{(0)}, v^{(2)}, \dots, v^{(K-1)}\}$ are an orthogonal family of eigenvectors of \mathbf{Q} , we obtain a multitaper representation of the quadratic spectral estimator as

$$\hat{S}(f) = \sum_{k=0}^{K-1} \mu_k \left| \sum_{n=0}^{N-1} x_n v_n^{(k)} e^{-i2\pi n f} \right|^2 \quad (\text{A.4})$$

where $v_n^{(k)}(N, W)$ or just $v_n^{(k)}$ is Slepian's notation for the DPSSs with the time index shifted by 1. When $K = 1$, this becomes the familiar direct estimate, and if $NW = 0$ so $v_n^{(0)} = N^{-1/2}$, it is the periodogram.

Formally, the components of the spectral estimator written

$$y_k(f) = \sum_{n=0}^{N-1} x_n v_n^{(k)} e^{-i2\pi f n} \quad (\text{A.5})$$

are called the *eigencoefficients* and are the discrete Fourier transform of the data multiplied by the k^{th} discrete taper. In the classic development, these tapers are the discrete prolate spheroidal sequence. As each eigencoefficient is computed by transforming the data multiplied by the k^{th} data window $v_n^{(k)}$, their absolute squares are individually direct spectrum estimates and are referred to as the k^{th} *eigenspectrum*. The Fourier transforms of the eigenvectors (tapers) $v^{(0)}, \dots, v^{(K-1)}$ alone are written

as $\{V_k(f)\}_{k=0}^{K-1}$. These functions are odd and even as k is odd or even, and have k zeroes in $(-W, W)$. Slepian (1978) writes

$$U_k(f) = \epsilon_k \sum_{n=0}^{N-1} v_n^{(k)} e^{i2\pi f(n - \frac{N-1}{2})} \quad (\text{A.6})$$

where $\epsilon_k = 1$ if k is even, and $\epsilon_k = i$ for k odd. We instead use the notation from Thomson (1982) which has

$$V_k(f) = \sum_{n=0}^{N-1} v_n^{(k)} e^{-i2\pi f n}, \quad (\text{A.7})$$

which is complex-valued, and more useful as it directly represents the form taken by an FFT implementation.

In `multitaper`, these eigencoefficients are obtainable by the user by setting the parameter `returnInternals=TRUE` in the `spec.mtm` call. We will show that the eigencoefficients are central to the tools that can be developed in the `multitaper` arena, and thus are critical for extensibility of this package.

The \mathbf{Q} matrix from Equation (A.3) was not specified, and the choice

$$q_{nm} = \frac{\sin 2\pi W(n - m)}{\pi(n - m)} \quad (\text{A.8})$$

that gives the best concentration (jointly) in time and frequency is used by default. The user should note that the default tapers (or windows) used in `multitaper` are the Slepian tapers, and that each Slepian taper has a corresponding eigenvalue λ_k that represents the concentration of that taper within the band $(-W, W)$. For low-order tapers, $\lambda_k \approx 1$ (although bounded above), with the concentration decreasing as k approaches $2NW$.

In Equation (A.4), the absolute squares of the eigencoefficients are combined additively, each weighted by a corresponding μ_k . One version of the spectral estimator

takes $\mu_k = 1$, i.e., the simple arithmetic average. There is an improved, adaptively weighted, version of the estimator, as detailed in Thomson (1982), which is the default for the `multitaper` package. The idea is that as the eigenvalues decrease (as k increases), the bias characteristics also degrade, since $(1 - \lambda_k)$ is the fraction of energy in the k^{th} Slepian function outside the band $(-W, W)$. To counter this, we apply adaptive weighting to the eigenspectra, thus decreasing contributions from the higher order eigenspectra in regions where the spectrum is small. In this form, the spectral density function is a solution of the equation

$$\sum_{k=0}^{K-1} \frac{\lambda_k \left(\hat{S}(f) - \hat{S}_k(f) \right)}{\left(\lambda_k \hat{S}(f) - \hat{B}_k(f) \right)^2} \quad (\text{A.9})$$

where $\hat{B}_k(f)$ is an approximation to the broad-band bias term, usually approximated as $\sigma^2(1 - \lambda_k)$. In practice, this equation is solved iteratively by using the average of $\hat{S}_0(f)$ and $\hat{S}_1(f)$ as a start point and iterating. The solution is positive and lies between the minimum and maximum of the $\hat{S}_k(f)$ s. Convergence is typically rapid, requiring no more than a few dozen iterations. The maximum number of iterations is user-tunable by setting the `maxAdaptiveIterations` parameter in the `spec.mtm` call.

A.3.4 Multitaper Example of an AR(4) Process

In this section, we demonstrate a basic example of multitaper spectral estimation, using the `multitaper` package. We use an AR 4 process with the AR coefficients

$$\phi = (2.7607, -3.8106, 2.6535, -0.9233)^T$$

as suggested in Percival and Walden (1993, p. 46), and analyzed throughout that text. Using the above coefficients, we can calculate the theoretical spectra as

$$S(f) = \frac{\sigma_w^2}{|1 - \sum_{i=1}^4 \phi_i e^{-i2\pi f}|^2}, \quad (\text{A.10})$$

where σ_w^2 is defined as the innovations variance. The following R code generates a realization of this AR(4) time series with standard normal innovations, loads the multitaper library, and displays the multitaper spectral estimate of the series similar to that in Figure A.1. Note that the following R commands show only the estimated multitaper spectrum, while Figure A.1 also includes the theoretical spectrum.

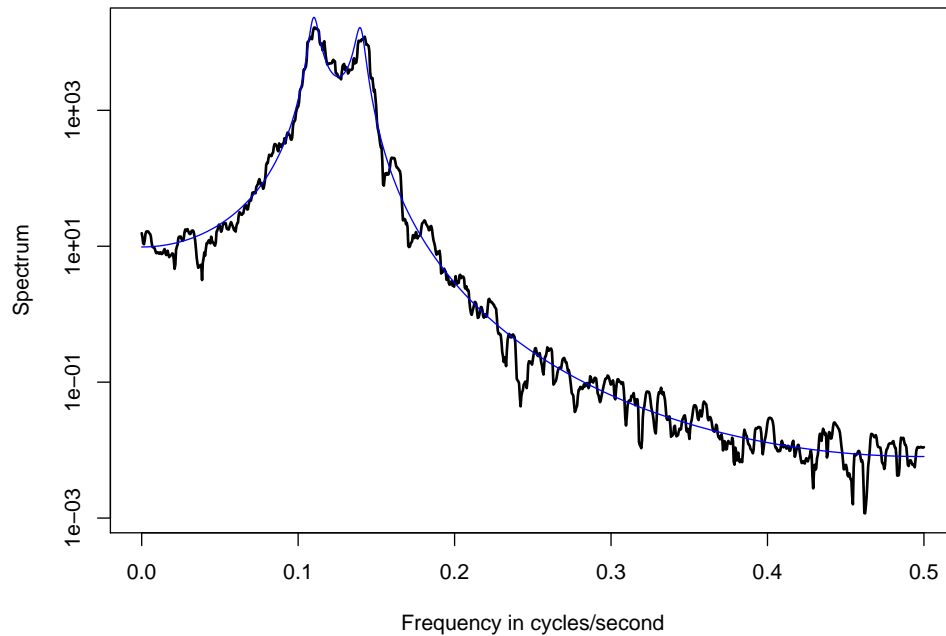


Figure A.1: Adaptive multitaper spectrum of the realization of an AR(4) time series (thick lines) plotted on top of the theoretical spectrum (thin line).

Function 1. spec.mtm

```
R> library("multitaper")
R> ar4Coef <- c(2.7607, -3.8106, 2.6535, -0.9238)
R> set.seed(60)
R> ar4.ts <- arima.sim(list(order = c(4, 0, 0), ar = ar4Coef),
+   n = 1024)
R> spec.mtm(ar4.ts, nw = 4, k = 8, dtUnits = "second")
```

A.3.4.1 Basic Options

The above command used a time-bandwidth parameter of $nw = 4$ with $k = 7$ tapers.³ It also used adaptive weighting on the tapers and used the Slepian sequences to centre (i.e., remove the mean of) the data as set by default.

In this case, with 1024 samples and a sampling period of $\Delta T = 1$, we find the bandwidth $W = 4/(1024\Delta T) = 0.00390625$. The reader will notice that we did not give units for the sample or the sampling period. If we were sampling at 1/sec, then the bandwidth could be written as 0.004 Hz (cycles/second). The frequency axis description can be modified by setting `dtUnits` as shown in this example.

³In this paper we use uppercase NW and K , but our software packages uses lowercase variables `nw` and `k`, and all listed code used the appropriate case.

A.4 Addressing Statistical Significance with Multitaper Tools

A.4.1 Jackknifing Multitaper Spectral Estimates

The jackknife is a classic statistical tool, covered in Efron and Gong (1983) and fully applied to the multitaper spectrum estimate in Thomson and Chave (1991b), with a more approachable overview in Thomson (2007). This tool is fully implemented in the `multitaper` package, and briefly reviewed here.

To jackknife multitaper spectrum estimates, begin with Equation (A.9), omit the j^{th} eigencoefficient from the weight, and take $\theta_{\setminus j} = \ln \hat{S}_{\setminus j}(f)$ (where the subscript $\setminus j$ is read in the set-theoretic meaning of “without j ”) at each frequency, where $\hat{\theta}_{\setminus j} = \{x_1, \dots, x_{j-1}, x_{j+1}, \dots, x_K\}$ denotes the estimate of the parameter θ omitting the j^{th} observation. This action treats the eigencoefficients as exchangeable data and is called “jackknifing over tapers.” We then compute the delete-one log-spectrum estimates as

$$\ln \hat{S}_{\setminus j}(f) = \ln \left[\frac{1}{K-1} \sum_{k=0, k \neq j}^{K-1} \hat{S}_k(f) \right]. \quad (\text{A.11})$$

Taking the average of these estimates,

$$\ln \hat{S}_{\setminus \bullet}(f) = \frac{1}{K} \sum_{j=0}^{K-1} \hat{S}_{\setminus j}(f), \quad (\text{A.12})$$

we then compute the variance estimate as

$$\hat{V}_J(f) = \frac{K-1}{K} \sum_{j=0}^{K-1} \left[\ln \hat{S}_{\setminus j} - \ln \hat{S}_{\setminus \bullet}(f) \right]^2. \quad (\text{A.13})$$

This gives all that we need to compute arbitrary confidence intervals for the Slepian-tapered multitaper spectrum estimate. An example is shown in Figure A.3.

A.4.2 The Harmonic F (Variance Ratio) Test

Harmonic analysis (in the context of spectrum estimation) has come to mean the study of line components in a spectrum, without regard to whether they are at multiples of a common frequency or not. To make sense of this, it is essential to recognize that the assumption of “pure” line components is a convenient fiction, and is rarely supported over long time spans. Thus, we can divide time series into two types: short series, in which our focus is primarily on detection and resolution of line components, and long series, in which our focus is typically on the structure of any line components present.

In addition to the basic multitaper approach, Thomson (1982) presented a new approach to the problem of “mixed” spectra, i.e., where line components are embedded in stationary background noise with a continuous spectrum. The process is typically described as being a stationary random process plus a non-zero mean value function, consisting of some number of sinusoidal terms at various frequencies, plus perhaps a polynomial trend. In terms of the spectral representation, this amounts to having the extended Munk-Hasselmann representation

$$\mathbf{E}\{dZ(f)\} = \sum \mu_m \delta(f - f_m) \quad (\text{A.14})$$

in place of the usual assumption that $\mathbf{E}\{dZ(f)\} = 0$. Under this assumption, the continuous part of the spectrum is the second absolute central moment of $dZ(f)$. As implemented in the `multitaper` package, the harmonic F -test assumes the simplest case of a single line component at frequency f_0 . In this case, the eigencoefficients, as defined in Equation (A.5), have non-zero expected value:

$$\mathbf{E}\{y_k(f)\} = \mu U_k(f - f_0). \quad (\text{A.15})$$

The assumption is made that the continuous component of the spectrum near f_0 is slowly varying (or locally white), resulting in the relationship

$$\text{Cov} \{y_k(f), y_j^*(f)\} \approx S(f) \cdot \delta_{j,k}, \quad (\text{A.16})$$

where $S(f)$ is the continuous spectrum and does not include the line power. One then uses point regression at $f = f_0$, where the relation

$$\mathbf{E} \{y_k(f_0)\} = \mu U_k(0) \quad (\text{A.17})$$

holds, and, remembering that both the $y_k(f)$ s and $\mu(f)$ are complex-valued, μ can be estimated by standard regression techniques (Miller, 1974):

$$\hat{\mu}(f) = \frac{\sum_{k=0}^{K-1} U_k(0) y_k(f)}{\sum_{k=0}^{K-1} U_k^2(0)}. \quad (\text{A.18})$$

Subtracting this result from the eigencoefficients gives an estimate of the continuous background spectrum, and comparing this value with the power in the line component results in an F variance-ratio test (Fisher et al., 1990) with 2 and $2(K - 1)$ degrees of freedom for the significance of the line component. Formally,

$$F(f) = \frac{(K - 1) |\hat{\mu}(f)|^2 \sum_{k=0}^{K-1} U_k(0)^2}{\sum_{k=0}^{K-1} |y_k(f) - \hat{\mu}(f) U_k(0)|^2}. \quad (\text{A.19})$$

This is the ratio of the variance in the band $(f - W, f + W)$ explained by the sinusoid to the residual, unexplained variance in the same band, scaled by the degrees-of-freedom.

Thus, as implemented in the `multitaper` package, the test results in an array of F statistics on the same frequency mesh as the spectrum. Significance levels can

be computed using the `qf()` function. When plotting an `mtm` object, the F -test is plotted in place of the spectrum by passing `Ftest = TRUE`, and significance lines can be added to the plots by passing `siglines = c(p1,p2,p3)` with `p1,p2,p3` user-defined significance levels, typically chosen on the basis of sample size. A typical rule of thumb is to set your minimum significance level at $\sim 1 - 1/N$ (Thomson, 1990b). In general, one must be aware of possible false detects, as the F -statistic is highly sensitive to violations of its underlying assumption of a locally white spectra. In practice, one would plot both the spectrum and the F -test; a statistically significant F -test statistic at a given frequency combined with a characteristic (approximately rectangular) multitaper peak centred at the same frequency gives much more credence to the detection.

We also note here that, for practical examples, the choice of zero-padding amount (when computing the FFT) can have significant impact upon the F -test. As is shown in Thomson (2001, pp. 364–365), the standard deviation of the F -test can often be very small. Thus, the number of zero-padded transform bins should be selected to be approximately equivalent to half this standard deviation in order to accurately determine estimated frequencies. This technique can be applied as part of an iterative process whereby a pilot estimate of the spectrum and F -test are computed, and then a refinement is made based on the maximum F -test value observed (approximately the signal-to-noise ratio).

A.4.3 Harmonic F -test Example

To explore the use of the harmonic F -test as included in the `multitaper` package, we use the Hadley Centre Central England Temperature daily series, available from

the United Kingdom Meteorological Office (2011), as documented in Parker et al. (1992). The series as used in this paper consisted of 87,566 daily observations, beginning January 1, 1772 and ending September 30, 2011. Figure A.2 displays the first six years of the daily temperature series. There are no missing values in this series. It is included with the multitaper package as dataset `CETdaily`.

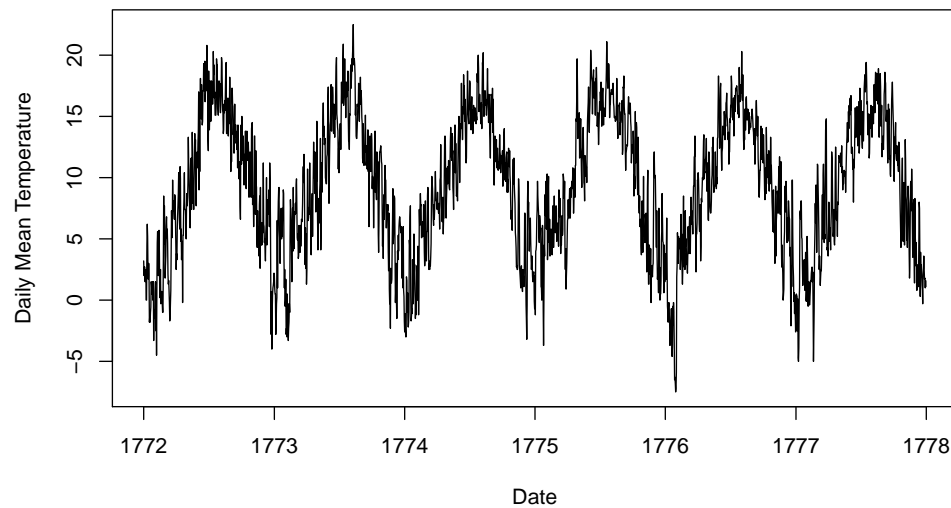


Figure A.2: First six years of the CET daily series.

We compute the multitaper spectrum of this series and display the portion around 1 cycle/year using the `dropFreqs` function, as detailed in Section A.7.1. As is commonly known, any long-run temperature series exhibits extremely strong response at 1 cycle/year, or $31.69n\text{Hz}$. This series is no exception, as can be seen in Figures A.3 and A.4. The jackknifed confidence intervals are included in these plots at 5% and 95%.

Function 2. spec.mtm (Ftest and Jackknife)

```
R> data("CETdaily")
R> cet.spec <- spec.mtm(CETdaily[, "Temp"], nw = 5, k = 10,
+ plot = FALSE,
+ Ftest = TRUE, jackknife = TRUE, dT = 86400, units = "second")
```

Function 3. dropFreqs

```
R> zoomSpec <- dropFreqs(cet.spec, 2e-08, 4e-08)
R> zoomSpec$freq <- zoomSpec$freq * 1e+09
```

Function 4. plot.mtm (Jackknife Confidence Intervals)

```
R> plot(zoomSpec, main = "", xlab = "Frequency in nHz",
+ jackknife = TRUE,
+ ylab = "Spectrum with Jackknife Confidence Intervals")
R> text(x = 36, y = 4e+09, "1 cycle/year", col = "black")
```

The F -test coefficients are contained in the `mtm` object, and can easily be extracted and plotted separately. The frequency array is also available, and can be modified to scale to user-selected units. As is shown in this example, the function `dropFreqs` acts on the entire `mtm` object, focusing (in frequency) the spectrum estimate, the harmonic F -test statistic (if computed), and possibly the coherence (see Section A.5.1). We also show the use of the `siglines` parameter, placing a 0.999 significance line on the plot.

Function 5. plot.mtm (Ftest)

```
R> plot(zoomSpec, Ftest = TRUE, siglines = c(0.999), xaxs = "i",
+      xlab = expression(paste("Frequency in ", mu, "Hz", sep = "")))
```

A.5 Bivariate Time Series: Magnitude-squared Coherence

Given two stationary stochastic processes $x(t)$ and $y(t)$, $t \in \mathbb{Z}$, the coherence between x and y , is the complex-valued function of frequency defined as

$$C_{xy}(f) = \frac{S_{xy}(f)}{\sqrt{S_x(f)}\sqrt{S_y(f)}}, \quad (\text{A.20})$$

where $S_x(f)$ and $S_y(f)$ are spectra of x and y respectively, and $S_{xy}(f) = \mathbf{E}[dX(f)dY^*(f)]$.

A related quantity is the MSC between x and y , denoted as $\gamma_{xy}(f)$ and defined by

$$\gamma_{xy}(f) = |C_{xy}(f)|^2. \quad (\text{A.21})$$

Now let x_t and y_t , $t \in 1, \dots, N$ be realizations of $x(t)$ and $y(t)$ respectively. Given these realizations, the multitaper estimate of the autospectrum of x , as per Section A.3, is

$$\hat{S}_x(f) = \frac{1}{K} \sum_{k=0}^{K-1} |x_k(f)|^2 \quad (\text{A.22})$$

where $x_k(f)$ is the k^{th} eigencoefficient of x , and the non-adaptively-weighted spectrum estimate is used. Analogously, the cross-spectrum can be estimated by

$$\hat{S}_{xy}(f) = \frac{1}{K} \sum_{k=0}^{K-1} x_k(f) \overline{y_k(f)}, \quad (\text{A.23})$$

where $x_k(f)$ and $y_k(f)$ are the eigencoefficients of x and y respectively, and the line indicates complex conjugation. Substituting these estimates of the auto- and cross-spectra into the definition of the coherence results in a multitaper estimate of coherence:

$$\hat{C}_{xy}(f) = \frac{\sum_{k=0}^{K-1} x_k(f) \overline{y_k(f)}}{\left(\sum_{k=0}^{K-1} |x_k(f)|^2 \cdot \sum_{k=0}^{K-1} |y_k(f)|^2 \right)^{0.5}} \quad (\text{A.24})$$

with the MSC then computed as

$$\widehat{\gamma}_{xy}(f) = \left| \widehat{C}_{xy}(f) \right|^2. \quad (\text{A.25})$$

The coherence between two time series can be computed by using the `mtm.coh` function, as demonstrated in Section A.5.1. The coherence can be computed between weighted or unweighted spectrum estimates, depending on how the user has generated the `mtm` objects.

A.5.1 Spectral Coherence Example

In similar fashion to Kuo et al. (1990), we examine records of atmospheric CO₂ from the Mauna Loa observatory, Hawaii, USA, and monthly northern hemisphere temperature anomalies from the Hadley Climate Research Unit, University of East Anglia, UK. The records were obtained from NOAA (2011); Hadley Climate Research Unit (2011) and were cleaned. The few missing points were linearly interpolated. Both records are of monthly data. As in the cited paper, we estimate the trend using a multitaper technique, with further details given in Section A.7.1. We use this method of trend estimation over the more traditional least-squares estimator for the favourable

frequency-domain aspects of the result. The residuals after this trend estimate are then passed through the `spec.mtm` function, and the resultant `mtm` objects are passed to `mtm.coh`.

Function 6. multitaperTrend

```
R> data("mlco2"); data("HadCRUTnh");
R> temp <- HadCRUTnh
R> nw <- 5; k <- 10; dt <- 1/12; N <- length(mlco2[, 1]);
R> time <- seq((mlco2[1, 1] + mlco2[1, 2]/12), (mlco2[N, 1] + mlco2[N,
+   2]/12), dt)
R> ttbar <- time - (time[N] + time[1])/2
R> trend1 <- multitaperTrend(mlco2[, "CO2"], B = 0.18, dT = dt,
+   t.in = time)
R> co2.resid <- mlco2[, "CO2"] - trend1[[1]] - trend1[[2]] * ttbar
R> co2.resid <- ts(co2.resid, deltat = 1/12)
R> trend2 <- multitaperTrend(temp[, "Temp"], B = 0.12, dT = dt,
+   t.in = time)
R> temp.resid <- temp[, "Temp"] - trend2[[1]] - trend2[[2]] * ttbar
R> temp.resid <- ts(temp.resid, deltat = 1/12)
R> plot(time, mlco2[, "CO2"], type="l", xlab="Date",
+   ylab = "CO2 in ppb")
R> lines(time, mlco2[, "CO2"]-co2.resid,type="l",col="red")
R> plot(time,temp[, "Temp"], type="l", xlab="Date",
+   ylab = "Temp. in Celsius")
R> lines(time, temp[, "Temp"] - temp.resid, type="l", col="red")
R> co2.mtm <- spec.mtm(co2.resid, nw = 5, k = 10, plot = FALSE,
+   returnInternals = TRUE, dtUnits = "year")
R> temp.mtm <- spec.mtm(temp.resid, nw = 5, k = 10, plot = FALSE,
+   returnInternals = TRUE, dtUnits = "year")
```

Function 7. `mtm.coh`

```
R> coh <- mtm.coh(co2.mtm, temp.mtm, plot = FALSE)
```

Function 8. `plot.mtm.coh`

```
R> plot(dropFreqs(coh, 0, 2.5))
```

The coherence plot shown here differs from Kuo et al. (1990) in that the first few yearly harmonics have not been removed from the individual series before computing the MSC. The two plots are similar in that their scales have been adjusted to be the same, and both have been detrended using `multitaperTrend`. For more details on the implications of this plot, see Kuo et al. (1990, pp. 711–713).

A.6 Complex Demodulation

Complex demodulation is a tool to analyze both the phase and the amplitude of a specific frequency component in a time series. A good reference on the general application of this theory is given in Bloomfield (2000, pp. 97–131).

In general, one is interested in locating periodic phenomena that have a simple representation in terms of cosine functions. However, even after the multitaper technique has been used to improve analysis, harmonic analysis has some limitations in describing the signal component of the time series. To overcome this, the technique of complex demodulation can be used to describe features of the data that could be missed with standard multitaper harmonic analysis, or to confirm that no such features exist.

The algorithm for complex demodulation involves two steps. First, a frequency shift is applied to the data such that the frequency of interest is centred at zero, and secondly, the centred frequency of interest is isolated using a low-pass filter. The objective is to expose small changes in amplitude or phase of a specific approximately periodic cycle.

Begin with an assumed model for x_t :

$$x_t = R_t \cos 2\pi(f_0 t + \phi_t) \quad (\text{A.26})$$

where $\{R(t)\}$ is the amplitude, and $\{\phi_t\}$ represents the slowly varying phase of a harmonic component at frequency f_0 . We will focus on isolating and graphing the slowly varying phase, ϕ_t . To develop the method, consider the complex analog of (A.26):

$$x_t = R_t e^{2\pi i(f_0 t + \phi_t)}. \quad (\text{A.27})$$

If, as we assume, f_0 is known, then we can construct

$$y_t = x_t e^{-2\pi i f_0 t} = R_t e^{2\pi i \phi_t}. \quad (\text{A.28})$$

In this case, $R_t = |y_t|$ and $e^{2\pi i \phi_t} = \frac{y_t}{|y_t|}$. The new series $\{y_t\}$ is said to be obtained from $\{x_t\}$ by complex demodulation. Returning to Equation (A.26), the real form of x_t can be written as

$$x_t = \frac{1}{2} R_t (e^{2\pi i(f_0 t + \phi_t)} + e^{-2\pi i(f_0 t + \phi_t)}) \quad (\text{A.29})$$

and is thus the sum of two complex terms, one similar to Equation (A.27) and the second its complex conjugate. We will use complex demodulation and filter the second (conjugate) component using convolution with a Slepian sequence.

Applying complex demodulation to the real form, Equation (A.29), we obtain

$$y_t = \frac{1}{2}R_t e^{2\pi i \phi_t} + \frac{1}{2}R_t e^{-2\pi i(2f_0 + \phi_t)}. \quad (\text{A.30})$$

The first term is our desired component, from which we can easily extract R_t and ϕ_t , while the second term must be removed, which we will do using an appropriate low-pass filter. In this case we opt to convolve the data with a Slepian sequence with relatively small bandwidth parameter, w , which acts as an effective low-pass filter. Using this method, the estimate of y_t is formed by

$$y_t = \sum_{j=0}^{N_c-1} v_j^{(0)}(N_c W) x_{t-j} e^{-i2\pi f_0(t-j)\Delta t}, \quad (\text{A.31})$$

where $v_j^{(0)}(N_c W)$ is a Slepian taper with appropriately chosen time-bandwidth parameter, N_c is the length of the convolution representing the time and frequency resolution trade-off and Δt is the time step. In previous equations where Δt was omitted, it was assumed to be 1.

After passing the complex demodulate through a low-pass filter, isolating the single component of interest, the result is smoothed to remove unwanted variation due to noise. The choice of smoother is an open one, so for consistency we use a short-length convolution Slepian filter. The parameter N_c for this filter should be considerably less than the length of the series of interest and is typically chosen to be approximately $1/f_0$ —i.e., the time-domain period of the frequency of interest.

A.6.1 Complex Demodulation and the CET Series

In this section, we examine the CET monthly means series originally compiled by Manley (1974), and updated in Parker et al. (1992); Parker and Horton (2005). The analysis we follow was originally published in Thomson (1995), and examined the phase

of the annual cycle in the monthly temperature series. The data consists of monthly mean temperature for CET from 1659 to 2011.⁴ We take t to represent the calendar year, with $t = 1$ representing January 1659, and take the time step, $\Delta t = 1/12$. We also use 12-year blocks for time resolution. We note that in the original analysis, the author found no visible difference in the phase plot when correcting for month length in his analysis.

⁴The CET monthly series begins in 1659 whereas the CET daily series begins in 1772.

Function 9. demod.dpss

```

R> data("CETmonthly")
R> nJulOff <- 1175
R> xd <- ts(CETmonthly[, "temp"], deltat=1/12)
R> demodYr <- demod.dpss(xd, centreFreq=1, NW=3, blockLen=120,
+   stepSize=1)
R> phase <- demodYr["phase"] [["phase"]]
R> offsJul <- 3*360/365
R> phaseAdj <- phase
R> phaseAdj[1:nJulOff] <- phase[1:nJulOff] + offsJul
R> yr <- (time(xd)+1658)[1:length(phase)]
R> plot(yr, phaseAdj, type="l", lwd=2,
+   ylab="Phase of the Year in Degrees",
+   xlab="Gregorian calendar date")
R> lines(yr[1:nJulOff], phase[1:nJulOff], col="red", lty=3)
R> fit <- lm( phaseAdj ~ yr)
R> abline(fit, lty=2, col="blue")

```

We include Δt in equation (A.31) and construct y_t and examine the phase by following the algorithm detailed in Section A.6 using $NW = 3$ as the time-bandwidth parameter. Figure A.8 shows the phase plot with the 2.96 degree correction accounting for the three-day calendar offset in September of 1752. The dotted red line shows the phase plot up to September of 1752 without the correction, and the dashed blue line shows a fitted regression line to the corrected phase line. This regression line has

a slope of 56.8 arcseconds per year, which is similar to the 51.1 arcseconds per year found in Thomson (1995) and slightly greater than the precession constant of 50.3 arc seconds per year. Note that Thomson's original paper used data only up to 1990, and this analysis uses data up to 2011. As noted in the paper, the phase begins to exhibit different characteristics after 1940.

A.7 Additional Tools and Extending Functionality

A.7.1 Miscellaneous Functions

There are four miscellaneous utility routines included in the `multitaper` package. Some are referenced within other routines (including `centre` and `dpss`), while others are not needed for the default operations of the package. The four routines are:

1. `dpss`: Generates Slepian tapers (discrete prolate spheroidal sequences) using the tridiagonal method of Slepian (1978) and returns the eigenvectors and eigenvalues for the user-provided `nw`, `k` and `N` parameters.
2. `centre`: Takes a time series, and estimates the mean, using one of: the arithmetic mean, the robust trimmed mean, or the Slepian taper-based mean method; see Thomson (1982). The function returns the residuals after the computed mean has been subtracted.
3. `dropFreqs`: Given an `mtm` or `mtm.coh` object, truncates all internal data objects to a frequency range specified by the user. Note that `mtm.coh` cannot act on objects that have first been passed through `dropFreqs`, instead requiring unmodified `spec` objects.

4. `multitaperTrend`: Given a time series, estimates a first-order polynomial trend using the Slepian tapers. This method has improved frequency-domain performance over least-squares estimation.

A.7.2 Extending Functionality

As detailed in Section A.3, the majority of the tools developed for multitaper spectral analysis work on the raw (or weighted) eigencoefficients $y_k(f)$. To provide functionality for extending this package, an option is provided in the `spec.mtm` call that returns the internal parameters of the spectrum estimation procedure. The option that produces these parameters is `returnInternals = TRUE`, and the parameters, for an `mtm` object named `test.mtm`, are:

- `test.mtm[["mtm"]][["eigenCoefs"]]`
- `test.mtm[["mtm"]][["eigenCoefWt"]]`

and consist of the eigencoefficients and their associated weights. Using these coefficients, and the related quantities (returned by default) of `nw`, `k`, `nFFT` and `dpss` (the tapers), it is possible to extend the package to produce any desired multitaper-based tool. As there have been numerous papers published since 1982 that contain suggestions or development of tools, the list of possible extensions is too long to fully list, but we do suggest several possibly useful options.

1. Multiple-line harmonic F -test statistic (see e.g., Thomson, 1990a),
2. High-resolution spectrum estimates (see e.g., Thomson, 1982, 1990a),
3. Bispectra and polyspectra (see e.g., Birkelund and Hanssen, 1999),

4. Canonical coherence and canonical bicoherence (see e.g., He and Thomson, 2009a,b).

Each of these tools is useful in certain applications.

A.8 Summary

The `multitaper` package implements the core functionality detailed (and implied) by Thomson (1982), with refinements from Riedel and Sidorenko (1995) and Percival and Walden (1993) among many others. Discrete prolate spheroidal sequences are generated in an efficient and accurate fashion and are provided as the default tapers for the multitaper spectrum estimation routine. Approximately unbiased adaptive sine tapers are also available, and the spectrum associated with them is easily produced. A number of core extensions, including the harmonic F -test, the magnitude-squared coherence, and jackknife estimates of significance, are also provided. Finally, high-level utility routines designed to make working with data easier have been implemented and are included in the package.

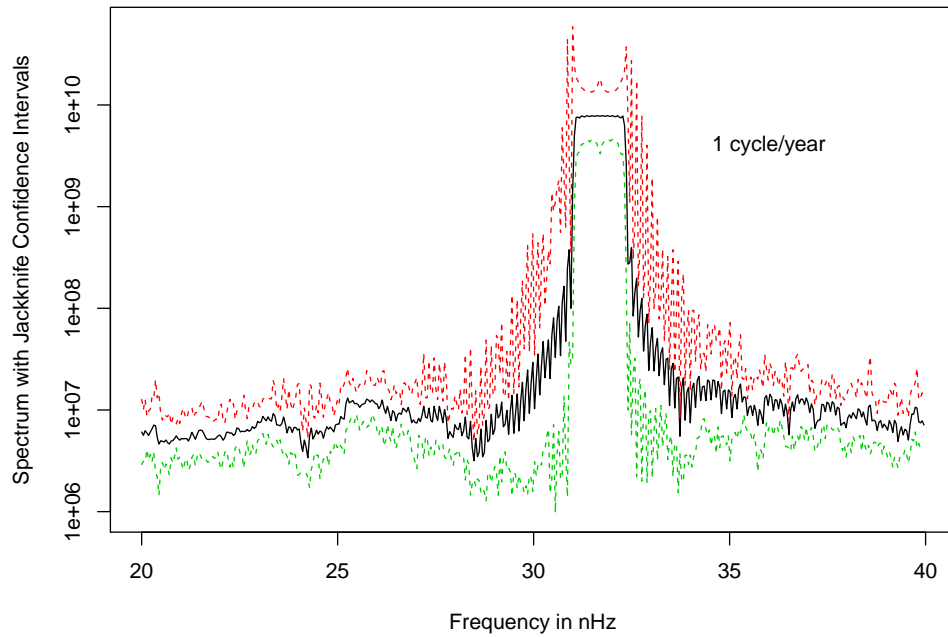


Figure A.3: Spectrum of CET series, zoomed to region around 1 cycle/year ($31.69\text{nHz} = 31.69 \times 10^{-9} \text{ Hz}$), with 95% jackknifed confidence intervals.

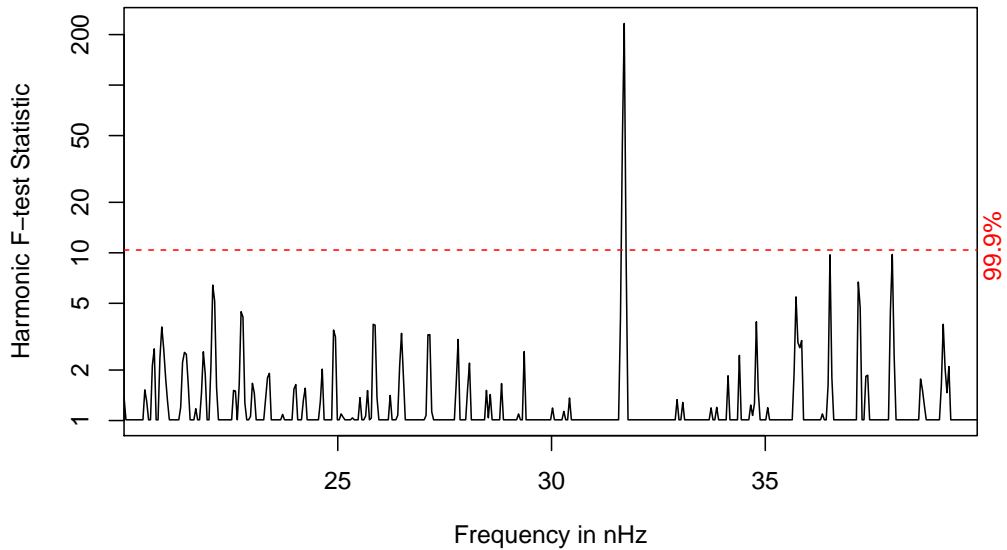


Figure A.4: Harmonic F -test statistic for the CET series, zoomed to low frequencies using the function `dropFreqs`.

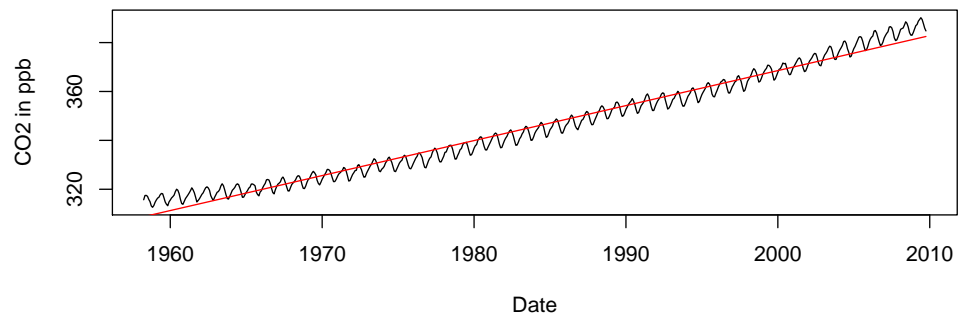


Figure A.5: CO₂ concentration time series in parts-per-billion with trend lines fitted.

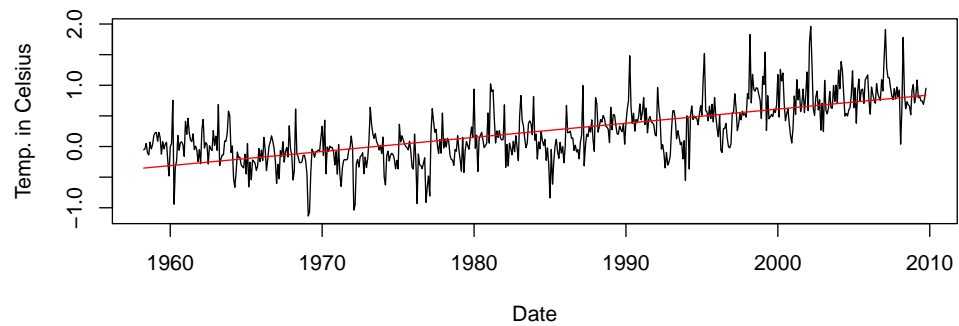


Figure A.6: Temperature deviations time series in degrees Celsius with trend lines fitted.

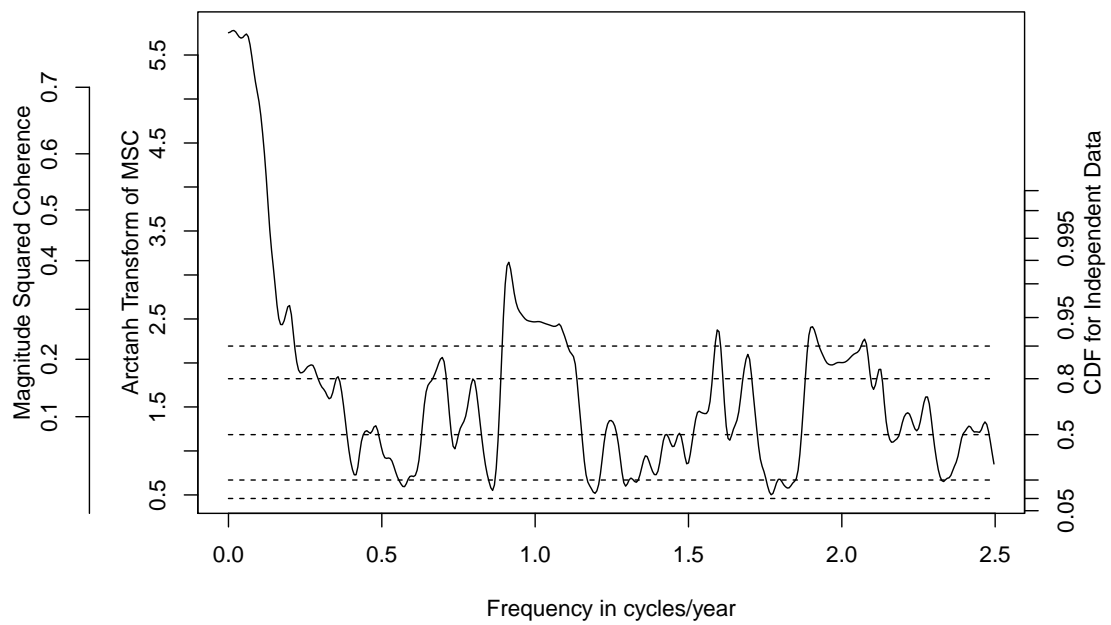


Figure A.7: MSC between monthly CO_2 measurements from Mauna Loa, and the global temperature series during 1958–2007. The Arctanh transform normalizes the MSC and each integer value on this scale represents approximately one standard deviation (Thomson and Chave, 1991b).

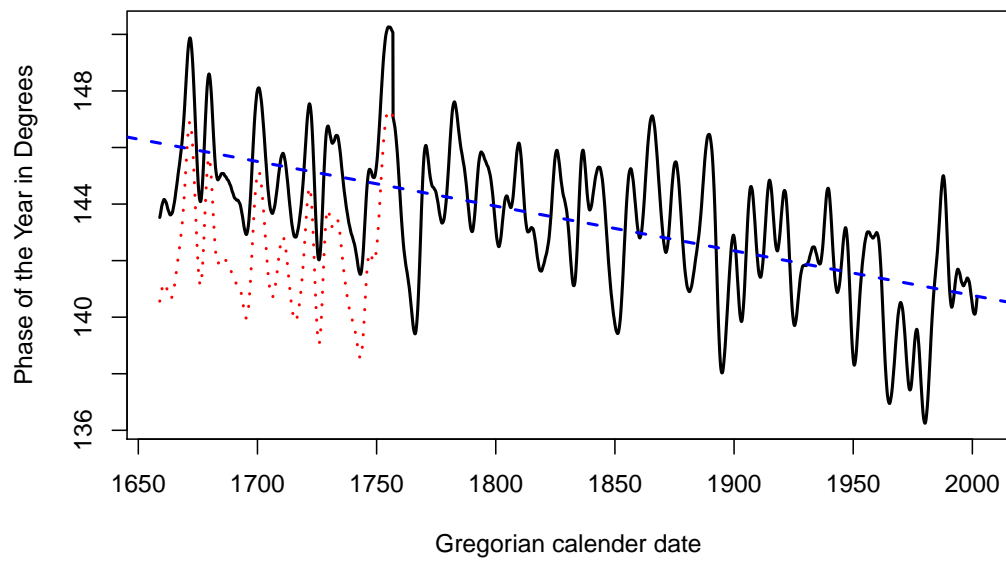


Figure A.8: CET monthly phase, thick (black) line. The dotted red line indicates the phase before the calendar correction, and the dashedblue line shows the least squares line with a slope of 56.8 arcseconds.