Spectral Analysis:
in theory and by example

with an introduction to the Morlet wavelet

P M Tate and C J Holden

Title
Spectral analysis: in theory and by example – with an introduction to the Morlet wavelet.

Authors
P.M. Tate
Director WQ Data Pty Ltd
Terrigal, New South Wales, Australia

C.J. Holden
Director Oceanographic Field Services Pty Ltd
Freshwater, New South Wales, Australia

ISBN 978-0-6455057-0-2: pdf version
ISBN 978-0-6455057-1-9: hard copy version

Key words: Fourier transform, auto spectrum, cross spectrum, rotary spectra, coherence squared, phase lag, phase difference, harmonic analysis, Fast Fourier Transform, short time Fourier transform, Morlet wavelet transform, principal component analysis, filtering, confidence limits, stationarity, tapering, aliasing, limiting frequencies, patching missing data.

This work is © Peter M. Tate and Clive J. Holden, 2022. You may use and reproduce the contents of this work for non-profit research and educational purposes only. No part of this work may be used, reproduced, translated or copied in whole or in part for commercial or for-profit purposes without the expressed written consent of the copyright owners.

Please reference this work as follows:
Tate, P.M. and Holden, C.J. (2022). Spectral analysis: in theory and by example – with an introduction to the Morlet wavelet. WQ Data Pty Ltd. ISBN 978-0-6455057-1-9, 494p.

Preface

Spectral analysis is a data analysis tool that lends itself well to examining signals that contain regular, repeating patterns. Examples include tides and tidal currents, water temperature and long term processes such as the El Nino / La Nina cycle. It is a commonly used tool in physical oceanography, but also has applications in the medical, behavioural, cognitive and other sciences. The authors' background is in physical oceanography and that is where the emphasis of this book lies.

Our target audiences are those starting out using spectral analysis as a data analysis tool. So, we're generally aiming at senior undergraduate or post graduate students; those commencing their spectral journey.

Spectral analysis should not be seen as a panacea for all data processing – it is a useful tool. Like any tool it can be abused, so we spend a fair bit of time discussing the work that needs to be done, questions that need to be asked before implementing spectral analysis. We say a few things to which the purists may object, particularly concerning stationarity. While accusations of 'heretic' are likely to be aimed at us, we generally discard the assumption of stationarity. Why we do this, read Section 3.8. The final (technical) chapter focuses on non-stationary data sets. In particular, we examine the Morlet wavelet, albeit not in as much detail as the preceding chapters on spectral analysis.

The focus here is on time series analysis, although there are some examples that use data from a spatial array of instruments. The only real difference between time and space data is that a spectral analysis of the former converts time to frequency, while that of the latter converts space to wave number.

Fourier theory underpins spectral analysis and it is from there that this work develops. So, yes, there is a fair bit of mathematics embedded here. Our intention is to provide the reader with the mathematical details of the theory behind each of the different tools. We've tried to not omit any steps in the theoretical development, and some will find this theoretical development a bit tedious. We've worked on the assertion that it is better to include all of the steps rather than omit a critical piece of mathematics i.e. we've tried to avoid any jumps in mathematical logic.

A pragmatic view is adopted by complementing the theory with multiple examples for each of the tools described. The examples cover a range of variables, time spans and delta-t values. Most examples use the 'radix-2' Fast Fourier Transform algorithm which requires there to be 'close to' 2^m number of data points. This is the most commonly used algorithm for solving the discreet Fourier transform.

We've tried to keep the structure of the examples uniform, hence the structure is repetitive and, well….., a bit boring. Broadly, this structure involves asking a question, pre-processing of the data and undertaking statistical tests to ensure that we are not violating any of the assumptions associated with spectral analysis. Considerable emphasis is placed on getting the data in the right form to validly undertake the spectral analysis.

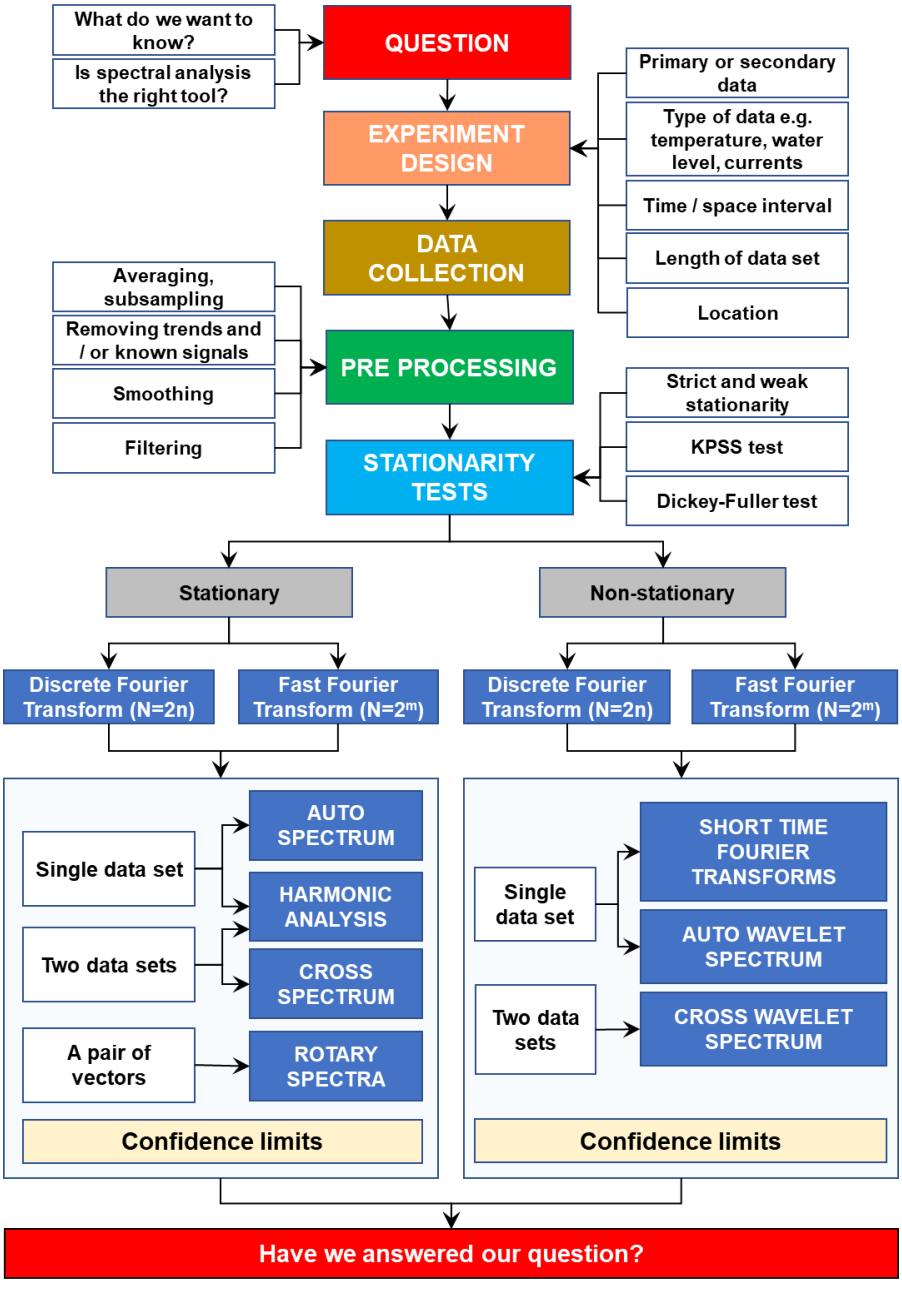
Most of the examples are from the Southern Hemisphere, where winter lies in the middle of the year and Coriolis deflects moving waters to the left.

To keep costs down, we've purposefully not engaged a publishing company and provide this book in an electronic format. It contains lots of equations and figures, which often equates to expense in a hard copy version. We want students to have a 'no cost' resource that helps build their understanding of the topic. However, one downside in not going through a publishing company is that there has been limited editorial review – so there may be (will likely be) some errors, our apologies.

Having said that, it is understood that what we've produced here may not suit everyone's taste. It's unlikely that this work and/or its layout will resonate with everyone. That's okay. If it helps just one person to better understand spectral analysis, then that's great.

All data are analysed 'as is'. Although we have quality assured most of the data used here, no guarantee of data quality is provided. The intention here is to demonstrate statistical techniques, not necessarily to draw conclusions or to provide detailed interpretations of the results.

The flowchart below outlines the steps that are described in this book to carry out spectral based analyses. It starts with a question, proceeds through a range of tasks to prepare the data, then ends with the appropriate spectral analysis tool to answer the question.


What do we want to know?
Is spectral analysis the right tool?
QUESTION
Primary or secondary data
Type of data e.g. temperature, water level, currents
Time / space interval
Length of data set
Location
EXPERIMENT DESIGN
DATA COLLECTION
Averaging, subsampling
Removing trends and / or known signals
Smoothing
Filtering
PRE PROCESSING
Strict and weak stationarity
KPSS test
Dickey-Fuller test
STATIONARITY TESTS
Stationary
Non-stationary
Discrete Fourier Transform (N=2n)
Fast Fourier Transform (N=2m)
Discrete Fourier Transform (N=2n)
Fast Fourier Transform (N=2m)
Single data set
AUTO SPECTRUM
HARMONIC ANALYSIS
Two data sets
CROSS SPECTRUM
A pair of vectors
ROTARY SPECTRA
Confidence limits
Single data set
SHORT TIME FOURIER TRANSFORMS
AUTO WAVELET SPECTRUM
Two data sets
CROSS WAVELET SPECTRUM
Confidence limits
Have we answered our question?

Acknowledgements

Most of the offshore oceanographic data was collected on behalf of Sydney Water Corporation. Their permission to use their data here is gratefully acknowledged.

Tide (water level), atmospheric pressure, Southern Oscillation Index and Tasman Sea surface temperature data were obtained from the Australian Bureau of Metrology. They are gratefully acknowledged for access to these data.

Fortran programs were written by the authors. The cross spectrum and rotary spectrum programs were modified by the authors from programs written by Dr. L.Padman and Prof. J.H.Middleton, respectively.

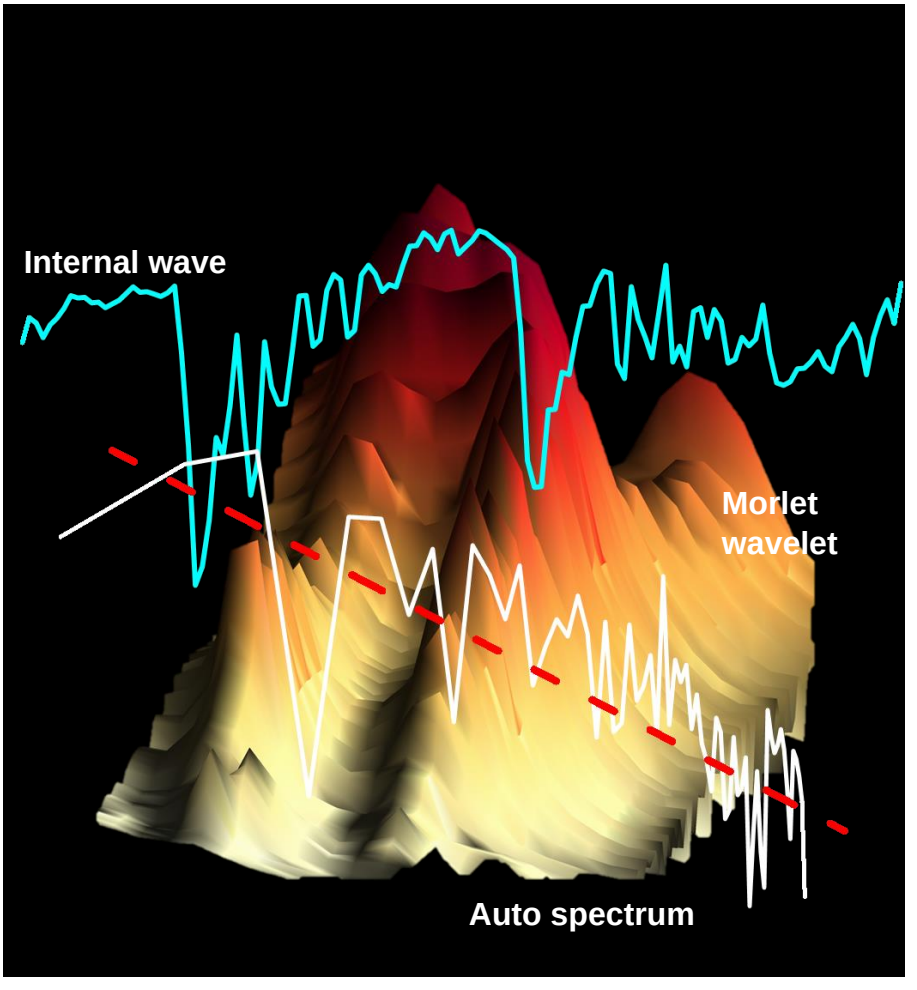
Wavelet analyses and the associated plots are based on the Matlab scripts by Torrence and Compo (1998) and Grinsted et al (2004). Modifications to these Matlab scripts were made by the authors to suit the work presented in this book.

Contents

Chapter 1. Introduction 1
1.1. Overview 1
1.2. The approach 11
Chapter 2. The Fourier Transform 12
2.1. Origins of Fourier analysis 12
2.2. Fourier theory 14
2.3. Discrete Fourier transform 17
2.4. The Fast Fourier Transform 21
2.5. What to expect 22
2.6. Summary 38
Chapter 3. Preparing the Data 39
3.1. Experimental design 39
3.2. Filling data gaps 41
3.3. Filtering and subsampling 44
3.4. Filter response functions 57
3.5. An even number of data points 59
3.6. Truncate or extrapolate 59
3.7. Aliasing and limiting frequencies 63
3.8. Stationary data sets 67
3.9. Independence of the data 79
3.10. Summary 80
Chapter 4. The Auto Spectrum 82
4.1. Removing trends, known signals, outliers 83
4.2. Confidence limits 83
4.3. Examples of the auto spectrum 91

4.4. Spectral analysis in the space domain 129
4.5. Summary 137
Chapter 5. The Cross Spectrum 140
5.1. Theory behind the cross spectrum 141
5.2. Examples of the cross spectrum 148
5.3. Summary 209
Chapter 6. Rotary Spectra 210
6.1. Theory behind rotary spectra 213
6.2. Examples of the rotary spectrum – a single vactor 236
6.3. Examples of the rotary spectrum – a pair of vectors 264
6.4. Summary of rotary spectra 285
6.5. Overview of spectral analysis 288
Chapter 7. Harmonic Analysis 302
7.1. Harmonic analysis using a scalar 306
7.2. Harmonic analysis with a vector 325
7.3. Harmonic analysis in space 338
7.4. Summary 351
Chapter 8. Short Time Fourier Analysis 352
8.1. Limitations of spectral analysis 354
8.2. Short time Fourier transform 371
8.3. Examples using the STFT 386
Chapter 9. The Morlet wavelet 398
9.1. Theory of the Morlet wavelet 398
9.2. Examples of the Morlet wavelet 410
9.3. The Morlet wavelet cross spectrum 423
9.4. Examples of the Morlet cross wavelet 425

9.5. Summary 438
Chapter 10. Tips and Tricks 440
10.1. General 440
10.2. Stationarity 441
10.3. Resolution versus practical interpretation 442
10.4. Excel FFT routine 443
10.5. Spectral frequencies 443
10.6. Sum of spectral estimates 444
10.7. Confidence limits 445
10.8. A single vector 446
10.9. A pair of vectors 446
Chapter 11. References 447
Chapter 12. Appendix A – The Fast Fourier Transform 455
12.1. Discrete Fourier transform 455
12.2. Fast Fourier transform 458
12.3. Examples 466
12.4. The sparse Fourier transform 473
Chapter 13. Appendix B – Principal Component Analysis 475
13.1. Axis Orientation 475
13.2. Dimension Reduction 476
13.3. Example 1: Axis rotation 483
13.4. Examples 486


Internal wave
Morlet
wavelet
Auto spectrum

Chapter 1. Introduction

1.1. Overview

The primary intention of this book is to provide a resource for those initiating a journey into spectral analysis. Specifically, we examine several different forms of spectral analysis and introduce the (Morlet) wavelet analysis. From its origins in Fourier theory, we show how the basic equations are derived and complement this with many examples for each of the areas of work. In this context, use of the term 'applied spectral analysis' is generic and refers to any or all of spectral analysis, harmonic analysis and wavelet analysis. Application is primarily to time series analysis, although exactly the same principles apply to the space domain. Some examples use spatial data sets, mainly to demonstrate that they too can be applied to spectral analysis.

Throughout this book we use a combination of Excel, Fortran and Matlab tools to undertake the applied spectral analysis. Excel is used because most personal computers have this function installed ('Fourier analysis' is an inbuilt routine in the Excel 'data analysis' tool). However, Fourier analysis within Excel does have limitations, for example:

- No cross spectral or rotary spectral analysis functions (although these can be built using the Fourier analysis tool).
- No in-built harmonic analysis or wavelet analysis tools.
- A limit of 4,096 on the maximum number of elements in its 'Fast Fourier Transform' routine. It appears that this can be expanded using VBA tools, but that is beyond the scope of this work.
- No in-built discrete Fourier transform package (although such packages appear to be available from web sites).

Not everyone has access to a Fortran compiler or to the Matlab package, hence the Excel package provides a good starting point for applied Fourier analysis. The other packages have been included here because

they can do all of the processing that Excel can do as well as the more advanced Fourier analyses outlined in the dot points above.

The use of Fortran as a programming tool has, perhaps, declined over the years. However, having multiple tools to do the same task means that we can confirm our results, which gives us more confidence in what we're doing.

Spectral analysis is not a panacea for all time series analyses. It is just one of many tools. It is an excellent analytical tool when there is a regular, repetitive number of cycles in the data set e.g. annual temperature signal, tide signals in water level. Hence it is a tool that is favoured by physical oceanographers. It is, perhaps, not so good for analysing data that contain irregular features or outliers.

Where spectral (or Fourier) analysis sits in the broader range of data analysis tools is shown in Figure 1–1. The primary objective of applied spectral analysis is to determine which variables (or processes) are important by reducing the dimension of the problem. By this, we mean that applied spectral analysis can be used to identify (and then remove) known process from the data set until we are left with a data set comprising only random variables.

The basic idea behind spectral analysis is to represent the data set (which can be either time based or space based) by a sum of sine and cosine curves. Each of the sine and cosine curves have different amplitudes and phases: the relative amplitudes telling us how important each sine / cosine curve is in terms of their contribution to the total energy (or variability) in the data set and the phase telling us the 'starting point' of each sine / cosine curve.

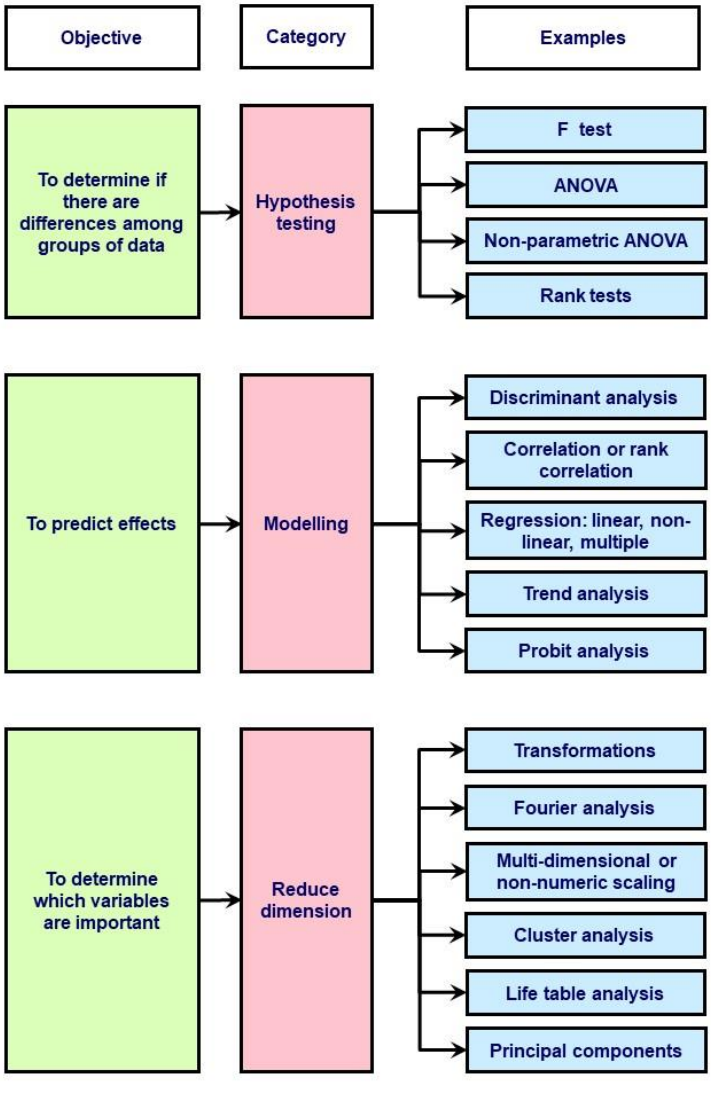


Figure 1–1. Schematic showing some of the different approaches and examples of tools for data analysis. Specifically, it identifies where applied Fourier analysis sits in a broader matrix of data analysis options.

We then try to allocate a process (using the frequency information for each sine / cosine curve) to the dominant signals (i.e. those with the greatest relative amplitude). In this way, we can quantify the relative importance of different processes in the data set. In a practical sense, it may not be possible to allocate a process to a particular signal. This would warrant further investigation into why this signal is present in the data set.

Applying spectral analysis (i.e. implementing the Fourier transform) to a data set changes the way in which we visualise the data. If we're dealing with a time series, the time is transformed to frequency via the Fourier transform. So rather than telling us when a feature occurs, the Fourier transformed data set tells us how frequently a particular feature occurs. If we're dealing with a spatial data set, the Fourier transform changes the data set to wave number. (Most of our focus will be on the former data sets – time series).

We'll restrict ourselves to processes in the physical environment. Generally, we (or at least most of our experiences) deal with time series, rather than spatial data sets. Therefore, much of the development of spectral analysis presented here will focus on time series of physical data sets (e.g. temperature, water level, ocean currents). We'll often use temperature and water level data because we all have a good 'feeling' for the patterns expected in both types of data. These processes repeat themselves regularly and applied spectral analysis is a good tool to analyse such repetitive processes.

We assume that the data set comprises (i) a trend (usually linear, but not necessarily), (ii) a set of (initially) unknown signals (or 'known, unknowns') that can be linked to physical processes and (iii) residual variations that are, in general, random and normally distributed. Ultimately, what we want to do is to remove any trends, identify and remove the initially unknown signals (e.g. annual signals from temperature records, tidal signals from water level records) then undertake our spectral analysis on the remaining data which we hope will follow a normal distribution or identify a previously unknown signal. Often, this requires multiple runs of the spectral analyses (an iterative process) to identify and sequentially remove signals that represent physical processes. However,

in practice, we usually don't arrive at a normal distribution of the residuals even though we've removed all of the known processes from the data set. Having said this, sometimes it is the identification of periodic signals and their relative importance in the data that are of most importance, rather than their removal.

Much of the work presented here is based on three different data sets (temperature, water level and ocean currents) and four different time scales (5-minute, hourly, daily and monthly) to demonstrate applied spectral analysis. We've chosen them because:

- Generally, we have a good feeling for these types of data, even though we may not necessarily understand the underlying processes associated with each signal in the data set.
- The temperature data sets generally have one dominant signal (the annual signal), which makes it easy to demonstrate the main features of spectral analysis. The water level and current data sets have multiple signals and their interactions in frequency (or wavenumber) space can be examined.
- Our area of specialisation is in physical oceanography, so we have considerable experience in using these types of data. Having said this, the approach shown here can be exported to other areas of study.

The data sets are taken from the locations in the Southern Hemisphere (mostly from the nearshore continental shelf of Sydney, Australia) so, water temperatures will be cooler in the middle of the year and Coriolis will deflect currents to the left. The basic data sets are:

- Water temperature. Usually, this is a single signal, the strong annual signal resulting from the elliptical path taken by the earth about the sun combined with the orientation of the axis of

rotation of the earth. We'll also use temperature as a surrogate for density in the analysis of internal waves.

- Water level. This has multiple strong signals associated with the tides and atmospheric pressure. They result from the gravitational attraction of the earth, moon, sun, and the planets and their relative positions / movements.
- Ocean currents. Understanding of ocean currents is, perhaps, not quite as thorough as that of water temperature and water level. We have included this to demonstrate how spectral analysis can be used on vectored time series.

The nearshore continental shelf of Sydney contains a number of oceanographic processes that operate at different time and space scales. These processes include:

- East Australian Current (EAC) – a poleward flowing western boundary current (e.g. Hamon, 1965; Boland and Hamon, 1970, Boland and Church, 1981; Tilburg et al, 2001; Roughan and Middleton, 2004; Schaeffer et al, 2017). Time and space scales associated with the EAC are of the order months and 100s of kilometres (Hamon and Kerr, 1968; Tate et al, 1989). It is noted that the data use here may only be exposed to the western boundary of the EAC, and that such exposure may be highly variable.
- Coastal trapped waves (CTW) flow towards the equator along the Australian East Coast (e.g. Church et al. 1986a, 1986b; Freeland et al. 1986; Church and Freeland, 1987; Morrow et al. 1990). Time and space scales associated with CTWs are of the order 7-20 days and 100's of kilometres (Church et al. 1986a).
- Wind driven circulation (e.g. Griffin and Middleton, 1990; MacDonald et al, 2009; Wood et al, 2012). Time scales associated with wind driven circulation are of the order hours and kilometres.
- Tides (e.g. Gibbs and Middleton, 1997). Time and space scales are of the order hours to days and 1000s of kilometres.

- Upwelling and downwelling (e.g. McClean-Padman and Padman, 1991; Roughan and Middleton, 2002). These processes tend to be event based and allocating time and space scales is difficult.
- Surface waves. (e.g. Short and Trenaman, 1992; Turner et. al., 2016). Time and length scales are of the order 10s of seconds and metres.
- Internal waves are often present in this region at a range of time scales. High frequency internal waves (periods less than about 30 minutes) often occur (e.g. Cresswell and Boland, 1981; Fang et al. 1984; Tate et al, 2019). In the warmer months their signals can be clearly identified in the temperature structure of the water column. Time and space scales are the order of 10's of minutes and 100's of metres. Internal tide packets with time scales of 3-4 days were discussed by Gibbs and Middleton (1991).

Primarily, four time frames are used: 5 minutes, 60 minutes, 24 hours and one month. Occasionally, we'll use different time frames to demonstrate a particular point. In general, we'll process the data using the radix-2 Fast Fourier Transform (FFT) with 4,096 data points. (This is the maximum sized FFT that Excel can handle). In so doing, the total length of the first three time series is: 14.2 days, 170 days and 11.2 years, respectively. For monthly data, we restrict the size of the FFT to 1,024 records (about 85 years), as there are few reliable data sets that exceed 100 years in duration.

The following figures provide some examples of the data we'll be examining. They show a range of variables and time scales and provide the reader with a broad view of the types of data that are available. There is no explanation of the data sets at this stage, except to note that there are different scales of temporal variability in the plots, some plots contain an obvious linear trend, while others show a clear cyclic pattern.

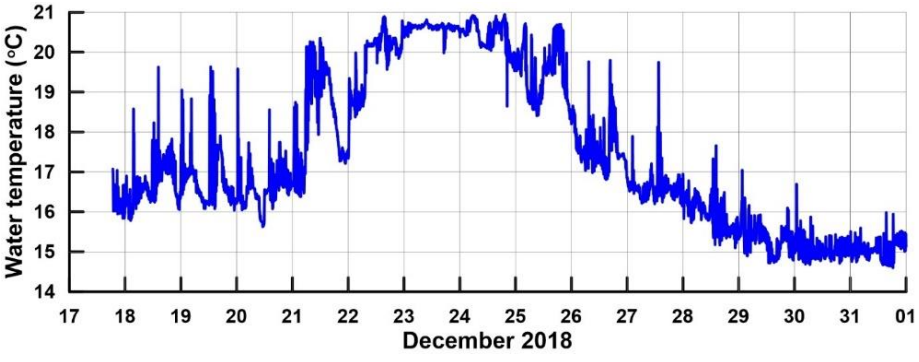


Figure 1–2. Water temperature at 5 minute intervals

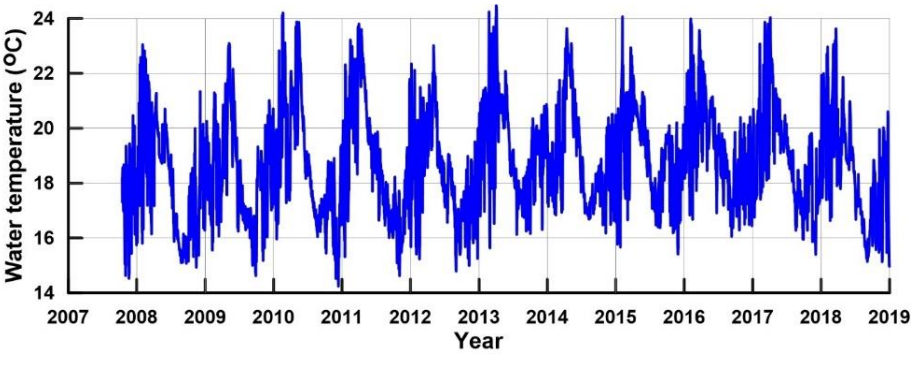


Figure 1–3. Daily water temperature over 11 years.

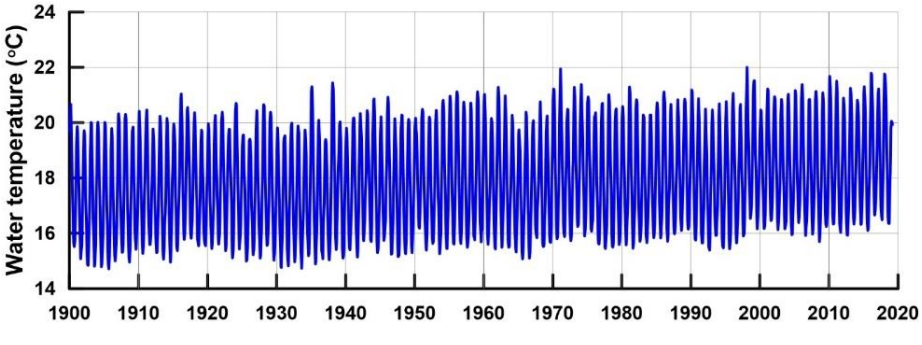


Figure 1–4. Monthly water temperature over 120 years.

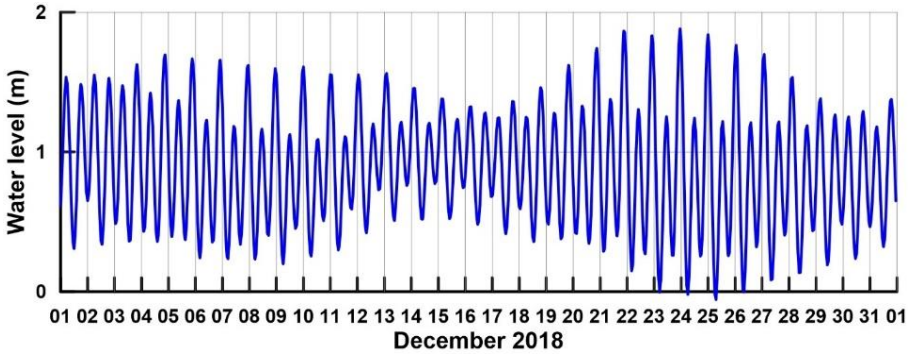


Figure 1–5. Hourly water level over 30 days.

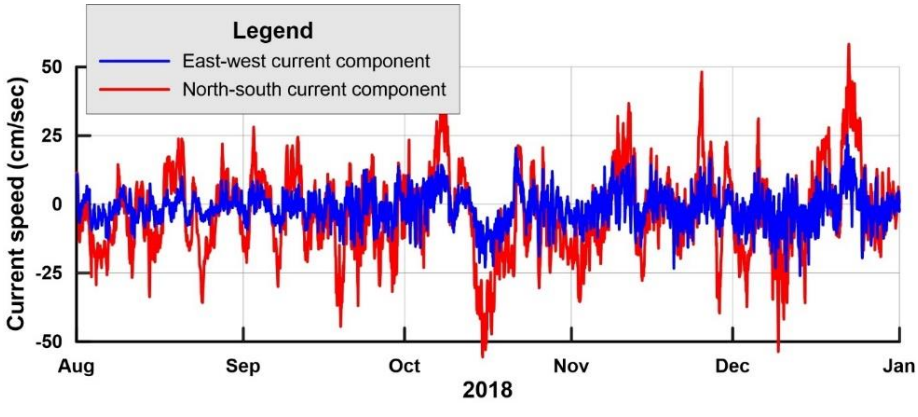


Figure 1–6. Hourly current speed and direction over 5 months.

The above examples are all taken from a fixed point in space and plotted as the relevant variable changes over time. Figure 1–7 (below) is perhaps not as intuitively obvious as the previous examples. In Figure 1–7 water temperature at 14 levels in the water column are plotted for 6 time steps, each separated by 5 minutes. The black triangles on the left hand side of the plot indicate the depths at which the temperature readings were taken. We can use a particular form of spectral analysis (known as harmonic analysis) to examine the vertical structure (in this case the vertical thermal structure) of the water and how it changes over time.

For much of this study we'll use the periodogram to estimate the spectrum.

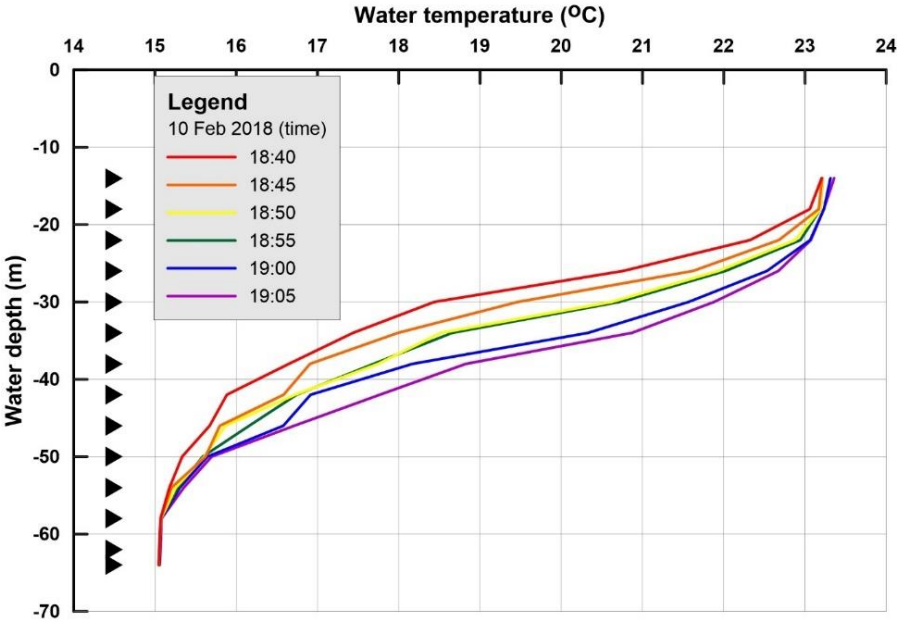


Figure 1–7. Water temperature from 14 depths in the water, taken at 5 minute intervals.

So, we have a range of data types and time scales that we'll use. These data will be used to examine the following:

- The auto spectrum (using a single time series).
- The cross spectrum (using two time series, which might include two scalars or the two components of a vector).
- Rotary spectra (using a pair of vectored time series).
- Harmonic analysis (using both a single time series and the two components of a vectored time series). This will use both temporal and spatial data sets.
- Wavelet analysis for analysing non-stationary time series (using both scalar and vectored time series).

1.2. The approach

A range of Fourier based applications using real world examples is demonstrated. Each of the main chapters starts out with some theory, generally a derivation of the equations used, followed by several examples.

Prior to undertaking a Fourier analysis, a discussion on data processing that may be undertaken to prepare the data for analysis is provided. The subsequent chapters then progress through a range of spectral analysis techniques that increase in their complexity. This is achieved through the use of many examples. Each example follows the same basic process:

- We start with a question to be addressed using spectral analysis. A clearly defined question helps focus the work required. Such questions need to be asked during the design phase of a field experiment and re-asked many times during the experiment and data analysis phases of the work.
- The data that will be used is then described, along with any processing that might undertake in the time domain.
- There are a number of tests that should be undertaken, the results from which are presented.
- A statement outlining the smoothing of the spectrum is given along with the confidence limits.
- Finally, a description of the spectrum is provided, which includes a link back to the original question(s) being asked.

We recognise that this approach will make the examples appear repetitive. We make no apologies for this. Rather than the blind application of a data analysis tool, this process helps us focus on the questions being asked, as well as the assumptions and limitations of the results that are obtained.

We should also point out here that we don't always follow the 'rules' that are often associated with applied spectral analysis. In our opinion, these rules need to be bent (or even broken) to provide practical guidance for the application of Fourier based analyses to real world problems.

Chapter 2. The Fourier Transform

This chapter provides an introduction to the Fourier transform, and includes:

- A mathematical description of applied spectral analysis and how it relates to the theory of Fourier transforms.
- Different definitions of the Fourier transform.
- The amplitude spectrum (or periodogram) and how it directly relates to the variance of the data.
- Some examples; including the spectrum of random data, what happens when we remove the trend and known signals prior to spectral analysis.

2.1. Origins of Fourier analysis

In 1807, the French mathematician Jean-Baptiste Joseph submitted a paper that would eventually become "The Analytic Theory of Heat". In this paper Fourier conjectured that any mathematical function could be represented as the sum of sine / cosine curves. Contemporary mathematicians, including Pierre-Simon Laplace and Joseph-Louis Lagrange were not convinced of Fourier's arguments supporting his conjecture. Among other queries, it appears that the objections centred around mathematical functions that contained 'sharp edges' (e.g. at the corners of a square). Earlier, Lagrange had used rational functions to identify multiple periodic signals in a data set.

Fourier resubmitted the work in 1808 and again in 1809. Again, there were objections to the paper. Fourier submitted a version of the paper to a mathematics competition in 1811-12, which he won. However, the reservations about the work remained. It's unclear what transpired with the manuscript over the next decade.

Eventually, on the recommendation of Jean Delambre, the Permanent Secretary of the French Academy of Sciences, the work was published in 1822.

It turns out that Fourier's conjecture was not entirely correct.

It was Peter Dirichlet's work in 1829 that described the conditions under which Fourier's conjecture converged. The data that we use are usually collected by instruments at discrete points in time. By fitting a sum of sine and cosine curves to the data we obtain a smooth, continuous curve to the data are largely overcome the convergence issues.

At this stage we discriminate between the Continuous Fourier Transform (CFT) and the Discrete Fourier Transform (DFT). The former generally applies to mathematical functions, while the latter refers to data collected at discrete point in time. It is the DFT that forms the basis of this work.

The computational effort in undertaking a 'Fourier analysis' was (and is) massive, repetitive and prone to errors when done by hand. It was not until the easy availability of computers that Fourier analysis could be effectively undertaken. Even so, the number of computations was large and relatively slow (compared with 'modern' computers). In the mid-1960s, Cooley and Tukey (1965) 'rediscovered' the Fast Fourier Transform (FFT). It appears that the FFT, or at least an efficient way to solve the Discrete Fourier Transform (DFT), can be traced back to the collected works of Carl Gauss in 1805 (Heideman et al 1984). (This year is interesting as it pre dates Fourier's publication date by some 17 years and his notes by 2 years). Between 1805 and 1965 contributions to efficient ways to solve the DFT were put forward by several authors including Thomson (1860), Runge (1903), Danielson and Lanczos (1942) and Good (1958).

Broadly, for a given number of data points, the greater the factorisation of that number, the more efficient is the FFT. If we have 2^n number of data points, then the FFT is at its most effective. This is referred to as the 'radix-2' FFT. (If we have 144 data points, we could set up an FFT such that its factors are 3^2 times 2^4. The coding for such an FFT would be complex and we might be better off simply using the DFT).

Not only did this speed up the analysis, it was also more accurate (fewer mathematical operations resulted in less accumulation of truncation errors). Most computer Fourier transform packages still use the 'radix-2

FFT' algorithm that Cooley and Tukey proposed, although the speed and accuracy of modern computers (perhaps) questions the need of using the FFT over the DFT.

2.2. Fourier theory

There are many good books on Fourier analysis that are available. Here we briefly introduce Fourier theory and its application to real world problems. Some of the books to which we refer the reader for additional information include: Chatfield (1996), Bloomfield (2000), Bendat and Piersol (2000) and Ramirez (1985). Warner (1998) is another favourite; it is more descriptive (i.e. fewer equations) than the other books. These are the books to which we refer most often. Having said this, everyone has different likes and dislikes – a book that resonates with us, may not do so for someone else.

Suppose we have a function, $f(\xi)$, that is continuous. The Fourier transform, $\mathcal{F}(\Lambda)$, of this function is defined as:

$$\mathcal{F}(\Lambda) = \int_{-\infty}^{\infty} f(\xi)\, e^{-i\omega\xi}\ d\xi$$

The effect of this operation is to convert the data set from one based on the variable ξ to one based on Λ. An inverse Fourier transform (which converts back from Λ to ξ) is then defined as:

$$f(\xi) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \mathcal{F}(\Lambda) e^{i\Lambda\xi}\ d\Lambda$$

In many (perhaps most) applications the dependent variable, ξ, is time (t) and the transformed variable, Λ, is frequency. We can rewrite the above function as, $f(t)$, and its Fourier transform, $\mathcal{F}(\omega)$, is defined as:

$$\mathcal{F}(\omega) = \int_{-\infty}^{\infty} f(t)\, e^{-i\omega t}\ dt$$

The effect of this operation is to convert the data set from one based on time t to one based on frequency $\omega = {}^{1}/_{t}$. An inverse Fourier transform (which converts from frequency ω back to time t) is then defined as:

$$f(t) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \mathcal{F}(\omega) e^{i\omega t}\ d\omega$$

This is the definition used in Chatfield (1996, Appendix A) and it will be used here. However, other definitions for the Fourier and inverse Fourier transforms are used by others, e.g.

Churchill (1963)

$$\mathcal{F}(\omega) = \frac{1}{\pi} \int_{-\infty}^{\infty} f(t)\, e^{i\omega t}\ dt \quad and \quad f(t) = \frac{1}{2} \int_{-\infty}^{\infty} \mathcal{F}(\omega) e^{-i\omega t}\ d\omega$$

Carrier, Krook and Pearson (1966)

$$\mathcal{F}(\omega) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} f(t)\, e^{i\omega t}\ dt \quad and \quad f(t) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \mathcal{F}(\omega) e^{-i\omega t}\ d\omega$$

Lighthill (1978)

$$\mathcal{F}(\omega) = \int_{-\infty}^{\infty} f(t)\, e^{i2\pi\omega t}\ dt \quad and \quad f(t) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \mathcal{F}(\omega) e^{-i2\pi\omega t}\ d\omega$$

Bendat and Piersol (1996)

$$\mathcal{F}(\omega) = \int_{-\infty}^{\infty} f(t)\, e^{-i2\pi\omega t}\, dt \quad and \quad f(t) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \mathcal{F}(\omega) e^{i2\pi\omega t}\, d\omega$$

Other definitions may also be used. The different definitions are largely a personal preference. While the principles of these definitions are the same, the values of the numerical results will be different (by a constant and/or the sign of the complex part). The only limitation on the constants outside the integral signs is that the product of the transform and inverse transform pair must be $\frac{1}{2\,\pi}$. Sometimes the 'symmetric' form of the Fourier pair is used: viz. the coefficients for the transform / inverse transform pair are each $\frac{1}{\sqrt{2\,\pi}}$ (e.g. Spiegel, 1971). It should also be noted that some authors omit these coefficients altogether – after all, it is only a constant offset to the transform / inverse transform pair and will not impact, in a meaningful way, the end result.

There are 8 definitions of the Fourier transform / inverse transform pair that are commonly used: $\left(1, \frac{1}{2\pi}\right)$, $\left(\frac{1}{2\pi}, 1\right)$, $\left(\frac{1}{\sqrt{2\pi}}, \frac{1}{\sqrt{2\pi}}\right)$ and $\left(\frac{1}{2}, \frac{1}{\pi}\right)$ each coupled with $\left(e^{-i\omega t}, e^{i\omega t}\right)$ or $\left(e^{i\omega t}, e^{-i\omega t}\right)$. Which definition is chosen is simply a matter of personal preference, however, the definition that has been used should always be explicitly stated.

One final point to note is that the Fourier transform (or the Fourier representation of the signal) has been integrated over all time. In other words, the frequency applies to all times. The frequency component of a signal does not change. The implication is that the data set is 'stationary' (more of that later) which is an important concept in, and the reason for, the development of the wavelet transform (Chapter 8).

If space is the dependent variable instead of time, the Fourier transform / inverse transform pair can be written as:

$$\mathcal{F}(\kappa) = \int_{-\infty}^{\infty} f(x)\, e^{-i\kappa x}\ dx \qquad and \qquad f(x) = \frac{1}{2\pi}\int_{-\infty}^{\infty} \mathcal{F}(\kappa) e^{ix\kappa}\ d\kappa$$

Where x is the spatial function and $\kappa(= {}^{1}/_{x})$ is designated the 'wave number' (analogous to frequency in the time domain). Irrespective of the type of dependent variable being used, the mathematics is identical. Focus here is on the time-frequency domain, which is (possibly) a little easier to understand and interpret than the space-wave number domain.

2.3. Discrete Fourier transform

Real data sets are rarely continuous. Mostly they consist of discrete values at specific points in time (or space) over a fixed duration (or a fixed space). In such cases, the integrals are replaced by summations. However, because real data sets are finite in length, the limits become finite.

In much the same way that 'n' data points can be represented exactly by a polynomial of order 'n-1', 'N' data points can be represented exactly by 'N' sine / cosine curves. The finite Fourier transform (or discrete Fourier transform, DFT) of the time series x_t or $x(t)$ can be defined as:

$$x_t = \sum_{p=0}^{\frac{N}{2}} \left[a_p\, cos\left(\frac{2\pi pt}{N}\right) + b_p\, sin\left(\frac{2\pi pt}{N}\right)\right] \quad for \quad t = 1, 2, 3, \ldots N$$

Note, the upper limit of the sum is $\frac{N}{2}$ rather than N. If $p > \frac{N}{2}$ we find that the trigonometric functions are repeated, albeit with a negative sign associated with the sin term.

If $p = 0$ then:

$$x_t = \left[a_o\, cos\left(\frac{2\pi\ 0\ t}{N}\right) + b_0\, sin\left(\frac{2\pi\ 0\ t}{N}\right)\right] \quad for \quad t = 1, 2, 3, \ldots N$$

$$\Rightarrow \quad \sum_{t=1}^{N} x_t = \sum_{t=1}^{N} [a_o] = N a_o$$

$$or \quad a_o = \frac{1}{N}\sum_{t=1}^{N} x_t \;=\; \langle x \rangle$$

That is, the leading term (i.e. when $p = 0$) associated with the cosine component is the average of the data.

If $p = N/2$ then:

$$x_t = \left[a_{N/2}\, cos\left(\frac{2\pi\frac{N}{2}t}{N}\right) + \; b_{N/2}\, sin\left(\frac{2\pi\frac{N}{2}t}{N}\right)\right] \quad for \quad t = 1, 2, 3, \ldots N$$

$$x_t = \left[a_{N/2}\, cos(\pi t) + \; b_{N/2}\, sin(\pi t)\right] \quad for \quad t = 1, 2, 3, \ldots N$$

$$x_t = \left[a_{N/2}\, cos(\pi t)\right] \quad for \quad t = 1, 2, 3, \ldots N$$

Hence:

$$\sum_{t=1}^{N} x_t = \sum_{t=1}^{N}\left[a_{N/2}\, cos(\pi t)\right]$$

$$\Longrightarrow \quad \sum_{t=1}^{N} cos(\pi t)\; x_t = \sum_{t=1}^{N}\left[a_{N/2}\, cos^2(\pi t)\right]$$

$$\sum_{t=1}^{N}(-1)^t\, x_t \;=\; a_{N/2}\sum_{t=1}^{N}[\, cos^2(\pi t)] \;=\; a_{N/2}\sum_{t=1}^{N}[\,1] \;=\; N a_{N/2}$$

$$or \quad a_{N/2} = \frac{1}{N}\sum_{t=1}^{N}(-1)^t\, x_t$$

Discrete Fourier transform summary:

$$x_t = a_o + \sum_{p=1}^{\frac{N}{2}-1}\left[a_p\, cos\left(\frac{2\pi pt}{N}\right) + \; b_p\, sin\left(\frac{2\pi pt}{N}\right)\right] + \; a_{\frac{N}{2}}\, cos(\pi t)$$

$$for \quad t = 1, 2, 3, \ldots N$$

Where:

$$a_o = \frac{1}{N}\sum_{t=1}^{N} x_t \; = \; \langle x \rangle = \; the \; data \; average$$

$$a_{N/2} = \frac{1}{N}\sum_{t=1}^{N} (-1)^t \, x_t$$

$$a_p = \frac{2}{N}\sum_{t=1}^{N}\left[x_t \, cos\left(\frac{2\pi pt}{N}\right)\right] \quad for \quad p = 1, 2, 3, \ldots . \left(\frac{N}{2} - 1\right)$$

$$b_p = \frac{2}{N}\sum_{t=1}^{N}\left[x_t \, sin\left(\frac{2\pi pt}{N}\right)\right] \quad for \quad p = 1, 2, 3, \ldots . \left(\frac{N}{2} - 1\right)$$

This is the expression used in Chatfield (1996) and it will be used here. The Fourier coefficients a_p and b_p can also be expressed as a complex pair:

$$a_p + ib_p = \frac{2}{N}\sum_{t=1}^{N} x_t \left[cos\left(\frac{2\pi pt}{N}\right) + i \, sin\left(\frac{2\pi pt}{N}\right)\right] \quad for$$

$$p = 1, 2, 3, \ldots . \left(\frac{N}{2} - 1\right)$$

$$a_p + ib_p = \frac{2}{N}\sum_{t=1}^{N} x_t \left[e^{i\frac{2\pi pt}{N}}\right] \qquad for \quad p = 1, 2, 3, \ldots . \left(\frac{N}{2} - 1\right)$$

The FFT using Excel

The Excel FFT routine appears to use a slightly different formulation. Viz:

$$a_p - ib_p = \frac{2}{N}\sum_{t=1}^{N} x_t \left[e^{-i\frac{2\pi pt}{N}}\right] \qquad for \quad p = 1, 2, 3, \ldots . N$$

The Excel output includes all frequencies, even though the (reverse order of the) second half of the output is simply the complex conjugate of the first half of the output.

The imaginary part of the Excel output is defined as 'negative', compared with Chatfield's (1996) formulation.

There is no problem with this, we just need to take care when comparing the output from different Fourier transform programs. This difference (or ambiguity) in the Fourier transform definition will impact the phase lag in the auto-spectrum, but becomes more apparent when we look at more complex applications such as cross-spectra and rotary spectra.

The frequency ω_p (in radians per unit time) is given for each 'p' value (called the p^{th} harmonic), by:

$$\omega_p = \frac{2\,\pi\,p}{N\,\Delta t} \qquad for \qquad p = 1, 2, 3, \ldots . \left(\frac{N}{2} - 1\right)$$

Often, we find it more convenient to use frequency in cycles per unit time (rather than radians per unit time) in which case the frequency f_p is given (in cycles per unit time), for each 'p' value, by:

$$f_p = \frac{p}{N\,\Delta t} = \frac{\omega_p}{2\,\pi} \qquad for \qquad p = 1, 2, 3, \ldots . \left(\frac{N}{2} - 1\right)$$

The amplitude (R_p) and phase lag (ϕ_p)at each 'p' value is given by:

$$R_p = \sqrt{a_p^2 + b_p^2} \qquad or \qquad R_p^2 = \left(a_p^2 + b_p^2\right)$$

$$\phi_p = tan^{-1}\left(\frac{-b_p}{a_p}\right)$$

Parseval's theorem relates the amplitude to the variance of the data set via the following:

$$\frac{1}{N}\sum_{t=1}^{N}(x_t - \langle x \rangle)^2 = \frac{1}{2}\sum_{p=1}^{\frac{N}{2}-1} R_p^2 + a_{N/2}^2$$

That is: the (square of the) amplitude of each harmonic provides the relative contribution of that harmonic to the (biased[1]) variance of the data set. Parseval's theorem may be thought of as a statement that 'energy is conserved irrespective of whether we're in the time domain (variance) or the frequency domain (the sum of the squared spectral estimates)'.

Plotting the amplitude as a function of the frequency results in the 'periodogram' or 'spectogram' or simply the 'spectrum'. (The period is just the inverse of frequency). There are other ways of estimating the spectrum but the periodogram is, perhaps, the most common approach. Using the Fast Fourier Transform (FFT) routine in Excel, the variance of the time series is given by:

$$\sigma^2 = \frac{1}{N^2}\sum_{p=1}^{N} R_p^2 = \frac{1}{N^2}\sum_{p=1}^{N}\left(a_p^2 + b_p^2\right)$$

The advantage in using this form of the amplitude (i.e. dividing the amplitude squared by N^2) is that we can allocate the energy at a particular frequency to its relative contribution to the variance in the data. For the periodogram shown in Figure 2–6, we can see that there is a large peak at one cycle per year (i.e. the annual signal). Examination of the spectral estimates shows that this signal contributes about 65% to the total variance in the data.

2.4. The Fast Fourier Transform

The computation effort in generating a Fourier transform is considerable. To perform a Fourier transform of a time series comprising N data points requires of the order N^2 mathematical operations. Doing it by hand is tedious and repetitive – transforming as few as 32 data points requires

[1] Usually, the divisor in the variance is $N-1$ (i.e. the variance has $N-1$ degrees of freedom. If we know the mean of the data set, we only need to know the values of $N-1$ data points to know the entire data set). Here the divisor is N are the variance is regarded as 'biased'. However, if N is large (as it often will be), dividing by $N-1$ or N makes only a very small difference to the variance.

more than 1,000 mathematical operations. The Fast Fourier Transform (FFT) is a technique that considerably reduces the number of mathematical operations that are required to run the DFT.

The FFT is simply an efficient way of coding the DFT when the number of data points in the time series can be factorised. FFT algorithms can be designed for any number of data points. The greater the factorisation, the more efficient (i.e. the fewer number of mathematical operations) in the algorithm. The most efficient FFT algorithm is when there are 2^n data points. This is known as a 'radix 2' FFT and is the most common of all the FFT routines.

Unless explicitly stated otherwise, the 'radix 2 FFT' will be used throughout this book and will be referred to simply as the 'FFT'.

There are several 'you-tube' sites that give a good description of the mechanics of the FFT. An outline, with examples, is provided in Chapter 12 (Appendix A).

2.5. What to expect

Where do we want to end up? What do we expect to see from our spectral analysis? We know that spectral analysis changes our data set from one in which time is the independent variable to one in which frequency is the independent variable. We can do this in one of two ways. We can:

(a) Remove the trend as well as any other repetitive signals from the data set (e.g. annual signals, tidal signals), resulting in a set of (ideally) random data that are normally distributed. The spectral analysis is then undertaken on the remaining data to identify any other signals of which we may not be aware.
(b) Remove the trend then undertake the spectral analysis to identify important, repetitive signals in the data set and the frequencies at which they occur.

Which approach we use will depend on the question that we're asking. Knowing, identifying and removing all of the known (or important) signals in a real data set is difficult, and it may not be achievable in a

practical sense. For the present, we'll focus on a theoretical data set and ask: 'What happens when we subject a data set of normally distributed random samples to a spectral analysis'? This will give us a feel for what to expect when we've removed all of the relevant signals from our data.

The Excel function 'NORM.INV(randbetween(1,9999)/10000,0,1)' was used to select 4,096 (i.e. 2^{12}) random numbers from a normal distribution with mean zero and standard deviation of one. (This random number generator is sufficiently random for the purposes here). For convenience, it is assumed that these data are 4,096 daily water temperature anomalies. These might represent the data we obtain after removing the linear trend, annual signal and any other known processes from the original data set, producing a set of data anomalies. These water temperature anomalies are plotted as a function of time in Figure 2–1. The red dashed line represents the linear trend (which has zero slope).

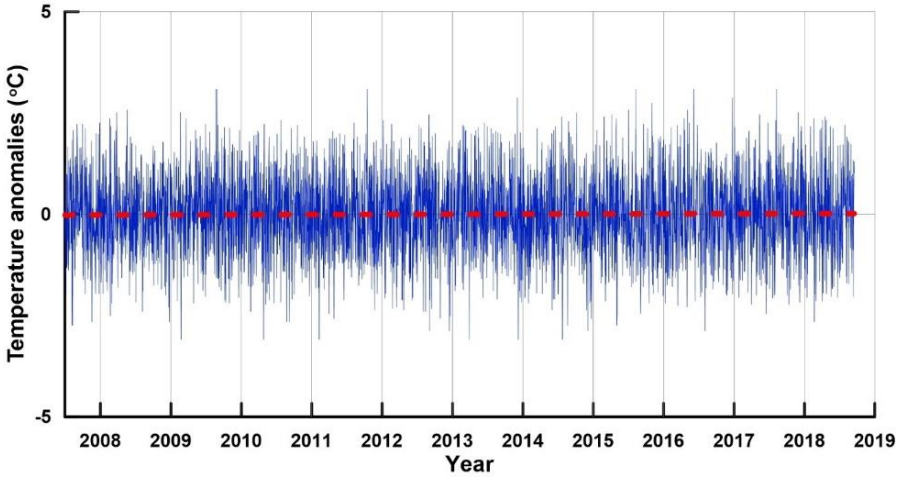


Figure 2–1. Time series of 4096 daily random 'water temperature anomalies'.

These daily water temperature anomalies were subject to a FFT (of length 4,096 days) using the Excel routine 'Fourier analysis' under the 'Data Analysis' tab. The results from the FFT are shown in Figure 2–2.

The first thing to note from the resulting periodogram is that there are two plots: the spectral estimates (also called the spectral energy, amplitude spectrum or power spectrum) and the phase lag (or phase) spectrum. The spectral estimates indicate how much energy is contained at each frequency while the phase lag spectrum tells us the 'starting position' of each sine curve. For example, the sine curve with a frequency near 0.1 cycles per year has a phase lag of 90^o i.e. the starting point for the sine curve with frequency of 0.1 cycles per year is 90^o in advance of zero.

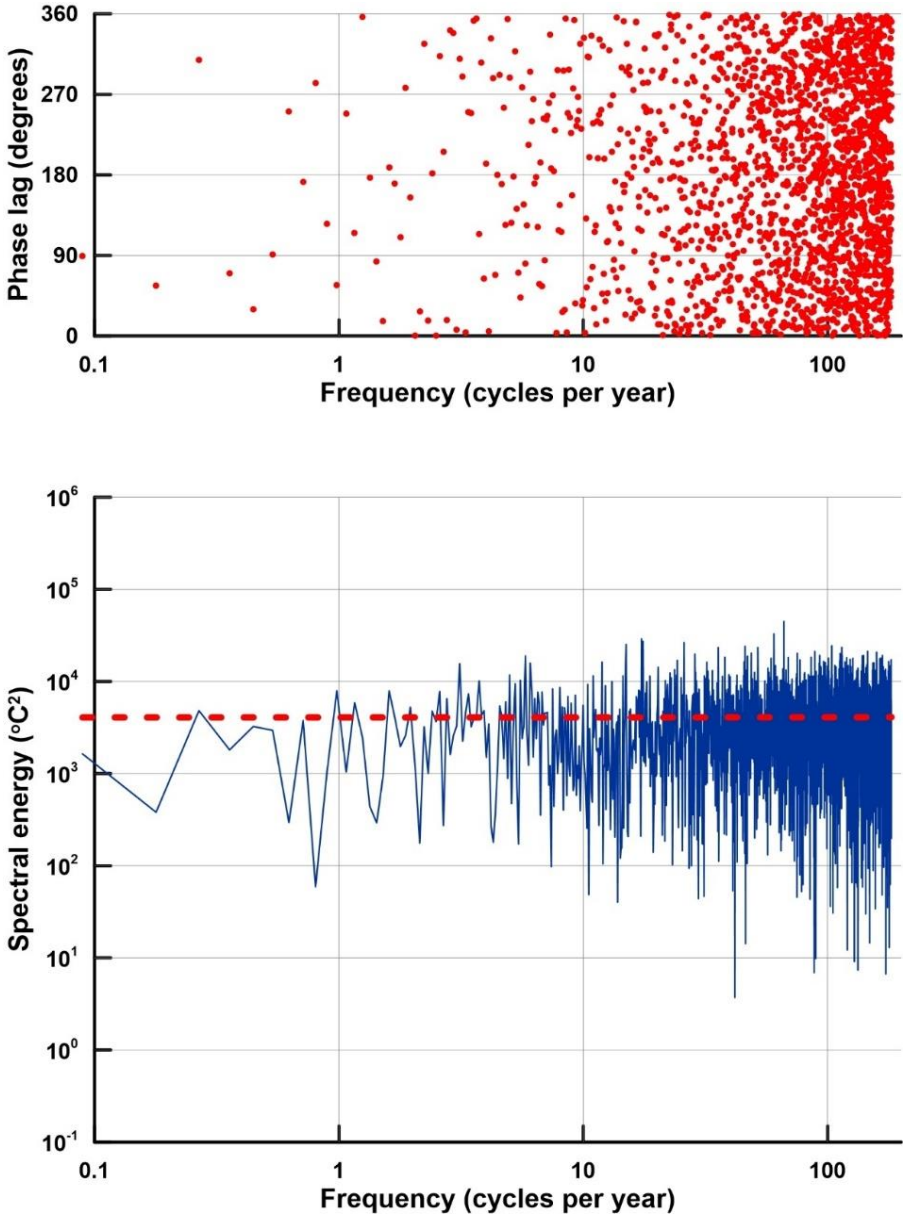


Figure 2–2. Spectral analysis of the temperature data shown in Figure 2–1.

The red dashed line in the spectral estimates plot is the best fit (in a least squares sense) power curve fit to the spectrum. The spectral estimates vary by about two-to-three orders of magnitude. However, the best fit power curve is relatively flat. This is indicative of what the Fourier transform of a random signal should look like – flat, featureless, no high energy spectral peaks.

The phase lag spectrum is also featureless with the phase lag seemingly random, or at least with no clearly observable pattern. In general, the phase lag spectrum is not used as often as the amplitude spectrum.

When we see a spectrum such as that shown in Figure 2–2 we know that there are no processes in the data set – it is purely random (a white noise spectrum). The energy spectrum is flat, although the variability may cover several orders of magnitude, and the phase lag spectrum has no discernible pattern. Ideally, this is what we want to target. However, it is almost impossible to identify, justify and remove all processes from a real data set. We're usually left with some residual signal in our data.

A limitation of the Excel FFT routine

The Excel FFT routine is limited to 4,096 data points*. At first glance this might appear restrictive. However, if we're interested in long period (i.e. low frequency) features, there is no need to sample frequently – we just need to sample over a long period of time. For example, if we're interested in annual signals, then we need to sample over many years. Sampling every few seconds or even daily does not really contribute to the understanding of our feature of interest. Hence, we can sample, say, monthly and cover almost 350 years of data.

If we're interested in high frequency features, then we need to sample frequently, not necessarily over a long period of time. For example, if we're interested in what's happening on a daily time scale, then we don't need to sample for many years. But we do need to sample at, say, hourly intervals, to make sense of what's happening at the daily scale.

Rarely, if at all, would we need to look simultaneously at both high and low frequency features. So, limiting the FFT to a maximum of 4,096 data

points is not really all that restrictive. It largely depends on the question we want to answer.

*Note: Apparently it is possible to extend the Excel Fourier Transform algorithm beyond this limit, although this is not something that we have tried.

So, in theory, we remove the trend and the processes from the data set and we should arrive at the FFT depicted in Figure 2–2. What happens in practice?

We repeat this analysis using a real data set – daily ocean water temperature from mid-2006 to mid-2017 (4,096 days), as shown in Figure 2–3. We'll examine the data set in three ways: (a) analysing the raw data set, (b) removing the trend from the data, and (c) removing the trend and the annual signal. Data are from a location in the Southern Hemisphere so the annual maximum ocean water temperature occurs in the early part of the year (noting that the ocean seasons lag the atmospheric seasons by about 1-2 months).

Several features are apparent in Figure 2–3.

- There is a strong signal in the data that appears at, or close to, the annual period.
- The data are quite noisy – smoothing the data will reduce the noise, although there is a danger that too much smoothing will remove part of the signal in which we are interested.
- Over the period examined, there is a clear, increasing, trend in the data (the red, dashed line in Figure 2–3.

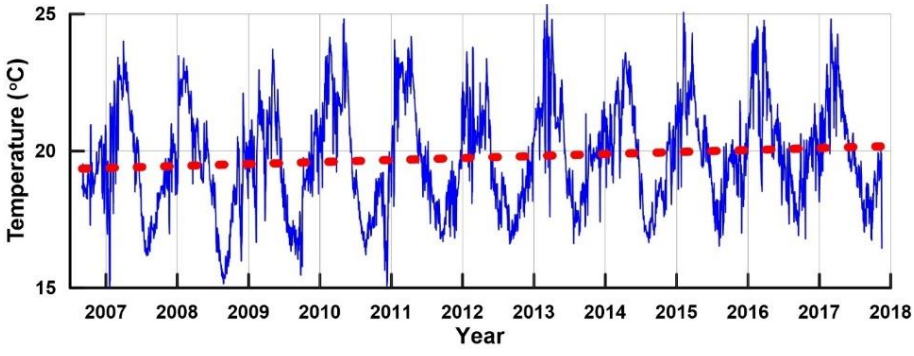


Figure 2–3. Time series of daily water temperature – 4,096 days from 2006 to 2017. The red dashed line denotes the linear trend.

For the moment, we'll show the results of the FFT on these data without any additional processing – just to give an initial feel for the resulting spectrum. Later, we'll repeat the analyses showing what processing should be (or must be) done before undertaking a spectral analysis.

The periodogram corresponding to the data in Figure 2–3 is given in Figure 2–4. There is large spectral energy at low frequency, a result of retaining the trend in the data. There is a clear peak in the periodogram at a frequency of one cycle per year (viz. the annual temperature signal). The size of this peak is about one order of magnitude greater than any other peaks (at higher frequencies) in the spectrum, but about two orders of magnitude less than the spectral energy at low frequency. From Parseval's theorem, the low frequency energy (i.e. that from the trend) accounts for >99% of the variance in the data set. At frequencies greater than one cycle per year the spectral energy decreases as the frequency increases. This is in contrast to the spectrum of random values (Figure 2–2) and suggests that there is something non-random happening here. It turns out that this spectral slope is common in ocean processes and is the result of energy 'cascading' from low frequencies to higher frequencies (but more on this later).

The phase lag spectrum (top plot in Figure 2–4) appears similar to the spectrum of random data shown in Figure 2–2. There appears to be little

structure in the phase lag spectrum and interpretation is more difficult than the energy spectrum.

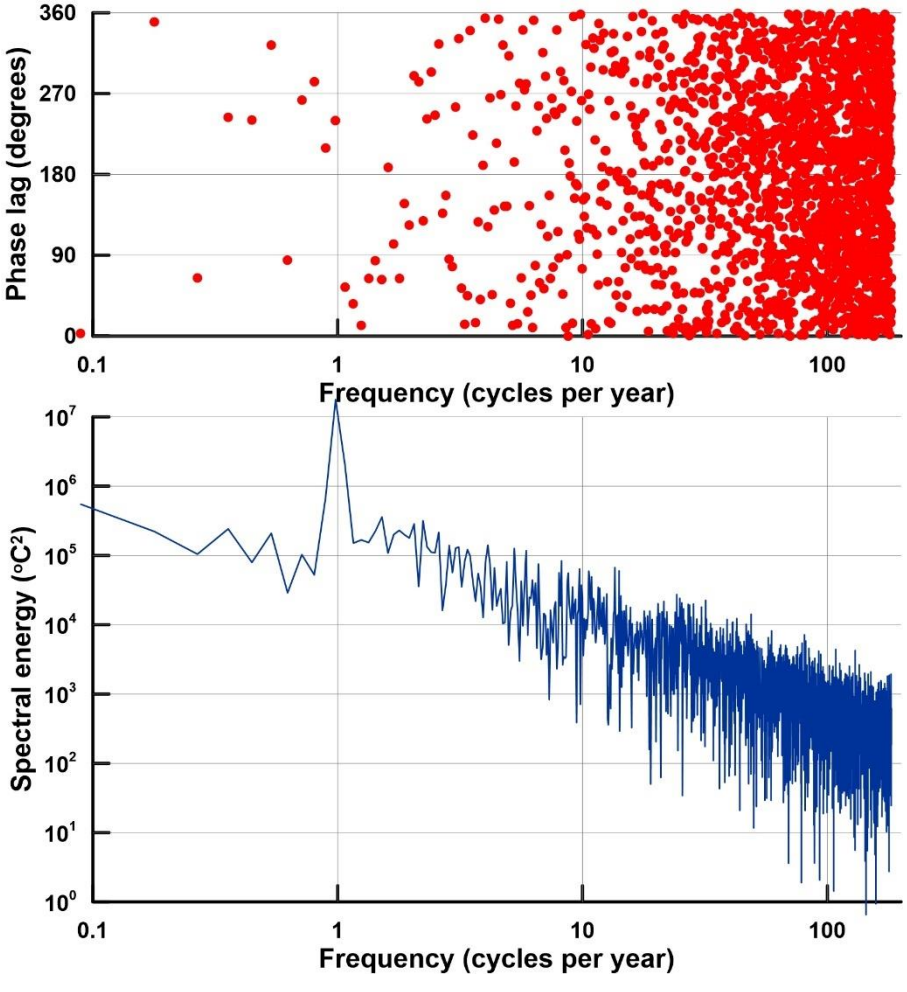


Figure 2–4. Spectral analysis of the temperature data shown in Figure 2–3.

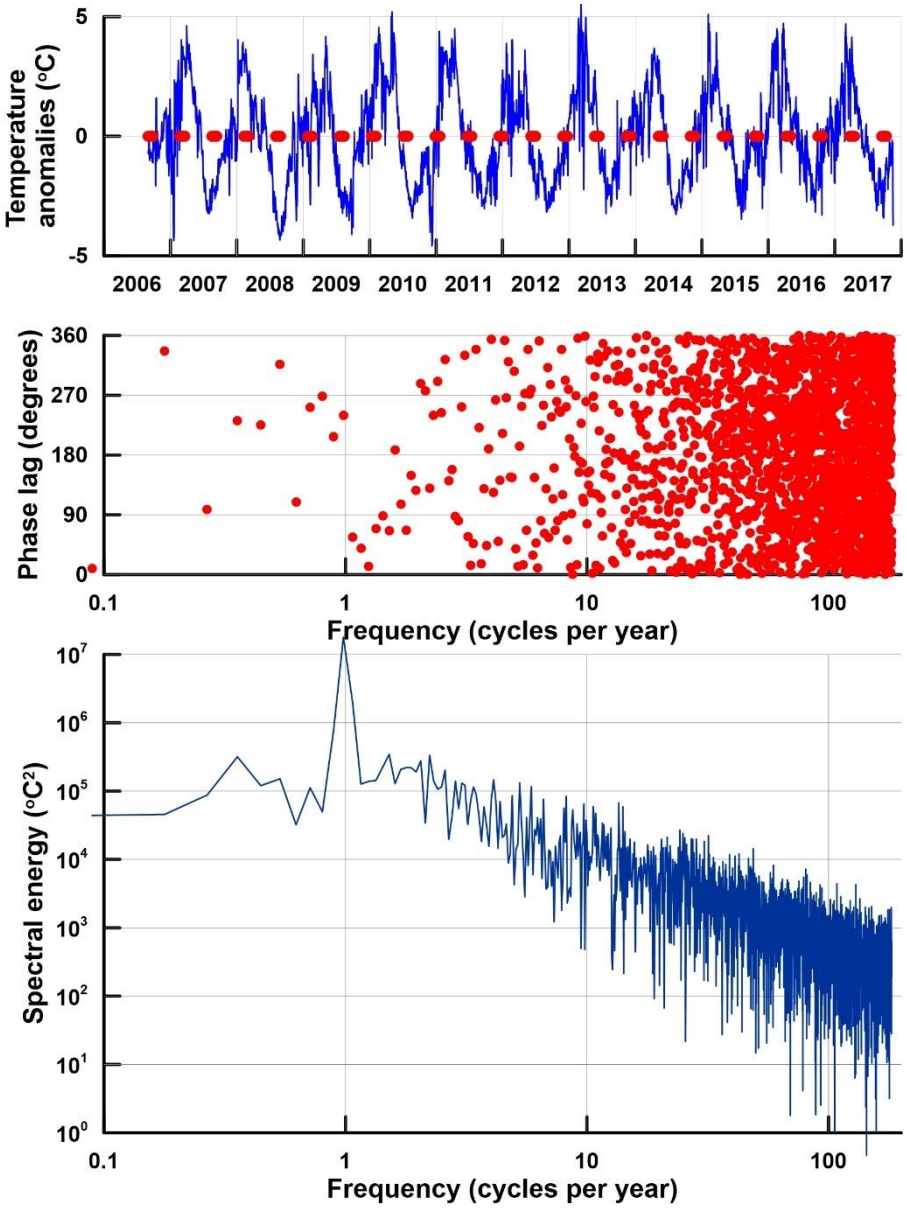


Figure 2–5. Top: Time series of daily water temperature – 4,096 days from 2006 to 2017 with the linear trend removed. Bottom and middle: Amplitude and phase lag spectra of the data.

The data from Figure 2–3 with the linear trend removed (resulting in temperature anomalies) are replotted in Figure 2–5 (top plot) together with the corresponding amplitude and phase lag spectra (bottom and middle plots). The seasonal signal is still apparent, however, the trend line (the red dashed line) in Figure 2–3 has been removed and the slope of the new trend line is zero. The resulting periodogram is also shown in Figure 2–5 (bottom and middle plots). Removing the linear trend mainly impacts the low frequency end of the spectrum. Comparing the low frequency ends of Figure 2–4 with Figure 2–5, it is seen that the spectral energy in the latter figure is considerably lower than that obtained when the trend is retained (Figure 2–4). Without the trend, the peak in Figure 2–5 is emphasised. From Parseval's theorem, 66% of the variance occurs near the dominant (annual) spectral peak. There is a marked difference to the contribution to the variance from different spectral estimates when the trend is removed.

The differences in the phase lag spectra (with and without the trend, top plot in Figure 2–4 and middle plot in Figure 2–5) are relatively small.

Finally, we remove the annual and semi-annual signals (as well as the trend) from the data time series. The time series of temperature anomalies is shown in Figure 2–6. The trend line (red dashed line) has a slope of zero and there are no consistent patterns in the data although excursions of up to +/-5^{o}C can be seen.

Why did we remove **both** the annual and semi-annual signals? Sometimes there is 'leakage' of energy from the dominant frequency to higher frequencies. To remove the annual signal (as in this case) it is advisable to remove the annual signal as well as the 'harmonics' i.e. the signals that occur twice per year, 4 times per year, 8 times per year and so on. In most cases removing the dominant signal (once per year or annual period) and its first harmonic (twice per year or semi-annual period) is sufficient to remove the annual signal.

The corresponding spectrum is also shown in Figure 2–6 (bottom and middle plots). At low frequency the spectral energy is low. There is a broad region of higher spectral energy between 0.1 and 10 cycles per year,

although no spectral peaks are obvious. At higher frequencies the spectral energy decreases as frequency increases.

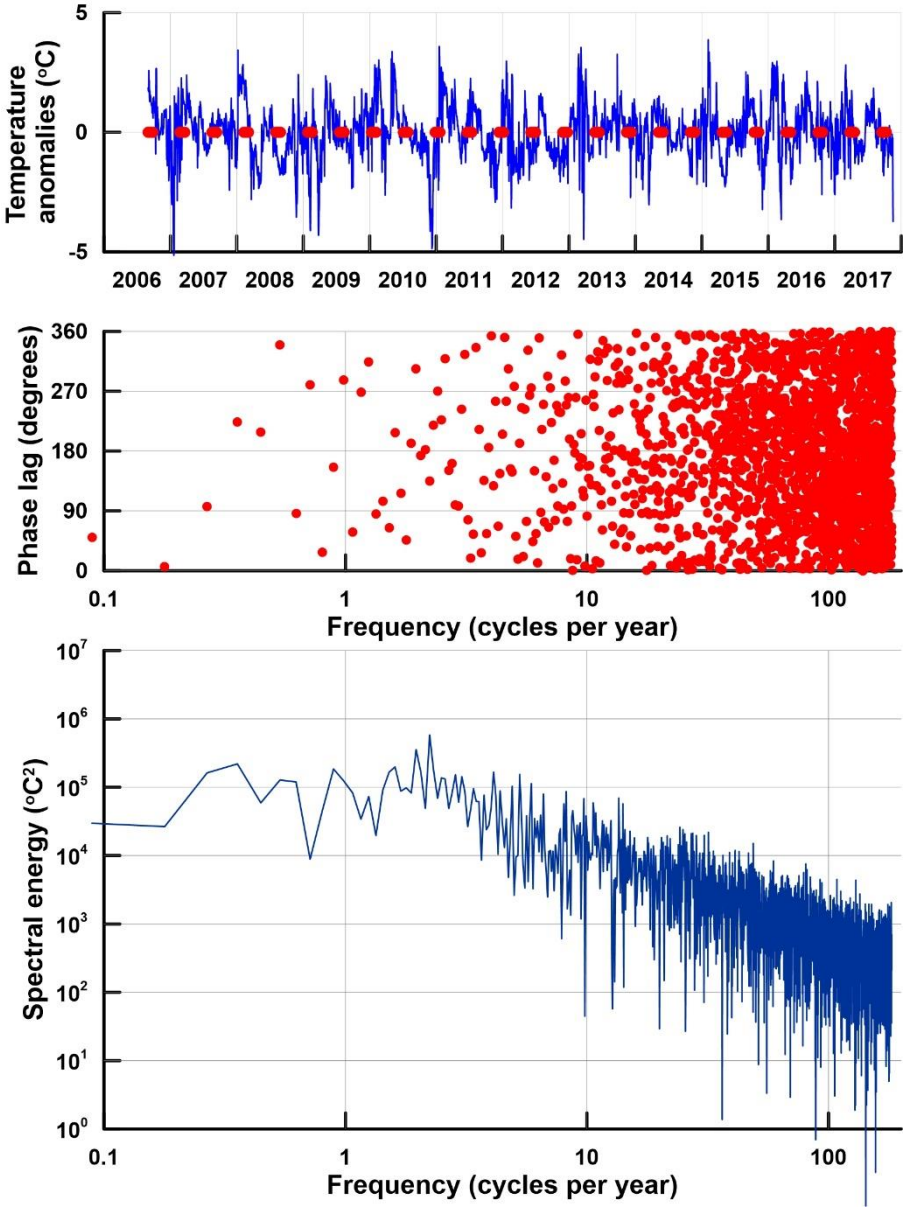


Figure 2–6. Top: Time series of daily water temperature – 4,096 days from 2006 to 2017 with the linear trend and the seasonal signals

removed. Bottom and middle: Amplitude and phase lag spectra of the data.

The flat spectrum obtained using random data (Figure 2–2) is not apparent suggesting that there is still some unresolved information in the data. Real data (e.g. temperature, water level) show spectra similar to those seen in Figure 2–5 and Figure 2–6. Even after removing the known processes, we still don't get a flat spectrum as was obtained using random samples from the normal distribution (Figure 2–2).

Differences in the phase lag spectra of Figure 2–5 and Figure 2–6 are small.

In summary, the spectrum using the 4,096 data point time series in Figure 2–3 is shown in Figure 2–6 with the major known signals identified and removed. Spectral energy is plotted as a function of frequency. It is common to plot the spectrum on a log-log axis because:

- Under most conditions, the spectral energy (and frequency) vary over several orders of magnitude.
- It tends to best highlights the important features of the spectrum.
- It is convenient when generating confidence limits, which we'll show later.

What we look for in the spectrum are peaks – regions of high spectral energy (compared with the surrounding parts of the spectrum) and the frequency at which they occur. We then want to interpret these features e.g. can we attribute a large spectral peak at a particular frequency with a physical process?

The frequency is given at each point 'p', by:

$$f_p = \frac{p}{N\,\Delta t} \qquad for \qquad p = 1, 2, 3, \ldots . \left(\frac{N}{2}\right) - 1$$

Excel returns the spectral estimates as complex numbers for all 4,096 data points. However, the second half of these are the complex conjugates of the first half (in reverse order), so they have the same amplitudes. It is usual to only plot for first half of the spectrum. The amplitude (Amp_p) squared or spectral energy at each 'p' value is given by:

$$Amp_p^2 = \left(a_p^2 + b_p^2\right) \qquad for \qquad p = 1, 2, 3, \ldots\ldots, \left(\frac{N}{2} - 1\right)$$



We're now in a position to generalise what we've found thus far. This is demonstrated, in part, using Figure 2–7 from which we note the following:

- The highest spectral energy occurs at low frequency. This is because we did not remove the trend – which we should do before we undertake a spectral analysis. Otherwise, the spectral energy at low frequency will mask other features in the spectrum.
- If the amplitude spectrum is 'relatively' flat and the phase lag spectrum does not appear to have any patterns, the data set comprises random features only. There are no important signals in the data. The amplitude spectrum may still vary by an order of magnitude or more, but a random process will result in a relatively flat spectrum. In most practical data sets, we'll rarely find a spectrum that is flat.
- There is a relatively large peak in the spectrum near the frequency of one cycle per year. Spectral peaks indicate an important process occurs in the data at the frequency at which the spectral peak is observed. In this case, the spectral peak coincides with the annual (temperature) signal and it is something that we should expect in a water temperature data set that spans 10 years.
- With a strong annual signal, we sometimes see a second spectral peak at two cycles per year and a third spectral peak near four

cycles per year. This results from 'leakage' of a dominant signal into its harmonics.

- The general slope of the spectrum follows the '-5/3^{rds}' power law. We'll discuss this later, but the main part of many spectra of physical processes follows this slope.
- There is considerable noise at higher frequencies in the spectrum – perhaps 2 or 3 orders of magnitude difference in spectral energy near any particular frequency.

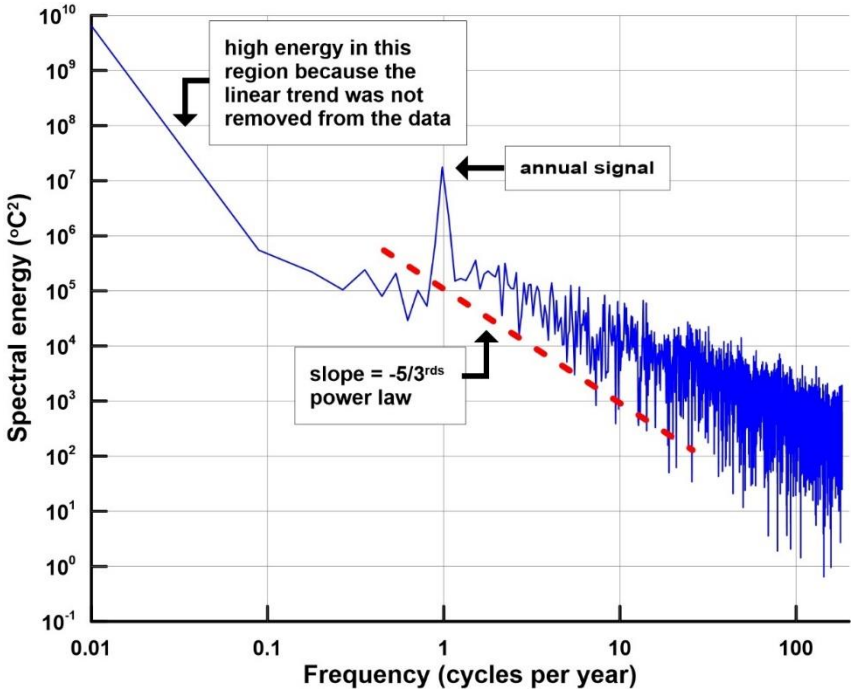


Figure 2–7. Spectral analysis of the temperature data shown in Figure 2–3.

Like many mathematical problems, the mechanics of spectral analysis is relatively straightforward. The real difficulty lies in how to prepare the data before we undertake the spectral analysis and how to interpret the resulting periodogram. Particularly, how do we know whether a peak is important or not? Much of this depends on the question that we're asking.

In practical terms, we're ultimately heading towards the periodogram shown in Figure 2–6, which is based on the time series shown at the top of the same figure. The trend has been removed, as has the annual signal.

At the low frequency end of the spectrum, and if the trend has been removed, the periodogram may have a positive slope or it may be relatively flat. At higher frequencies, the periodogram loses energy as the frequency increases (the slope of the spectrum is negative). Between two limits (known as the inertial subrange) the rate at which the spectral energy decreases as a function of frequency follows, on average, a '$-5/3^{rds}$ power law'.

Inertial subrange

The rate at which energy is lost obeys the '$-5/3^{rds}$ power law'. That is, the slope of the spectrum falls off such that:

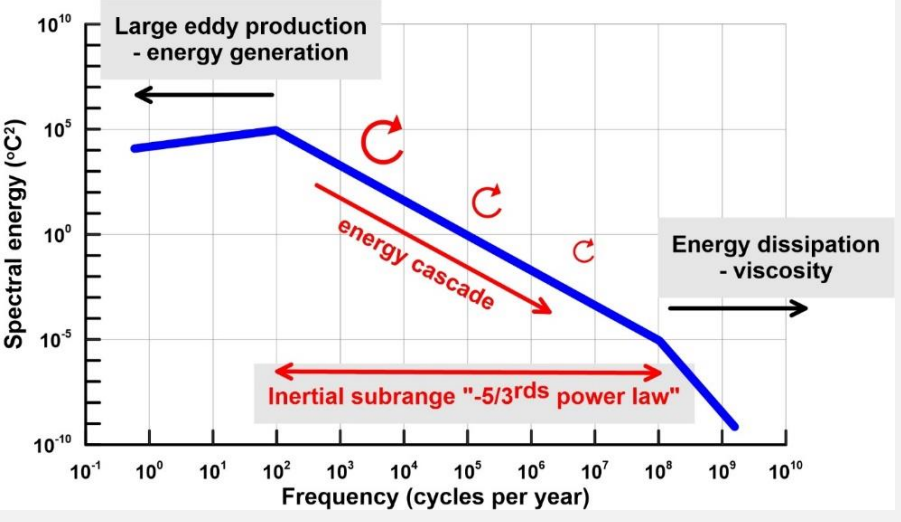
$$Spectral\ energy \propto frequency^{-5/3}$$

This power law has its origins in fluid dynamics. In the 1920s L.F. Richardson (among others) hypothesised that energy enters a system at the low frequency end of the spectrum and is dissipated at high frequency via viscosity. The spectral region in between these two limits effectively transfers the energy from low to high frequency and the process is referred to as the 'energy cascade'.

In the 1940s A. Kolmogorov made the assumption that the energy-generating part of the spectrum (i.e. where large eddies exist) was anisotropic (fluid properties are not the same in all directions). As the energy cascades down to higher frequencies, Kolmogorov assumed that these large eddies lose their structure and become isotropic. At some stage (viz. at high frequency) the eddies become small and their energy is dissipated via viscosity. Kolmogorov showed that by assuming isotropy, energy in the region between these two limits (known as the 'inertial subrange') follows this power law.

The limiting values of the 'inertial subrange' are a bit difficult to quantify as they will differ for different problems – depending on (among other things) length and time scales, the turbulent Reynolds number. However,

for many oceanographic problems, the inertial subrange lies (very generally, approximately, broadly, loosely) between about one cycle per day and one cycle per second.



So how did we arrive at these numbers? In the open ocean mesoscale eddies might have length scales of the order 200 km. Kolmogorov suggested the length scale separating the eddy producing region and the inertial subrange is, approximately, the eddy length scale divided by 6 i.e. 200/6=33 km. Assuming a speed scale of 1 m/s, we can estimate the time scale that separates these two regions is about 35,000 s or about 2.5 cycles per day (of the order of 1 cycle per day – the lower frequency limit of the inertial subrange). Kolmogorov also suggested the time scale of $\left(\frac{\nu}{\varepsilon}\right)^{1/2}$ separates the inertial subrange from the viscous dissipation region. Assuming $\nu = \varepsilon = 10^{-6}$ the period is about 1 sec (or a frequency of 1 cycle per second – the upper frequency limit of the inertial subrange).

2.6. Summary

Above, we have shown how Fourier theory is applied to a data set comprising discreet sampling in time (or space). Fourier analysis converts a time series, where the amplitude of a signal is a function of time, to a spectrum (or periodogram) where the energy is given as a function of frequency.

We've provided a broad indication of what to expect in a spectrum. If the energy spectrum is relatively flat and the phase spectrum contains no discernible pattern, then the signal is random. A dominant signal in the spectrum will appear as a spectral peak, the size of which gives the relative importance of that signal compared with other signals in the data. In the inertial subrange, the slope of the spectrum approximately follows the '-5/3^{rds} power law'.

Chapter 3. Preparing the Data

Real data sets are rarely perfectly suitable for spectral analyses without some form of pre-processing. Usually, we need to prepare the data before we can subject it to a spectral analysis. This chapter will discuss:

- Collecting data – experimental design
- Filling data gaps
- Filtering the data
- Truncating or extrapolating data sets
- Aliasing and limiting frequencies

After obtaining our data, the first step is to plot the data. This will help identify outliers, gaps and problems in the data and whether there are any trends or regular patterns that we can identify and remove. Before we can visually examine the data we first need to collect the data and this means designing a data collection program – an experimental design.

3.1. Experimental design

Data are generally available in one of two forms: primary data that we collect for our specific purpose and secondary data that have been collected by others.

(a) Primary data collection. We design and execute an experiment to collect the data we require. We have total control over the variables that we measure, the location of the measurements, the time (or space) intervals and the total time (or space) over which the data are collected. However, primary data are relatively expensive to collect.

(b) Secondary data have already been collected by someone else. We have no control over what was collected, where, when and how often the data were sampled. The data may have been collected for a different purpose which may, or may not, suit our needs. If we use secondary data we may need to compromise. We might

not be able to undertake the analysis that we originally planned or answer the exact question that we have posed. Data may have been collected from a sub-optimal location (from our perspective); the time step or total duration or time of year may not be ideal for our purpose, different variables may have been measured (e.g. total nitrogen data may have been collected although our interest may be in nitrate and nitrite). The data may be old and not now relevant. Nevertheless, these data are available and we may glean some information from them. At the very least, secondary data are useful in helping to design a primary data collection program. The gathering of secondary data is relatively inexpensive. As secondary data have already been collected, we may only need pay for a data extraction and processing fee which represents only a small fraction of the cost of collecting these data.

In designing a data collection program, our first consideration is to be very clear on the question that we want to answer. Secondly, we need to decide on how we want to analyse the data. Given that this manuscript concerns spectral analysis, we tacitly assume that this (or a closely related analysis) is how we want to analyse the data. It is emphasised that spectral analysis may not be the most suitable approach to answer some questions.

The question that we want to answer will largely govern the variables that we need to measure, the location from which we want to collect the data, the time (or space) step and the total duration (or spatial separation) of the data. In an oceanographic application, some of the questions we may need to consider are:

- Are we using the appropriate instruments to answer our question (e.g. accuracy, precision, resolution)?
- Will boundaries (lateral, bottom, surface) effect our results?
- Are we in an internal wave field? How will they impact our temperature, salinity or density data?
- Are sand-waves, mega-ripples or other bottom boundary features important?

- What will be the impact of other oceanographic processes on our results e.g. tides, waves, ocean currents, discharges from rivers etc?

Once we understand what the impact that these questions may have in answering our original question, we can design and execute the data collection exercise. Usually, we try to collect about 10% more data than we think are needed to answer our question. This also enables us to smooth the data without having to compromise the data processing (see Section 3.3).

3.2. Filling data gaps

Real data sets are often incomplete. Particularly, long data sets will almost certainly contain gaps. Such gaps can arise from: downtime due to servicing, instrument failure, battery failure, human error (e.g. incorrectly setting up the instruments), vandalism, marine growth on sensors and so on. If we're interested in high frequency features, we generally don't need a long data set and we can effectively ignore the missing data and analyse, separately, the data on either side of the gap. However, if we're interested in long period features then we need to fill in the missing data.

There are many ways in which gaps in the data can be filled; from linear interpolation to neural networks. The purpose of such data patching is simply to obtain a long, continuous data set that appears consistent throughout. The purpose of patching in our applications here is NOT to accurately fill the missing data, although it is a bonus if this occurs. Below, is described one technique that may be used to patch missing data[2].

[2] Knowing whether a data patch is accurate is not possible. By definition the data are missing hence there are no data against which the patched results can be compared. Yes, we can poke holes in a data set, fill them using a patching technique then determine how the patch compares with the actual data. This is a good test for the effectiveness of the data patching technique.

Firstly, what should the patching process look like? Below are some of the characteristics that we think a good patching process should have. The patching process should be:

- Simple, consistent across all gaps and easy to implement.
- Collapse to linear interpolation when there is only one data point missing.
- Applicable to data gaps of any size (in either time or space).
- Reflect the variability of the surrounding data.

Further, much of the data analysis undertaken in physical oceanography is based on spectral analysis (e.g. spectral analysis, harmonic analysis, wavelet analysis). Therefore, a gap-filling process that uses spectral techniques is desirable.

The following is a process that we often use to patch a data gap (Tate et. al., 2019).

Suppose there is a gap of size 'n' (time records) starting at time 't+1'. That is, the data gap extends from time 't+1' to time 't+n' inclusive.

- At time 't+1' (i.e. the first missing data) average the two data points at times 't-n+1' and 't+n+1'.
- At time 't+2' (i.e. the second missing data), average the two data points at times 't-n+2' and 't+n+2'.
- The gap at time 't+3' is the average of 't-n+3' and 't+n+3', and so on until the whole gap is filled.

This technique is equivalent to averaging the spectral estimates (at each frequency) of the 'n' data points before the start of the data gap with the 'n' data points after the end of the data gap, then applying an inverse spectral analysis to the averaged spectral estimates. It is a relatively simple patching technique and can be easily applied to a data gap of any size. Like many data patching processes, (a) it cannot produce an effective data patch near the start or the end of a data set, and (b) the larger the data gap, the less confident we will be in the patch.

Cubic splines at the start and end of the patched data can be used to ensure a smooth transition from the real data to (and from) the patched data.

Some examples of data patching using this technique are shown in Figure 3–1. A time series of water temperature over several hours is shown. Holes have been punched into a time series to enable a comparison of the patched and real data These data are represented by the black circles connected by the black line. Four sets of holes have been punched into these data and the gaps filled using the above technique. These are labelled 1 to 4 and identified by the arrows.

The first is a gap of size one. In this particular case the patch matches the data very well. The second is a gap of size two and the missing data happen to occur at the crest of a feature (or event). In this case, the patch does not identify the peak and the patch is poor. Gap three is five records long and the patch does a reasonable job in matching the missing data (although there does appear to be a time shift of one record). The final gap (number four) is 10 records long. The first three and last three data points in the gaps have been replaced by a cubic spline. The patch provides a fair estimate of the 'missing' data. Recall that the purpose of the patch is to obtain a continuous data set where the variability of the patched data reflects the variability of the surrounding real data.

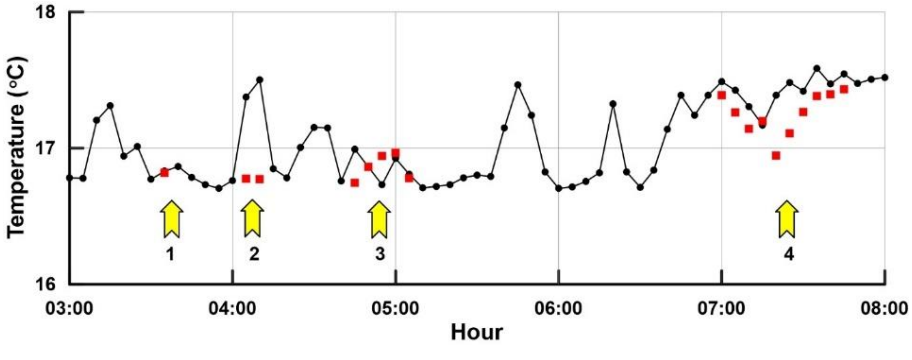


Figure 3–1. Examples of patching data gaps.

This approach enables data to be patched across a gap of any size up to one third of the length of the data set (although we would NOT recommend patching such a long data gap). The accuracy of the result depends on several factors, including: the size and location of the data gap and the variability of the surrounding data. The patching produces a continuous data set and enables the spectral analysis to be carried out over a long period of time (or space).

Outliers and other problems with the data may be difficult to handle and their identification is usually subjective. There is a temptation to remove any data that 'looks odd', or is substantially different to the surrounding data. However, we may be removing a genuine data point(s). If we're using a raw data set (i.e. a data set that has not been smoothed or manipulated in any way) look for gradients up to, and down from, the point(s) in question. If such gradients exist, then it is likely that the data point(s) in question is real. Without such a gradient, the data point(s) may be an outlier. Treat outliers in exactly the same way that you treat data gaps.

3.3. Filtering and subsampling

Filtering is a process used the smooth the data or to reduce the number of data points to be analysed. A filter comprises a series of 'weights' or fractions of unity. The number of weights that are used is defined by the used, while the values of the weights depend on the characteristics of the filter being used. The sum of the weights must equal unity. If the sum of the weights exceeds unity, the total variance (or energy) in the data set will increase; if the sum of the weights is less than unity the total energy (or variance) in the data set will decrease. Weights are usually positive, but can be negative.

Many (perhaps most) filters are symmetric. They have an odd number of weights (say, $2w + 1$ weights). There is one central weight and the same number (w) of weights on either side of this central weight. In general, the central weight has the greatest numeric value, while the magnitudes of the weights decrease the farther we are from the central weight. Further,

weights that are the same distance from the central weight (on either side of the central weight) have the same numeric value.

The filter selected depends largely on the question that is being asked and personal preference. Filters can be generic or designed for a specific purpose – to remove features of known frequency. Several examples of filters are given below, many more filters are available (e.g. Emery and Thomson, 1998).

Filtering or smoothing can be done either in the time / space domain (i.e. before we undertake the spectral analysis) or in the frequency / wavenumber domain (after the spectral analysis). Which domain is used is largely a matter of personal preference. This section will mainly focus on filtering in the time domain. Filtering (or smoothing) in the frequency domain will be examined in the next chapter and is used in that chapter for a different purpose. One key point to remember is that the filtered data should remain representative of the original data. In other words, don't smooth so much that the basic information in the data set is lost. How much smoothing is the right amount, is a matter of experience.

There are three main reasons for filtering a time series.

- Firstly, data may be sampled at a frequency that is not warranted for our purpose. For example, if we're interested in tides (which have time scales of the order several hours and longer) then sampling at 1 minute intervals will not give more information than data sampled at, say, hourly intervals.
- Secondly, there is a limit to the size of FFT routines (the Excel limit is 4,096, the Fortran program we use has a limit of 131,072 and the Matlab version we use also has a limit of 131,072). Even using a FFT of size 131,072 limits us to about 90 days of 1 minute data. This number of days is insufficient to properly interpret cycles of about two weeks duration or longer. If we averaged the data into 60 minute bins, we could then use the Excel FFT routine over 170 days, sufficient to effectively interpret a feature of two weeks duration.

- Thirdly, we may want to remove known signals (or signals in which we're not interested) prior to undertaking the FFT.

One outcome of filtering is to smooth the data – its variability is reduced, it 'chops off the peaks in the data and fills in the troughs'. Therefore, the spectral energy will also be reduced. This is not a problem *per se*, but if we're comparing the spectral results of different data sets, they should be filtered in the same manner.

The results of filtering depend on two main features: the type of filter being used (or the weights applied to each data points) and the length of the filter. In general, longer filters are more effective (than short filters) at smoothing the data, however, they lose more data at either end of the time series. (This is one reason to collect more data than you think you'll need).

There are three types of filters:

- Low pass filters that are used to smooth (or remove) high frequency features from the data set.
- High pass filters that are used to smooth (or remove) low frequency features from the data set. One way to implement a high pass filter is to generate a low pass filter then remove the low pass filtered data from the original data set.
- Band pass filters. These are used to retain only those data that lie within a specific frequency band. For example, if we're interested in the annual signal in a data set, we could filter out those data that have a period longer than, say, 18 months and also filter out data that have a period shorter than, say, 6 months.

Filters can be broadly divided into two groups: generic and specific. Generic filters have no specific function except to smooth the data. These may include: simple box-car average, Gaussian averages, running averages, cosine Lanczos filters, Savitzky-Golay filters etc. In contrast, specific filters have been designed to smooth the data in a very particular manner. Examples include: annual filters (to remove the annual signal from a data set) and the so-called 'tide killing' filters (to remove features that repeat themselves at the tidal periods) such as Munk's filter, Godin's filter, Thompson's '120i913' filter to remove tidal and inertial periods.

Spectral analysis can also be used to filter the data by supressing peaks in the spectrum. Suppose we wish to remove the annual signal from a data set. We undertake the spectral analysis, identify the peak at the annual period (or frequency) then reduce the size of that peak to the energy levels of the surrounding data. (Note: do NOT set the energy of the peak to zero, as the subsequent inverse transform will likely have problems). The inverse transform can then be taken and the resulting time series will not contain information at the annual period.



In addition to filtering, we can simply subsample the data to the required frequency. For example, taking one in every 60 values allows us to convert 1 minute data to hourly data. However, the results may be affected depending on which of the 60 data points in each hour are selected for analysis.

If we smooth the data, the consecutive data points (at the original time interval) will not be independent. Suppose we want to smooth the data using a running average. That is, we apply the filter to data points $t_1 \; to \; t_N$, resulting in a smoothed centred on time $\frac{t_1+t_N}{2}$. The filter is then applied to data points $t_2 \; to \; t_{N+1}$, resulting in a smoothed centred on time $\frac{t_2+t_{N+1}}{2}$. The two resulting data points will be closely related because most of the same original data were used to calculate both 'averages'. To ensure filtering process does not result in serial dependency, the second set of sampling should take place using the data points $t_{N+1} \; to \; t_{2N}$ i.e. the smoothed data should only be used every N records. For example, suppose we have 5 minute data that we smooth using a running average of 12 consecutive records. We still have 5 minute data, but each smoothed data point is related to the surrounding data – they are not independent. We need to subsample at hourly (or longer) intervals to ensure the filtering process does not result in sequential dependence. Note: the underlying data my still be serially dependent.

Below are some examples of filtering. The chosen data set comprises water temperatures sampled every 5 minutes. The objective is to use these data to generate a set of 60-minute data using several techniques.

3.3.1. Subsampling

Subsampling is probably the simplest process by which we can obtain hourly data. We simply take every 12th data point to produce hourly values (Figure 3–2). The subsampling in Figure 3–2 provides a reasonable fit to the original 5 minute data in the first part of the record, but much high frequency information is lost by the subsampling during the second part of the record. Having said this, if we're interested in high frequency information, then we would not be filtering in the first place. Taking subsamples will not introduce any new serial dependence in the resulting data set.

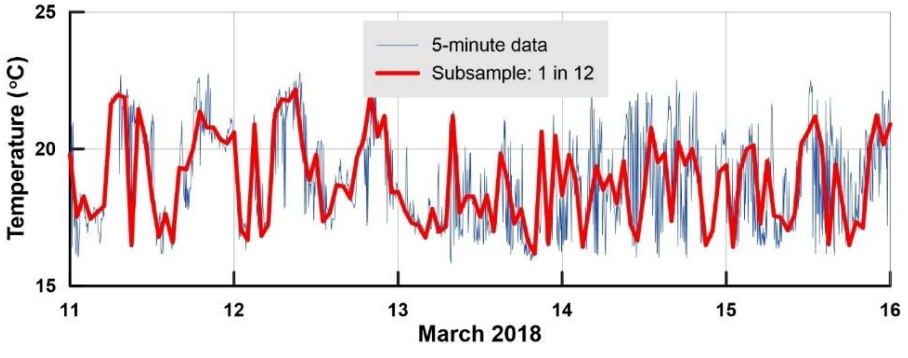


Figure 3–2. Subsampling – every 12th data point is selected.

In this example, the data selected were 'on-the-hour'. Selecting data from other times makes little difference on the overall result. Variances of the subsampled data (ranging between 3.31 to 3.53), selected from different times, were not statistically different from the variance of the 5 minute data ($\sigma^2 = 3.43$). Unlike other data smoothing techniques, subsampling may result in the variance of the subsampled data exceeding that of the original 5 minute data.

3.3.2. Box-car average

This is a simple average over n consecutive data points. Weights for the box-car average are all the same ($= \frac{1}{n}$ where n is the number of data points over which the average is undertaken). All data are given the same emphasis. The result is lower variability than the 5 minute data and that of the subsampling. The plot is noticeably smoother than that obtained from the subsampling (Figure 3–3). A 7-point average (i.e. $n = 7$) is shown. The larger the value of n, the smoother (and smaller variance) will become the resulting plot. In this example, the variance of the 7-point average ($\sigma^2 = 3.04$) is about 10% lower than that of 5 minute data ($\sigma^2 = 3.43$).

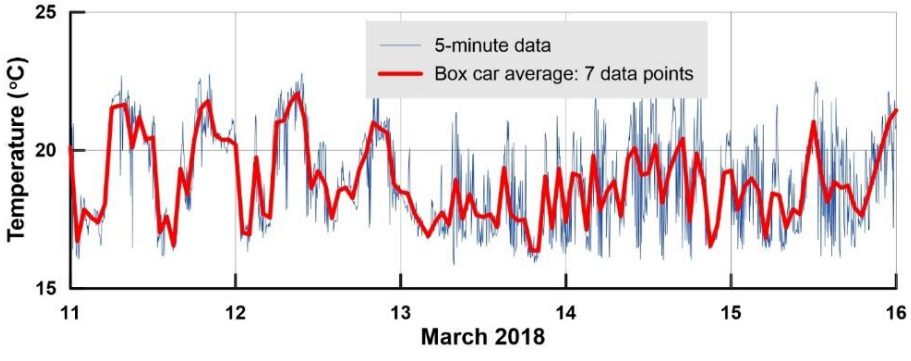


Figure 3–3. Box car average – 7 data points.

3.3.3. Gaussian average

The Gaussian average places less emphasis on data points that lie farther from the central point about which the average is taken. The weights are given by:

$$\binom{n}{k} = \frac{n!}{k!\,(n-k)!}$$

Where $n+1$ is the size of the average and k is the weight number. The example (Figure 3–4) uses a 7-point Gaussian average. Weights were generated using $n = 6$ and $k = 0\ to\ 6$ and are $(1, 6, 15, 20, 15, 6, 1)$. Each weight divided by 64 (the sum of all the weights).

The resulting plot (Figure 3–4) is similar to the box-car average (Figure 3–3) however, the peaks and troughs are better represented using the Gaussian average. The variance of the 7-point Gaussian average ($\sigma^2 = 3.12$) is closer to the variance of the 5 minute data ($\sigma^2 = 3.43$) than is the variance of the 7-point Box-car average ($\sigma^2 = 3.04$).

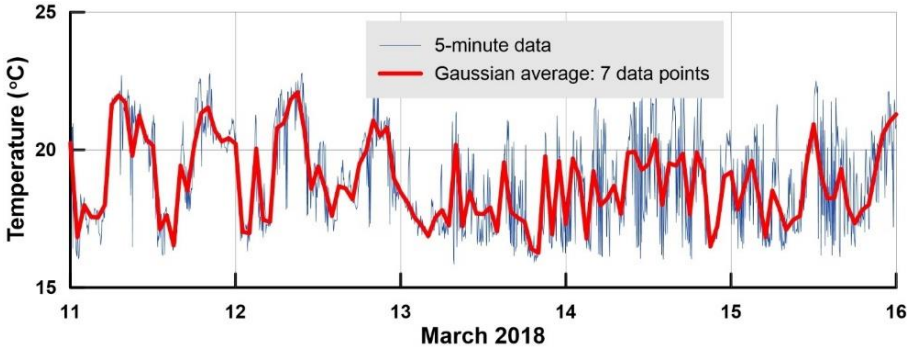


Figure 3–4. Gaussian average – 7 data points.

3.3.4. Savitzky-Golay average

The Savitzky-Golay class of filters are less commonly used filters (Press et al, 1992). One big advantage in using the Savitzky-Golay filters, is that they provide a good fit to peaks of different width and height. This occurs because the filter window is fitted with a high order polynomial – rather than the average. The filter in this section uses a 4^{th} order polynomial with 16 weights on either side of the central value (Figure 3–5). This filter is referred to as a 4,16,16 Savitzky-Golay filter – the first number referring to the order of the polynomial used, while the second and third numbers refer to the number of weights in each wing of the filter. Other combinations of polynomial order and the number of weights can be used. The filter does not need to be symmetric. Using zero weights after the central data point, it is possible to use the filter to extrapolate future values.

The variance of the 'Savitzky-Golay 4,16,16' filter (Figure 3–5) is ($\sigma^2 = 3.03$), less than that of the 7-point Gaussian average ($\sigma^2 = 3.12$) and about the same as the 7-point Box-car average ($\sigma^2 = 3.04$).

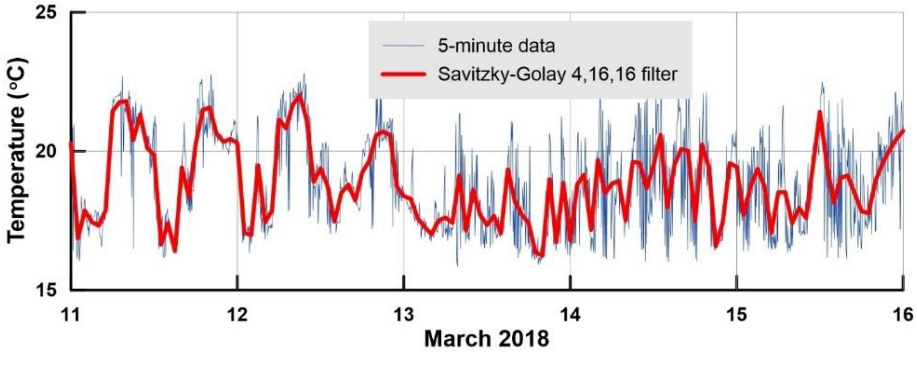


Figure 3–5. Savitzky-Golay filter – 4,16,16.

3.3.5. Cosine-Lanczos filter

The cosine-Lanczos filters are a set of filters based on the product of a cosine and a sinc function (i.e. a cosine function multiplied by a sine function that decreases in amplitude with distance from the central data point). They are a popular set of filters in the Fourier based time series analysis (especially in the oceanographic sciences).

In this example, a symmetric filter is used i.e. there are the same number of weights on each side of the central data point. The weights W_k are given by:

$$W_k = \frac{0.5\left[1+\cos\left(\frac{\pi k}{M}\right)\right]\frac{sin(\pi\, F\, k)}{\pi\, F\, k}}{\left[1+2\sum_{i=1}^{M-1} x_i\right]}$$

Where:

x_i are the data points inside the weight window,

$2M+1$ is the total number of weights (i.e. M weights on each side of the central data point),

F is the half amplitude response frequency (given in Nyquists[3]), and

k is a counter from 0 to $M+1$.

The example shown in Figure 3–6 uses 7 weights ($M=3$, i.e. 3 weights on either side of the central weight) and ($F=$) 0.5 Nyquists. The variance of the filtered data is $\sigma^2 = 3.19$. In general, the fit to the original data appears to agree slightly better than any of the above filters, particularly in days 13-15. This is reflected in the variance of this cosine-Lanczos filter ($\sigma^2 = 3.19$) which is closer to the variance of the 5 minute data ($\sigma^2 = 3.43$) than any of the above filters with an equivalent number of weights.

[3] The Nyquist frequency (or period) is the highest frequency (or shortest period) that can be resolved in the data set. Numerically, the Nyquist frequency is $1/2\Delta t$. (or Nyquist period is $2\Delta t$, twice the sampling interval). We expand on the Nyquist frequency in Section 3.7.

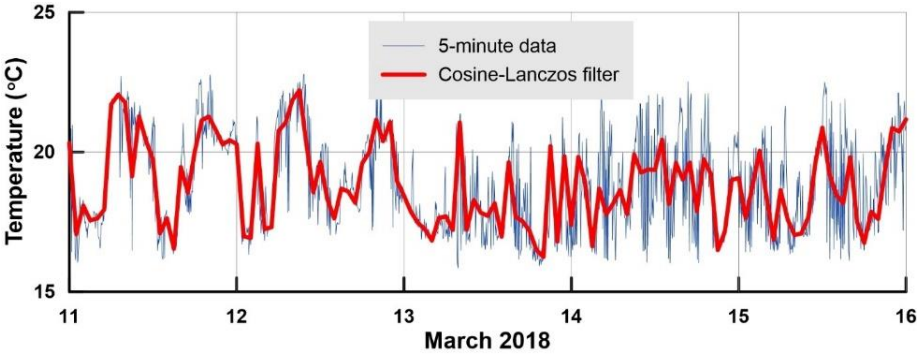


Figure 3–6. Cosine-Lanczos filter.

Overall, the above generic filters all provide similar hourly approximations to this 5 minute water temperature data set. They all provided a reasonable approximation when the data peaks were relatively broad, as observed in the first half of the record. The approximation is not as effective when high frequency variations occur in the data (e.g. days 13 and 14 of the data).

Below, some specific filters are examined.

3.3.6. Annual filters

Removal of the annual signal is undertaken in this section using a data set comprising 11 years of daily average water temperature data. The trend has been removed from these data and the resulting data plotted in Figure 3–7 (top plot). Superimposed on this plot is the annual plus semi-annual curve. (To remove an annual signal, it best to remove both the annual and semi-annual signals). The generic equation of the annual curve is:

$$x(t) = a + b\,sin(2\pi\,\omega_1\,t) + c\,cos(2\pi\,\omega_1\,t) + d\,sin(2\pi\,\omega_2\,t) + e\,cos(2\pi\,\omega_2\,t) + \varepsilon$$

Where:

a, b, c, d, e are coefficients to be determined,

$\omega_1 = \frac{1}{365.25}$ cycles per day, $\omega_2 = \frac{2}{365.25}$ cycles per day,

t time in days, and

ε is an error terms that is to be minimised in a least squares sense.

The equation can be solved using linear regression. The resulting daily temperature anomalies are shown in the bottom plot of Figure 3–7. This is a simple way to remove the annual signal.

An alternate approach to removing an annual signal uses the Fourier transform / inverse Fourier transform pair. The Fourier transform is undertaken on the data that has had the trend removed. The spectral estimates at the annual and semi-annual periods are supressed and the result is subject to an inverse Fourier transform. (When supressing unwanted signals, set the spectral energy at the relevant frequency, to a value similar to the spectral energy near the surrounding frequencies. Do not set the spectral energy to zero as the inverse Fourier transform usually struggles to resolve a signal with zero energy). The result is shown in Figure 3–8. The smoothed data are similar to that shown in Figure 3–7. The variances are slightly different: $\sigma^2 = 1.47$ using the trigonometric approach and $\sigma^2 = 1.58$ using the Fourier transform / inverse Fourier transform approach. The root-mean squared average difference between the two filtered data sets is 0.34. Results from the different approaches have little impact on subsequent data processing.

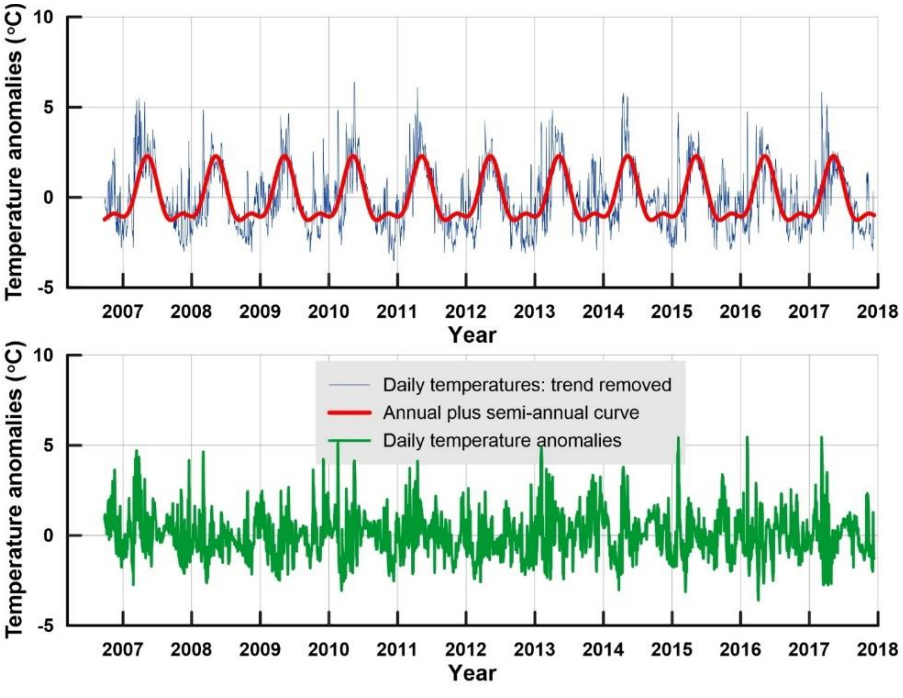


Figure 3–7. Removing annual signal – trigonometric approach.

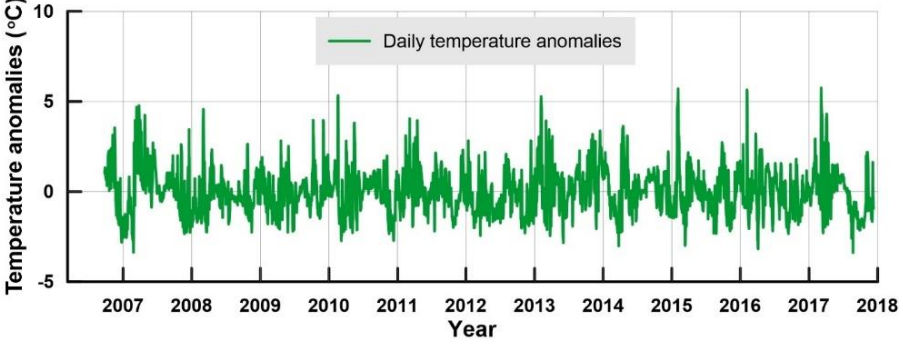


Figure 3–8. Removing annual signal – Fourier transform / inverse Fourier transform approach.

3.3.7. Tide killing filters

The final set of filters are specific filters designed to remove the tidal components from the data. A set of 170 days of hourly water level data are used to demonstrate the effect of these filters (Figure 3–9, top plot). Three common tide-killing filters are applied to the data: Munk's filter (Hamon, 1982), Godin's filter (Walters and Heston, 1982) and the 120i913 filter (Thompson, 1983). The three filters are of different length: 49, 71 and 210 hours, respectively. The longer the filter, the more data is lost at each end of the time series. Longer filters will, generally, provide a better representation of the low frequency component of the original data. Note that, for these three filters, the filtered data are serially independent when the temporal spacing between records is 49, 71 and 210 hours, respectively.

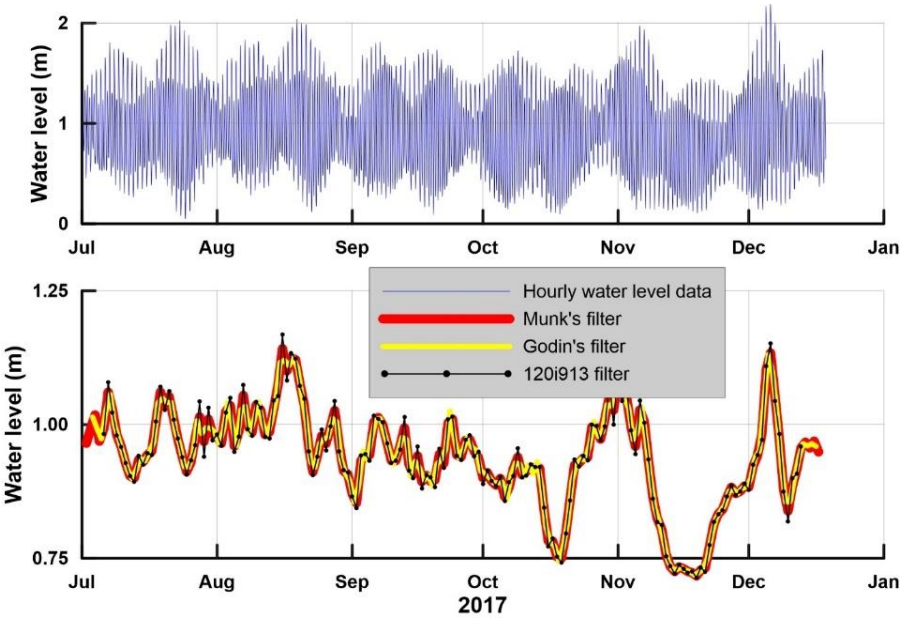


Figure 3–9. Tide killing filters: Munk, Godin, 120i913. Note that the vertical scales differ for the two plots.

The three filters shown in Figure 3–9 (bottom plot) all show similar features, although the 120i913 filter does have slightly greater variability

than the other two filters. All filters are effective in removing the tidal components from the data. In general, if a long time series is available, the 120i913 filter is preferable. If the duration of the time series is relatively short, then the shorter, Munk or Godin filters might be more appropriate.

3.4. Filter response functions

A filter response function can help with the design characteristics of a filter. Such a function identifies the frequency below which the filter has no effect on the data (i.e. the frequency below which 100% of the data pass unaffected) and the frequency above which the data are completely removed (i.e. the frequency above which 0% of the data are allowed to pass through the filter). The percentage of the signal that is passed by the filter is called the 'amplitude response'. The generally accepted reference is the 'half amplitude response' (or 'quarter power point') – that frequency at which the filter passes 50% of the amplitude of the original signal.

In this particular example (Figure 3–10), the filter response function applies to the cosine-Lanczos filter (defined in Section 3.3.5). Recall that the weights W_k for the cosine-Lanczos filter are given by:

$$W_k = \frac{0.5\left[1+\cos\left(\frac{\pi k}{M}\right)\right]\frac{sin(\pi\, F\, k)}{\pi\, F\, k}}{\left[1+2\sum_{i=1}^{M-1} x_i\right]}$$

Where the important parameters are:

M the number of weights in each wing of the filter, and

F the half amplitude response frequency (given in Nyquists).

Suppose we have a data set with 5-minute sampling (hence the Nyquist frequency is 6 cycles per hour) and we're only interested in low frequency features, features that have a frequency of less than 1 cycle per hour (cph). We set the half amplitude response frequency at 1 cph. Ideally, we'd like the filter response to look something like the black, dashed line in Figure 3–10. For frequencies less than 1 cph, the filter returns 100% of the original signal while for frequencies less than 1 cph, the filter returns 0%

of the signal. To achieve this 'perfect filter', we'd need an infinite number of filter weights. Not overly practical.

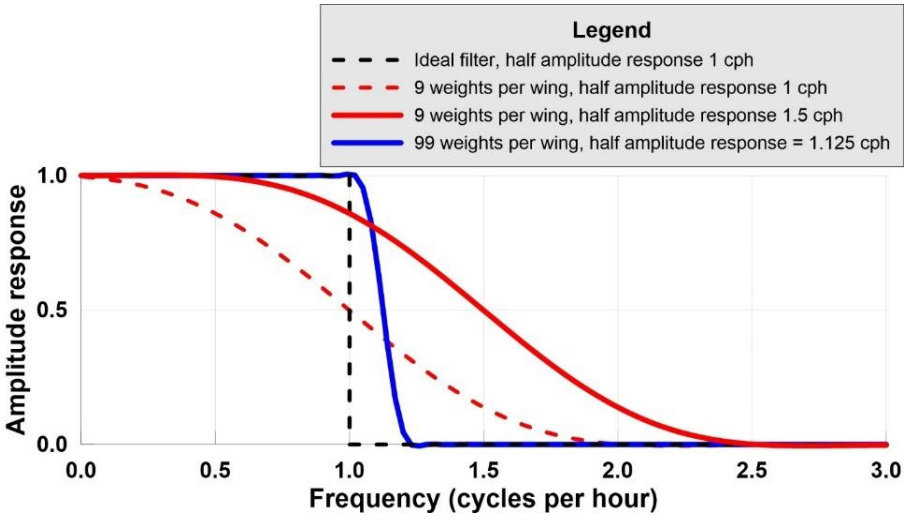


Figure 3–10. An example of the filter response function (a) black, dashed line – ideal filter and half amplitude response of 1 cph, (b) red, dashed line: 9 weights in each wing of the filter and half amplitude response of 1 cph (c) red solid line: 9 weights per wing and half amplitude response of 1.5 cph, and (d) blue line: 99 weights per wing and half amplitude response of 1.125 cph.

So, we need to compromise, and the compromise is to use a finite number of weights. If we choose $M = 9$ weights in each wing of the filter but retain the half amplitude response frequency of 1 cph (or $F = 1/6$ Nyquists), the result is the red, dashed curve in Figure 3–10.

Two problems are immediately obvious (a) there is much energy lost at frequencies in our region of interest, and (b) the slope of the curve is not particularly steep. The first problem can be solved by shifting the curve to the right (the red, solid line in Figure 3–10). To do this we chose a 'half amplitude response' frequency of, in this case, $F = 1.5$ cph. To steepen the slope of the curve we need to increase the number of weights in each wing of the filter. The blue curve in Figure 3–10 represents the filter response with $M = 99$ weights in each wing and a 'half amplitude response' frequency of $F = 1.125$ cph. The slope of this curve approaches that of the ideal filter, while the whole curve has been shifted slightly to

the right. The result is a filter that approximately meets our needs. Further trial-and-error can be used to align this curve closer to the ideal filter response. In applying this filter to the data, we lose 99 data points at each end of the data set.

3.5. An even number of data points

From the Fourier transform algorithm we note that the summation is over $\frac{N}{2} - 1$. If N is even, then $\frac{N}{2} - 1$ is an integer and the summation is easy to implement. If N is odd, then the set up of the algorithm is a little more complex. When N is odd we also find that there is no spectral estimate at the 'Nyquist' frequency (see Section 3.7). This is, perhaps, 'untidy' more than anything else, and it is certainly not a serious problem.

Our preference is to analyse the data using an even number of data points. If the time series contains an odd number of data points, just omit one data point (at either the start or end of the time series). If N is large, and for many (perhaps most) problems this is the case, then omitting one data point will not impact the results.

3.6. Truncate or extrapolate

If we plan to use one of the 'radix 2' FFT algorithms, there is a potential problem. (Recall that 'radix 2' uses a number of data points that is 2 raised to an integer power. This is the most efficient FFT that is available). Let's say we deploy an instrument for one month, collecting data hourly. This gives 720 data points, almost half way between 2^9 (=512) and 2^{10} (=1024). We're now left with a dilemma: do we lose $1/3^{rd}$ of our data points or do we add an additional $1/3^{rd}$ data points (and if so, how do we add in these additional data). Ultimately, we should have planned the experiment better; either by retrieving the instruments after 22 days and having slightly more than 512 data points or leaving the instruments in the field to collect data for 43 days and having slightly more than 1024 data points for the FFT. Which we choose depends on the question we wish to answer. If we're only interested in the high frequency end of the spectrum, then we can truncate the data set (i.e. lose $1/3^{rd}$ of our data, although we

might be better off looking at 3 x 256 point FFTs). If we're interested in the low frequency end of the spectrum, we want our data set to be as long as possible.

Truncating the time series is easy. We just need to decide on whether we use the first part, middle part of end part of the time series as input to the FFT. Adding additional data is trickier. A general rule that we use is not to add any more than about 5% to the length of the data set. So, in the above example, we'd have to add in about 30% of the data – which we think is too high to produce reliable results. We could analyse two sets of 512 data points with an overlap of about 200 data points or we could analyse three set of 256 data points with an overlap of 24 data points. Having three sets of FFT results would enable us to generate confidence limits (see Section 4.2). There is really not much else that we can do.

Let's say we have a data set comprising 1000 data points and we want to undertake an FFT using 1024 data points. All we need to do is add 24 data points to the data set. We could add these data to the start or end of the data set, but our preference is to add 12 data points to each end of the data set. We first remove the trend from the data set (Section 3.8) and add 12 'zeros' at each end. This process is known as 'padding' the data set.

3.6.1. Tapering

When zeros are added to the trend adjusted data, there is a jump from zero to the first (or last) data point. Spectral analysis does not handle jumps in data particularly well (more of this when we look at wavelet analysis). In the early days of spectral analysis, it was felt that a smooth transition from zero to the first data point would avoid any such jumps. Tapering is a gradual transition from the real data to the zeroes. Generally, tapering is often applied to the first and last 5% of the data, although this is not a universal percentage.

We are not fans of tapering, mainly because it changes valid data. It is important to note that the need for tapering is removed if the periodogram, rather than other techniques to estimate the spectrum, is used. Nevertheless, tapering is a process adopted by some and a brief description is given below.

Hourly water temperature anomaly data from Figure 2–5 are used to demonstrate the effect of tapering. The first and last 100 data points have been removed from the data in Figure 2–5 and replaced with zeros. The first and last 5% of the data have been subject to a linear taper (i.e. 195 data points at each end of the, non-zero, time series are tapered). The linear tapering formulae are (for this problem):

$$T_{new} = \frac{(n+1)}{195} T_{old} \quad for\ n = 0\ to\ 194 \quad starting\ at\ time\ t = 101$$

$$T_{new} = \frac{(195-n)}{195} T_{old} \quad for\ n = 0\ to\ 194 \quad starting\ at\ time\ t = 3802$$

Zeros were added to the data at times $t = 1\ to\ 100$ and at times $t = 3997\ to\ 4096$. Tapered values were added in at times $t = 101\ to\ 295$ and at times $t = 3802\ to\ 3996$.

For clarity, results for the tapering at the start of the time series only are shown (Figure 3–11). The black line indicates the original daily water temperatures anomalies. We've assumed here that we've had to add in 100 data points to each end of the time series to simulate padding or extrapolation of the data to bring the anomalies to the required 4,096 data points (the red line). The blue line includes both the 100 points of extrapolated data and linear tapering of the first 5% of the data (i.e. from data points 101 onwards). Close to the extrapolated section there is, in general, a large difference between the red and blue lines. This difference becomes less as we move farther from the extrapolated section.

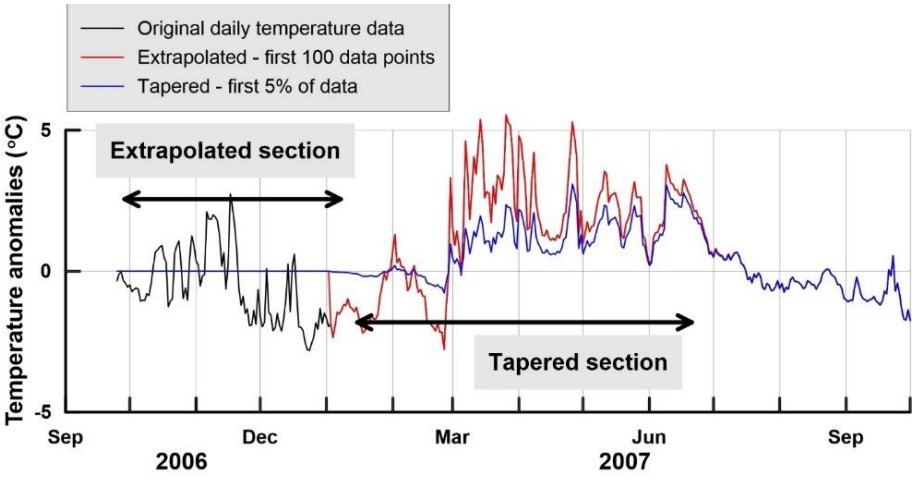


Figure 3–11. Example of tapering showing the extrapolated section (where zeros have been added) and the tapered section for the start of the time series.

So, how does tapering impact the FFT? The FFTs for the original temperature anomalies, the extrapolated data (adding 100 zeros to each end of the data set) and a 5% linear taper at each end of the data set are shown in Figure 3–12, by the black, red and blue lines respectively. The dominant (annual) feature of the spectrum remains unchanged for all three spectra as do their harmonics. The major differences among the three spectra occur at the low frequency end of the spectrum. However, as the linear trend has already been removed, the relative energy at the low frequency end of the spectrum is about two orders of magnitude less than the annual signal. For this example, the differences among the three spectra are negligible. This may not necessarily be the case for other combinations of extrapolation and tapering.

The tapering shown here is based on a linear taper. Other tapers are also used, a common one being the cosine taper. There is also a school of thought that uses tapering irrespective of whether the data are extrapolated. The idea is that spectral analysis assumes the same time series for all times, past and future. Within the time series under consideration, some frequencies will result in an integer number of sine /

cosine curves. However, most frequencies will not result in an integer number, and this leads to a jump or discontinuity at the ends of the time series. Tapering is one way to reduce the effect of this apparent discontinuity.

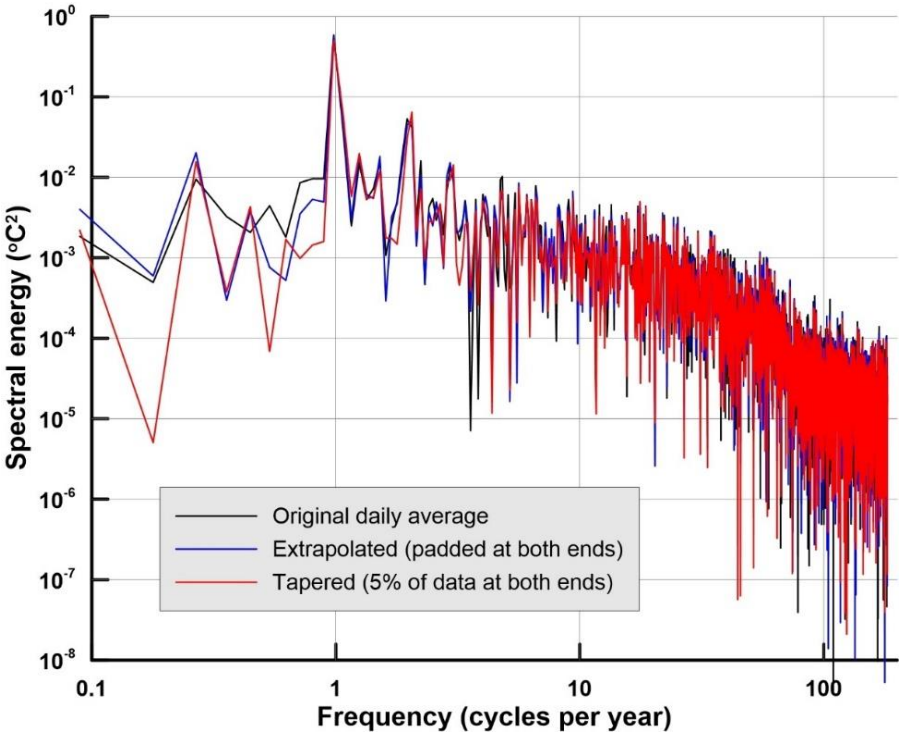


Figure 3–12. FFT of daily water temperature anomalies, extrapolated data and tapered data.

3.7. Aliasing and limiting frequencies

In analysing any data set it is important that we understand the limitations of the data. This applies equally to data that are collected as part of an experiment that we've designed and to data that were collected as part of a different experiment. We need to clearly define what we're seeking from the data and we need a broad understanding, *a priori*, of the processes of interest.

For example, let's assume that we're interested in the annual signal in a sea surface temperature data set. An example using data sampled monthly is shown in Figure 3–13. We have 120 data points and they clearly show the rise in fall of the sea surface temperature over time. There are 10 peaks and 10 troughs, as we might expect of our existing understanding of such data.

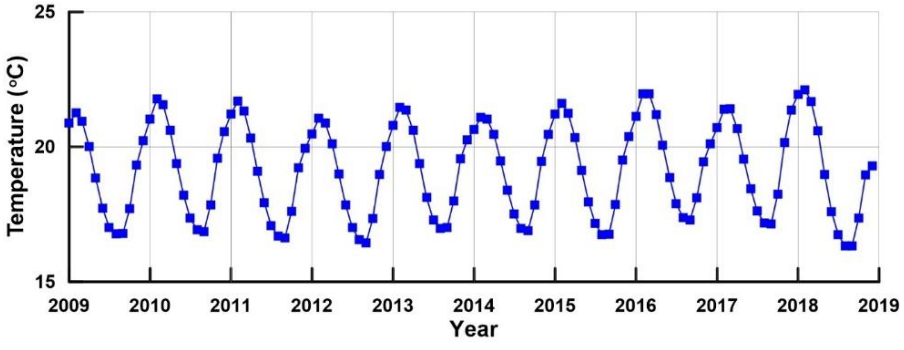


Figure 3–13. Monthly sea surface temperatures between 2009 and 2018.

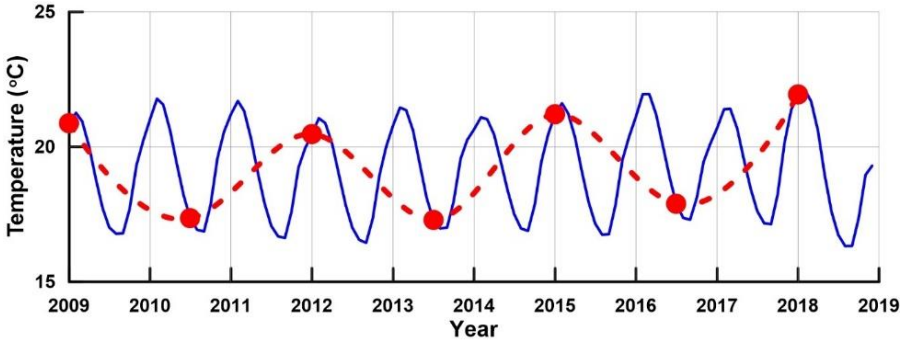


Figure 3–14. Monthly sea surface temperatures between 2009 and 2018 (solid blue line). Sampling sea surface temperatures every 18 months (red circles and the dashed line).

But what would happen if we didn't know that there was an annual signal in our data set? What would happen is we decided to sample, say, once

every 18 months? The result is shown in Figure 3–14 (red circles linked by the red dashed line). We still see a temperature range comparable with the original data. However, there are only 4 peaks and 3 troughs in the new data set. Clearly, it does not represent the annual signal. Monthly sampling clearly defined the annual sea surface temperature (Figure 3–13), but sampling every 18 months (Figure 3–14) did not shown the annual signal at all. So, how frequently should we sample to resolve (or effectively interpret) this annual signal?

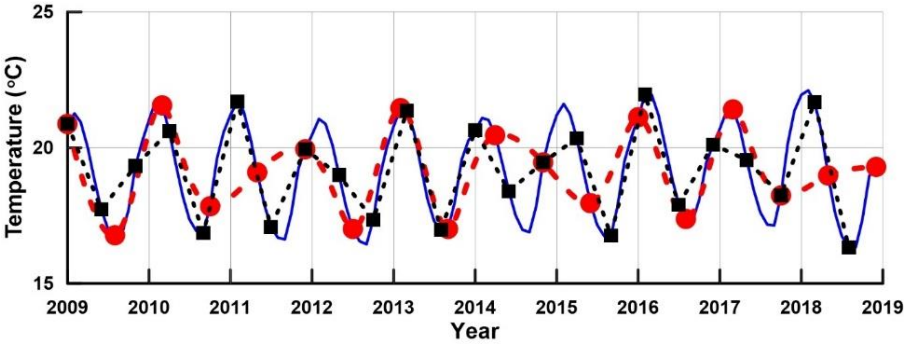


Figure 3–15. Monthly sea surface temperatures between 2009 and 2018 (solid blue line). Sampling sea surface temperatures every 7 months (red circles and the dashed line) and every 5 months (black squares and the dotted line).

Let's examine a few options. Sampling every 7 months (Figure 3–15, red circles and the dashed line) gives a better representation of the annual signal than at the 18-month sampling. However, it still only gives us 7 peaks and troughs – less than we require. Sampling every 5 months (Figure 3–15, black squares and the dotted line) gives us what we're looking for – 10 peaks and 10 troughs, although the annual signal is still not particularly well defined. So, it looks like we need to sample somewhere between every 5-7 months to resolve the annual signal. It actually turns out that we need to sample at least once every 6 months to resolve a 12 month signal. The more frequently we sample, the better we'll be able to represent the annual signal.

The 18-month and 7-month sampling periods represent 'aliasing' at the annual period. Generalising this, if we want to resolve a signal that has a frequency of $\frac{1}{T}$ cycles per unit time (i.e. is T time units long, or has a period of T units), then we must sample at a frequency of $\frac{1}{2\,T}$ cycles per unit time (or more frequently). This limiting frequency is known as the 'Nyquist frequency' – it is the highest frequency that can be resolved in a data set. If we're interested in what happens once per year, then we must sample at least twice every year. If we're interested in what happens once per second, then we must sample at least twice every second.

There is also a limit to what can be resolved at the low frequency end of the data set.

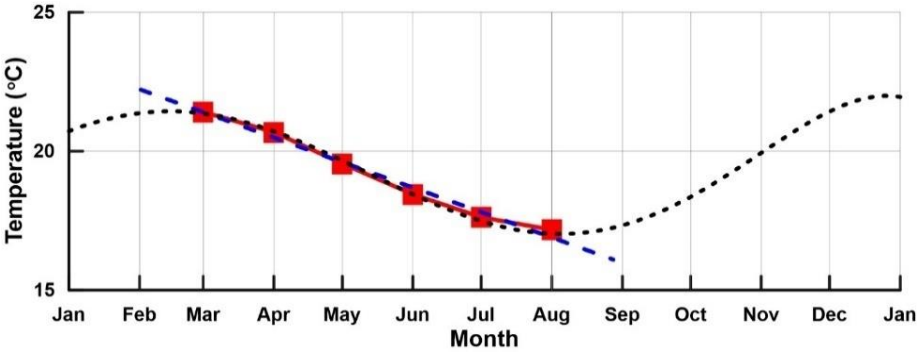


Figure 3–16. Monthly sea surface temperatures between March and August (red squares). Linear representation (blue dashed line) and an annual cyclic curve (black dotted line).

If we had, say, 6 months of sea surface temperature data as shown by the red squares in Figure 3–16, we could not explicitly determine whether the data are part of a trend (Figure 3–16, blue dashed line) or part of a cyclic signal (Figure 3–16, black dotted line). However, with 12 months of such data we can determine one way or the other. Therefore, if we have a feature that is N time steps long, then the lowest frequency that can be resolved is $\frac{1}{N}$ cycles per unit time. This is a problem that has appeared in 'climate change' discussions – is what we're seeing in the data a trend (i.e.

climate change) or just part of a very long period cycle (i.e. a natural process that is not yet known).

So, we need to have a broad understanding of what we want to see in our data set before either we design our own experiment or use data that someone else has collected.

Generally, if we're interested in high frequency processes, then we need to sample more frequently. If we're interested in low frequency processes then we need to sample over a long period of time. However, it is often better to be more conservative than this to ensure an effective interpretation of the data. Our general rules of thumb are:

(a) Sample at about 1/5 times the Nyquist frequency. For an annual sea surface temperature signal, the Nyquist frequency is one sample every 6 months. We would then sample at about 5 samples every 6 months (i.e. monthly sampling).
(b) Sample length should be (at least) 5 times the length of the process we're trying to resolve. However, this would depend on the strength of the signal that we're trying to resolve. Weak signals might require 10 (or more) times the length of the process we're trying to resolve. So, for the strong annual sea surface temperature signal, we would like at least 5 years of data.

So, to effectively interpret the annual signal in the sea surface temperature example of Figure 3–13 using spectral analysis, we would ideally like monthly (or more frequently) sampling over 5 years (or longer).

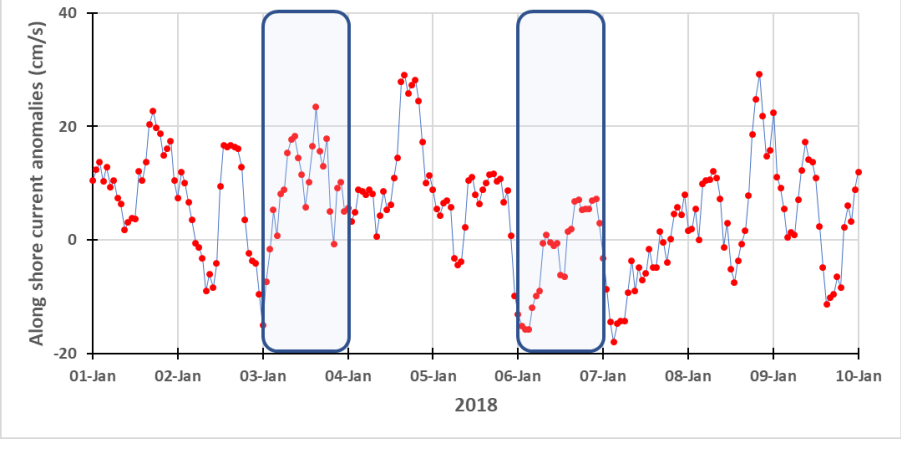
3.8. Stationary data sets

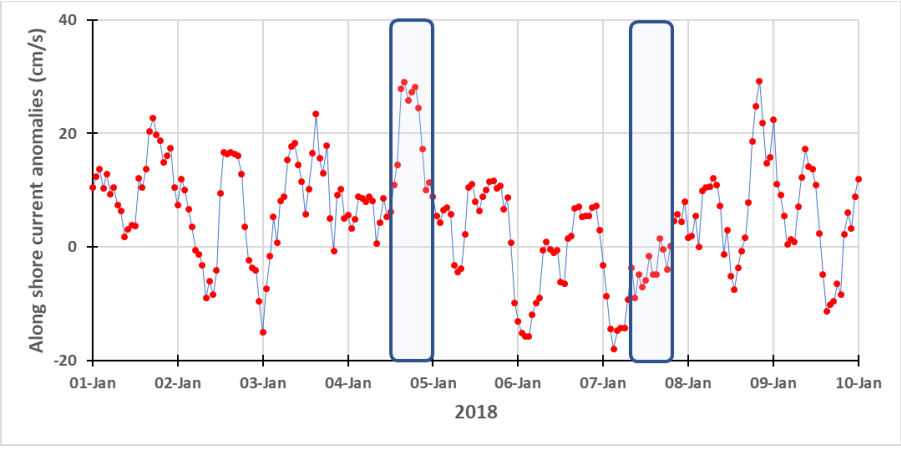
The Fourier representation of a signal forces an amplitude for the sine and cosine components to be constant throughout the whole data set. The implication is that a 'necessary' condition for spectral analysis is that the time series be stationary. Essentially, this means that, for any given subset of the data (each comprising 'm' number of data points), the statistics of those subsets should be the same. If we undertake a Fourier transform on a data set that is not stationary, we may compromise the

frequency information in the spectrum i.e. we assign spectral energy to a particular part of the data set that does not contain such spectral energy.

However, Zurbenko (1991) points out that '… *in practice this condition is possible only within some limited time interval or not possible at all*'. A strictly stationary process is one in which the '*moments and joint moments are time invariant*' (Bendat and Piersol, 2000). 'Moments' include statistical properties such as the mean, standard deviation, skewness, kurtosis etc. Strictly speaking, stationarity requires that those properties, calculated using different time scales, should remain unchanged.

Suppose we have a fixed width window containing a subset of the time series (figure). Stationarity basically says that the statistics of the data within that window should remain the same, irrespective of where that window occurs in the data set. This should be applicable to windows of any size.





For real data sets, this is unlikely to occur, hence this condition also needs to be relaxed – in general the mean and variance of different subsets of a real data set will be different, and statistically different at some level of significance. If the mean and variance of subsets of the data are approximately the same, then the data set may exhibit 'sufficient stationarity' to proceed with the spectral analysis. Such data sets are described as being 'weakly stationary'.

Many statistical tests are available to check that means are different (e.g. analysis of variance) and the variances are homogeneous (e.g. Bartlett's test or Levene's test) for groups of data within the time series.

Many of these tests actually fail, or the test assumptions (e.g. independence or normality) are violated. This is certainly our experience. Such tests are usually based on hypothesis testing. Often, when large numbers of samples are involved, small effects become statistically significant. The statistical power is very high and even a small departure from the expected value provides a statistically significant result.

Hypothesis testing also requires that the data be independent. In time series analysis, the data may not be serially independent and the use of any hypothesis test may be questionable.

Suppose we generate a time series using three frequencies: ω_1, ω_2 and ω_3 such that:

$$x(t) = a.\cos(2\pi\,\omega_1 t + \phi_1) + b.\cos(2\pi\,\omega_2 t + \phi_2) + c.\cos(2\pi\,\omega_3 t + \phi_3)$$

This time series is stationary because the three frequencies (ω_1, ω_2 and ω_3) are present in the data set, at all times.

However, in a real data set, this may not necessarily be the case.

Consider air or water temperature. This has a very strong annual signal. However, the 'annual' signal may not always be exactly 12 months. In one season, the maximum (Southern Hemisphere) temperature may occur in December, in February the next season and in January the following season. On average, the frequency is one cycle per 12 months, but this value may vary (albeit slightly) from year to year. Even the strong annual temperature signal may not be strictly stationary.

3.8.1. One approach to stationarity

This is based on the approach described in Warner (1998). Warner (1998) suggests that the number of data points in the time series should exceed 100, which is easily achieved in most oceanographic applications.

To test for stationarity, we select subsamples from our data set then test that the mean and variance in these subsamples could have come from the same population. There are a number of statistical tests available to undertake to achieve this, including Bartlett's test for homogeneity of variance (e.g. Kanji, 2001) and Analysis of Variance (ANOVA) to test for the homogeneity of the means (e.g. Sahai and Ageel, 2000).

The null hypothesis is that the variances and means have come from the same populations. The tests will be carried out using a set of randomly generated data, representing daily water temperatures and repeated using a set of real daily water temperatures from which the trend and annual signals have been removed. The 5% level of significance will be used; a commonly used percentage, but an arbitrary choice nonetheless. Assumptions are that the data are independent and are normally

distributed. Data can be tested for normality using a χ^2 goodness of fit test. In this case we've taken random samples from a normal distribution so we know that the population is normally distributed (the test statistic was 11.4 and the critical value was 25.0, hence we accept the null hypothesis that the data are normally distributed).

A time series of 4,096 random selection from a normal distribution (mean zero and variance unity) and shown as a time series in Figure 3–17. The slope of the trend line is zero (the red, dashed line in Figure 3–17). The data are divided into 16 groups of 256 data points each. Tests are carried out using the 16 groups. The null hypotheses are that (a) the variances are obtained from the same population and (b) the means are obtained from the same population. The null hypotheses are rejected if the test statistic exceeds a critical value.

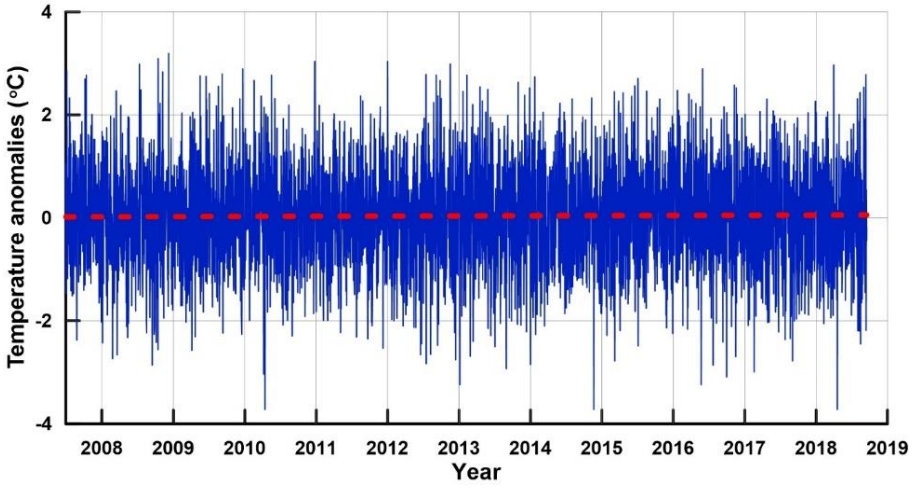


Figure 3–17. Random daily water temperatures.

Using 16 groups of 256 data points and Bartlett's test (Kanji, 2001), the test statistic is 14.6 and the critical χ^2 value ($\alpha = 0.05, \nu = 15$) is 25.0. Therefore, we accept the null hypothesis that the variances are all from the same population at the 5% level of significance. (Bartlett's test for homogeneity of variance is sensitive to departures from normality). If these same random samples are grouped into 8 groups of 512 data points

each, then the test statistic becomes 19.2 and the critical χ^2 value ($\alpha = 0.05, \nu = 7$) is 14.1. As the test statistic exceeds the critical value, we would reject the null hypothesis in favour of the alternate hypothesis that the variances are from different populations. This result may also suggest that the random number generator is Excel is not appropriate (unlikely for our purposes) or Bartlett's test is too sensitive (other tests are available that are less sensitive to departures from normality). Irrespective, we need to take care in how we interpret these results and our subsequent actions. In this case, we would be comfortable in assuming that the variances in the different groups resulted from data that were randomly selected from the same population.

ANOVA is a powerful tool for testing if sample means could have originated from the same population. Using the same 16 groups of 256 data points as above and ANOVA (Sahai and Ageel, 2000), the test statistic is 0.82 and the critical χ^2 value ($\alpha = 0.05, \nu_1 = 15, \nu_2 = 4080$) is 1.7. Therefore, we accept the null hypothesis that the means are all from the same population at the 5% level of significance. Repeating the analysis using 8 groups of 512 data points each, produces the same result, viz. test statistic is 0.72 and the critical χ^2 value ($\alpha = 0.05, \nu_1 = 7, \nu_2 = 4088$) is 1.8. In both cases the statistical power is relatively low (0.2-0.5) and there is a reasonable chance that a Type II statistical error will be made. Nevertheless, we are probably safe in saying that the data means are obtained from the same population.

Even though we know that the data are random samples from the same normal distribution, the above analysis places some doubt on the veracity of that claim. So, if we are in doubt when we have a synthetic data set that we are confident it should meet the criteria, a real data set will place even greater doubts on the homogeneity of variances and means. Let's examine this using a data set of daily water temperature anomalies.

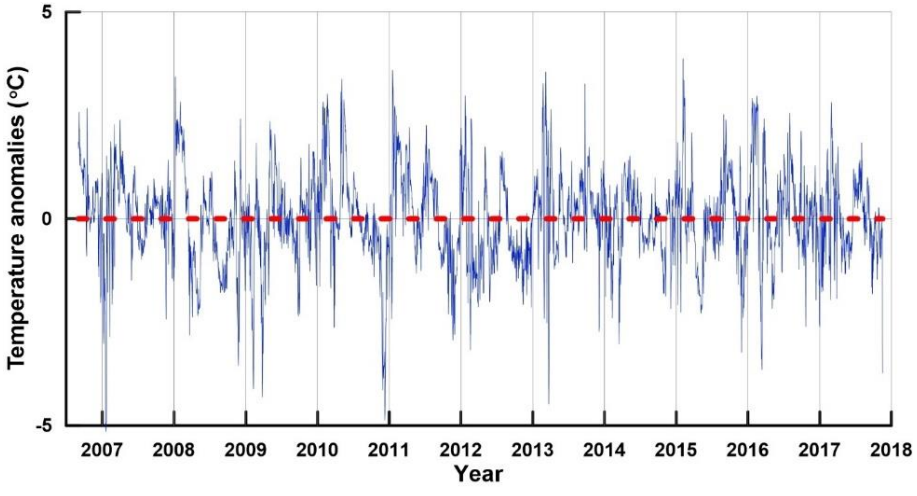


Figure 3–18. Daily temperatures anomalies. Linear trend and seasonal signal removed.

These analyses are repeated using the data set is that shown in Figure 2–6 – the trend and the known annual signal have been removed. We don't expect there to be any other (significant or important) signals in the data set. The data are plotted in Figure 3–18.

The Chi-square goodness of fit test was used to check whether the data follow a normal distribution. Results indicate that the data do not follow a normal distribution: the test statistic is 534, while the critical χ^2 value ($\alpha = 0.05, \nu = 15$) is 25.0. Other goodness of fit tests (the 'w/s' test, the Jarque-Bera test and D'Agostino's K-squared test) produced similar results. Therefore, we conclude that the data do not follow a normal distribution, despite the fact that the data 'look normally distributed' and, visually, are consistent with the expected values (Figure 3–19). Bartlett's homogeneity of variance test is sensitive to departures from normally and based on these results it is unlikely that the null hypothesis will be accepted. ANOVA is fairly insensitive to departures from normality and the means of the different groups may yet be representatives from the same populations.

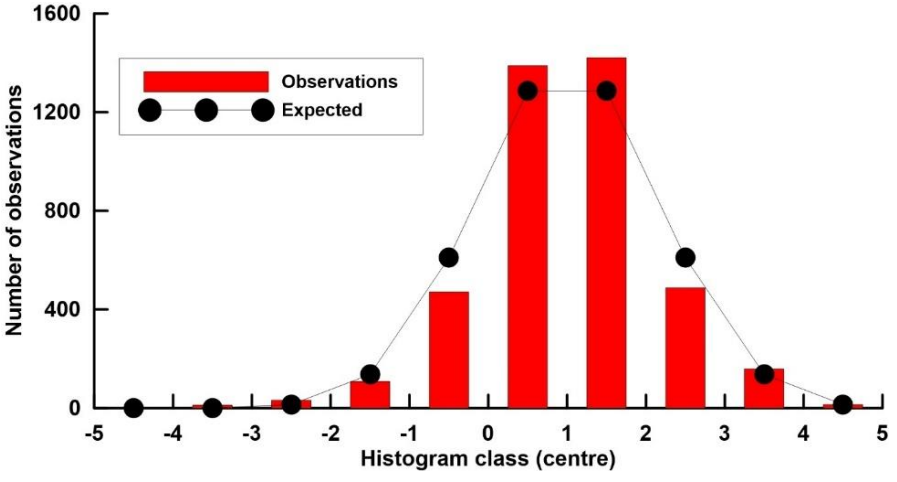


Figure 3–19. Histogram of observed data and line plot of expected data.

It is noted that we have 4,096 data points in this analysis. The Chi-square goodness of fit test does not perform particularly well when there are more than 'a few hundred' data points. A very large number of data points results in very large statistical power. This means that very small departures from normality will appear as significant (but may not be meaningful). The Chi-square goodness of fit test works best when we have between about 30 and 300 data points.

Proceeding, from Bartlett's test (Kanji, 2001), the test statistic is 392 and the critical χ^2 value ($\alpha = 0.05, \nu = 15$) is 25.0. As anticipated, we reject the null hypothesis that the variances are all from the same population at the 5% level of significance in favour of the alternate hypothesis that the variances are from different populations.

ANOVA (Sahai and Ageel, 2000) results in the test statistic being 15.1 and the critical χ^2 value ($\alpha = 0.05, \nu_1 = 15, \nu_2 = 4080$) is 1.7. Again, we reject the null hypothesis that the means are all from the same population at the 5% level of significance in favour of the alternate hypothesis that the means come from different populations.

The outcome is that the data are not even weakly stationary, despite the frequency histogram looking approximately normal.

This results in a dilemma: we either forgo the spectral analysis because the data are not (weakly) stationary, or we proceed with the spectral analysis based on the knowledge that the data look normally distributed and that the statistical tests used (for large sample sizes) are sensitive to small departures from expected values. Strictly, we should forgo the spectral analysis. However, in so doing, we will likely find that very few data sets would be suitable for spectral analysis.

3.8.2. Dickey-Fuller test for stationarity

An alternate approach is to use the Dickey-Fuller test for stationarity of a time series. The null hypothesis for this test is that the time series is not stationary. The alternative hypothesis is that the time series is stationary. The essence of the test is to regress the data with the first differences of the data and then determine the significance of the departures of the data from the regression model.

There are three common models used in the Dickey-Fuller test. In undertaking time series analysis, we usually remove the linear trend as part of the data pre-processing. Therefore, the Dickey-Fuller model we'll use is the 'constant, no trend' model, viz.:

$$\Delta t_i = \beta_0 + \beta_1 t_{i-1} + \varepsilon_i$$

Where;

t_{i-1} is the $i-1^{th}$ element in the time series,

Δt_i are the first differences in the data (i.e. $\Delta t_i = t_i - t_{i-1}$),

β_0, β_1 are the regression coefficients, and

ε_i is the error of the i^{th} element in the time series.

The mechanics of the test are relatively straight forward, although a little tedious to implement.

- Calculate the slope of the regression line between the data and the first differences of the data, β.
- Calculate the standard error estimate: $A = \sqrt{\frac{\sum_{i-1}^{n}(t_i - t_{i-regression})^2}{n-2}}$
- Calculate: $B = \sqrt{\sum_{i=1}^{n}(t_i - \langle t \rangle)^2}$
- Determine Tau calculated:

$$\tau_{calculated} = \beta \frac{B}{A}$$

If the calculated Tau value ($\tau_{calculated}$) is less than the relevant critical value in Table 3–1 (at the requested level of significance) then we reject the null hypothesis that the time series is not stationary in favour of the alternate hypothesis that the time series is stationary. In this case, $\tau_{calculated} = -17.34$ is less than $\tau_{critical} = -2.86$ (at the 5% level of significance) and we reject the null hypothesis and conclude that the time series is stationary.

Note that the null hypothesis for the Dickey-Fuller test is that the time series is not stationary. When the data set contains a large number of data points, the null hypothesis will usually be rejected. Hence the 'default' result from the Dickey-Fuller test for stationarity is that the data is stationary.

Table 3–1. Dickey-Fuller table of critical values for the 'constant, no trend' model.

	Constant, no trend model			
	Level of significance			
n	**0.01**	**0.025**	**0.05**	**0.1**
50	-3.568	-3.213	-2.921	-2.599
100	-3.498	-3.164	-2.891	-2.582
200	-3.457	-3.136	-2.873	-2.573
500	-3.443	-3.127	-2.867	-2.570
>500	-3.434	-3.120	-2.863	-2.568

Table is an extract from the augmented Dickey-Fuller table provided by: C.Zaiontz, Real Statistics using Excel (https://www.real-statistics.com/statistics-tables/augmented-dickey-fuller-table/).

3.8.3. KPSS test for stationarity

An alternative to the Dickey-Fuller test is the KPSS (Kwiatkowski, Phillips, Schmidt, Shin) test; Kwiatkowski et al (1992). The reason that it is mentioned here is because it is complementary to the Dickey-Fuller test in that the null hypothesis is: that the time series is stationary while the alternative hypothesis is: the time series is not stationary.

Again, the mechanics of the test are relatively straight forward. Calculate the following:

- Residuals at each time. Two models are commonly used, Model 1 removes the mean while Model 2 removes the linear trend. (We generally use Model 2).
- Cumulative residuals (R) at each time.
- Sum of the squares of the cumulative residuals (SSCR).
- Standard error (SE): $SE = \sqrt{\frac{\sum_{i=1}^{N} R^2}{N-2}}$ where N is the number of observations. If Model 1 is used then the denominator is $N - 1$ (not $N - 2$).
- The KPSS statistic is given by: $\frac{SSCR}{N^2(SE)^2}$.

For Model 2, the critical values are (for 95% confidence levels) = 0.146 and (for 99% confidence levels) = 0.216. If the KPSS statistic exceeds the critical value then we reject the null hypothesis and conclude that the time series is not stationary at the relevant level of confidence.

A significant result for the KPSS test and a non-significant result for the Dickey-Fuller test provides increased confidence that the time series in stationary (and the converse for non-stationarity). Clearly, a problem arises if results from the two tests indicate a different outcome for stationarity. This has prompted some practitioners to simply remove the linear trend (and possibly the annual or other known signals) from the data and assume that this is sufficient to approximate stationarity. It is a pragmatic approach that is often used for spectral analysis.

The above tests all involve hypothesis testing. When hypothesis testing is undertaken for many samples (as is often the case with time series analysis) the result is usually statistically significant. In other words, the Dickey-Fuller test will reject the null hypothesis that the data set in non-stationary in favour of the alternative hypothesis that the data set is stationary. The KPSS test will reject the null hypothesis that the data set is stationary in favour of the alternative hypothesis that the data set is non-stationary. A contrary outcome.

Hence, for many of our applications in which we have a large number of data points, none of the tests for stationarity are strictly applicable. Returning to the suggestion by Zurbenko (1991), stationarity in real data sets may not be achievable.

So, the whole concept of stationarity is vexing. On one hand we're saying that stationarity is a necessary condition for using spectral analysis, on the other hand, we're saying that the stationarity condition is rarely met and it has to be relaxed i.e. 'weakly stationary'. To be brutal, what is being suggested is to undertake a number of statistical tests then effectively ignore the results if they don't meet the test requirements. If we don't use such a pragmatic approach, we simply wouldn't use spectral analysis at all.

Our approach is to remove the linear trend (and other signals as dictated by the questions we ask) from the data. Alternatively, if the frequency histogram looks approximately normal, then proceed with the spectral analysis. Purists will likely object to this approach, but at the very least we will recognise the limitations and can allow for this in the description of the results.

3.9. Independence of the data

Statistical independence is critical for many applications. Formally, two events (or two data points) are independent if the joint probability of that event occurring is equal to the product of the probabilities of the individual events occurring. For example, using an unbiased coin, the probability of tossing two consecutive heads is 1-in-4. The probability of a head on the first toss is 1-in-2 and the probability of a head on the second toss is also 1-in-2. Their product is ¼ i.e. a probability of 1-in-4, the same as the probability of tossing two consecutive heads. The coin tosses are independent. Without statistical independence some statistical tests may be invalid and there is the risk that we will arrive at an incorrect conclusion from that test.

If we have a purely random process then the data are independent: the value of one data point does not depend on any other data points. However, for a real data set this is often not the case. Let's suppose we have a data set of comprising monthly water temperatures. Because of the strong annual signal, data points that are 12 months apart will be strongly correlated. They are not independent – the water temperature in, say, January of a particular year is strongly related to the water temperature in January of any other year (at least in a statistical sense).

So, do we need the data to be independent to undertake a spectral analysis and correctly interpret the output?

(a) The 'yes' case. Ideally, before we undertake a spectral analysis, we need to remove the trend and any known signals. Spectral analysis should be undertaken only on a set of random data. Further, spectral analysis requires the data to be stationary (see Section 3.8). The statistical tests often used for stationarity (e.g. homogeneity of variance and analysis of variance or t-test) require the data to be independent.

(b) The 'no' case. It is extremely unlikely that we can identify and remove all of the signals from a time series. Often, it is these very signals that are of most interest e.g. what is the relative contribution of the annual signal to the data (how important is the

annual signal). Stationarity can rarely be achieved in real data sets (Zurbenko, 1991).

Therefore, in theory, the data should be independent. In practice, it is unlikely that serial independence will be achieved. So, any requirement for data independence in spectral analysis probably needs to be relaxed.

3.10. Summary

This chapter focused on what needs to be done to the data prior to undertaking a spectral analysis. The main points are summarised below.

- Think carefully about the suitability of existing data or your experimental design before collecting the data. Consider the type of variables you need to answer your question, the sampling interval and the duration of the experiment. Think also about how to best analyse the data to answer your question.
- Data often have gaps (or bad data), particularly if long data sets are being examined. Such gaps / bad data need to be identified and suitably patched.
- Often, we need to smooth the data before we undertake a spectral analysis. There are many different types of filters available, the choice of which is largely personal. Choose those filters that best fit the problem.
- There are advantages in using an even number of data points in a spectral analysis. If an odd number of data points exist simply remove the first (or last) data point from the time series.
- Many spectral analysis programs use the FFT, the most popular of which is a 'radix 2' function i.e. the number of data points in the FFT is 2^n where n is an integer. If we're short of a particular value of 2^n we can pad in additional 'data'. If we have more than 2^n data points we can remove the extraneous data to obtain exactly 2^n data points. If the former is needed, we can taper some of the data (say, 5% at each end) to ensure a smooth transition from the added 'data' and the observations.

- There is a limit to the resolution of features in a data set. The highest frequency that can be resolved is known as the Nyquist frequency (or Nyquist period) and it is: $\frac{1}{2\,\Delta t}$ cycles per unit time (or $2\Delta t$ time units for the Nyquist period). Our preference is to sample at a much smaller frequency, about $\frac{1}{10\,\Delta t}$ time units to effectively interpret features at a period of Δt. At the low frequency end, the limiting frequency is given by the length of the data set. However, in practical terms we should have about 5-10 cycles to effectively interpret the lowest frequency feature. Therefore, the lowest frequency feature that can be practically interpreted in a time series of length N is about: $\frac{5}{N\,\Delta t}$.

Chapter 4. The Auto Spectrum

The previous chapter examined some of the procedures that need to be undertaken to prepare the data for a spectral analysis. This chapter focuses on the mechanics associated with the spectrum (or periodogram) of a single time series and some of the features that the spectrum may exhibit. Examples will be taken from water temperature and water level data. A simple flow chart of the auto spectrum is shown in Figure 4–1. (This flowchart will become increasingly complex as we examine more detailed spectra). Data from a single set of values as a function of time (or space) is transformed and the outcome is the amplitude and phase lag as a function of frequency (viz. the periodogram or spectrum).

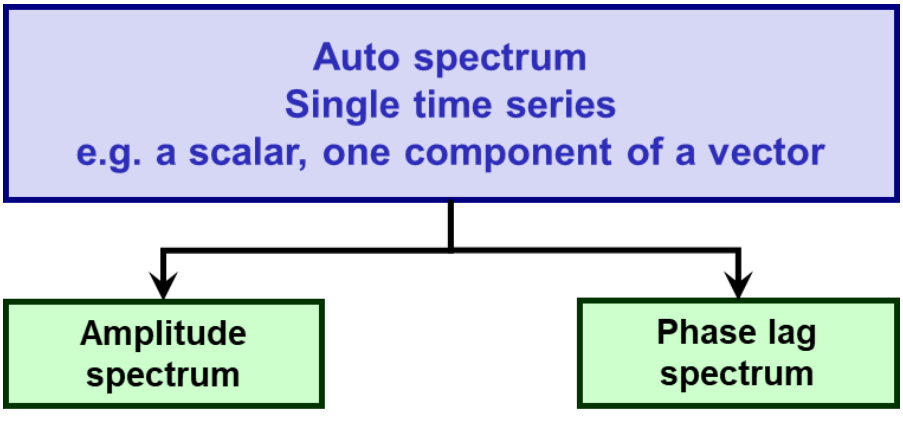


Figure 4–1. Flow chart of the auto spectrum.

The primary focus of this chapter lies in the following:

- Two important aspects of Fourier analysis – stationarity and independence.
- Generating confidence limits for the spectral energy and phase lag.
- Several examples showing results from different time scales and highlighting different features in the spectra.

4.1. Removing trends, known signals, outliers

In theory (to achieve stationarity), we need to remove the trend, any known signals and outliers from the data set prior to undertaking a spectral analysis.

- Removing the trend should always be done.
- Removing known signals will depend on the question being asked. If we're interested in finding 'hidden' features in the data set, then yes, remove as many known signals as can be identified. However, if we're interested in the relative importance of different signals in the data set, then we don't remove the known signals.
- Outliers are tricky – mainly in deciding whether a data point is a genuine outlier. If outliers are of primary interest, then perhaps spectral analysis is not the most appropriate analytical tool to use.

4.2. Confidence limits

The periodogram produces a point estimate of the spectral energy (or spectral density) at a particular frequency. Also, the periodogram is a biased estimator of the spectrum. By 'averaging' or smoothing in the frequency domain we can estimate the statistical confidence of the spectral energy in the averaged window and reduce the bias.

There are many forms of smoothing in the frequency domain – in much the same way that there are many forms of smoothing in the time domain (see Section 3.3). This is a process known as 'windowing', which was a subject of great interest with the popularisation of spectral analysis during the 1950s and 1960s.

Without going into the details, we are $100(1-\alpha)\%$ confident that the point value $s(f)$ of the spectrum at frequency f, (Chatfield, 1996) lies between:

$$\frac{\nu\,\hat{s}(f)}{\chi^2_{\nu,\alpha/2}} \qquad and \qquad \frac{\nu\,\hat{s}(f)}{\chi^2_{\nu,1-\alpha/2}}$$

That is:

$$\frac{\nu\,\hat{s}(f)}{\chi^2_{\nu,\alpha/2}} \;\leq\; s(f) \;\leq\; \frac{\nu\,\hat{s}(f)}{\chi^2_{\nu,1-\alpha/2}} \qquad or$$

$$\frac{\nu}{\chi^2_{\nu,\alpha/2}} \;\leq\; \frac{s(f)}{\hat{s}(f)} \;\leq\; \frac{\nu}{\chi^2_{\nu,1-\alpha/2}} \qquad or$$

$$log_{10}\left(\frac{\nu}{\chi^2_{\nu,\alpha/2}}\right) \leq\; log_{10}[s(f)] - log_{10}[\hat{s}(f)] \;\leq\; log_{10}\left(\frac{\nu}{\chi^2_{\nu,1-\alpha/2}}\right)$$

Where:

ν is the number of degrees of freedom,

$\hat{s}(f)$ is the 'average' value of the spectrum across the windowed part of the spectrum, and

χ^2 is the chi-squared distribution.

Using the first two forms of the confidence limits, above, means that the confidence intervals depend on the spectral estimates at different frequencies. However, if the spectrum is plotted on a log scale, then a single confidence interval applies across the whole spectrum. It is for this reason that the spectrum is often plotted on a log-log scale.

The final expression above (in the shaded box) uses the logs of the information. In so doing, we find that the argument between the two inequalities is the difference in the spectral estimate from some 'average' value i.e. it is effectively the size of a spectral peak, compared with its surrounding 'average'. If the size of the spectral peak lies outside the two limits above, then the spectral peak is important (or statistically significant).

Upper and lower confidence limits are produced for the energy spectrum. Spectral peaks whose height (compared with the nearby spectral estimates) exceeds the distance between these two confidence limits are regarded as statistically significant. In general, we'll use the 95% confidence limits, which we'll assume is equivalent to the 5% level of

significance. The choice of this value is arbitrary, although generally accepted.

The simplest form of smoothing is a simple box-car average. The equivalent averaging in the frequency domain this is called Daniell windowing. However, in contrast to the box-car average, the first and last Daniell weights are half the value of the other weights in the window. The Daniell weights are given by:

$$w_i = \frac{1}{m-1} \qquad for \quad 1 < i < m \qquad and$$

$$w_i = \frac{1}{2(m-1)} \qquad for \quad i = 1 \;\; and \;\; i = m$$

If we average over m number of spectral estimates, then $\nu = 2m$ and the confidence limits become:

$$log_{10}\left(\frac{2m}{\chi^2_{2m,\alpha/2}}\right) \;\leq\; log_{10}[s(f)] - log_{10}[\hat{s}(f)] \;\leq\; log_{10}\left(\frac{2m}{\chi^2_{2m,1-\alpha/2}}\right)$$

As the size of m increases, the 'distance' between upper and lower and lower confidence limits is reduced (although the size of spectral peaks is also reduced).

One difficulty lies in choosing a suitable value for m. If m is too small, the confidence limits will be large, but if m is too large then spectral peaks that otherwise may be important or of interest, will become small, not statistically significant or be eliminated. One option is to choose, say, three values of m, one small, one large and one in between, then select the value of m that produces a spectrum that 'looks best'.

Examples of smoothing using the Daniell windows and the corresponding 95% confidence limits are shown in Figure 4–2. Daily average water temperature over 11 years is used in this analysis (4,096 data points). A linear trend has been removed from the data. Three values of m are chosen: 4, 8 and 16 i.e. spectral estimates are averaged into bins of sizes 4, 8 and 16 frequency units wide. Assuming that the average spectral energy, $\hat{s}(f)$, is representative of the actual spectral energy, $s(f)$, the three

95% confidence limits apply across the spectrum and are given in the box at the top right hand corner of Figure 4–2. Spectral peaks that exceed the length of these confident limits (respectively) are regarded as statistically significant.

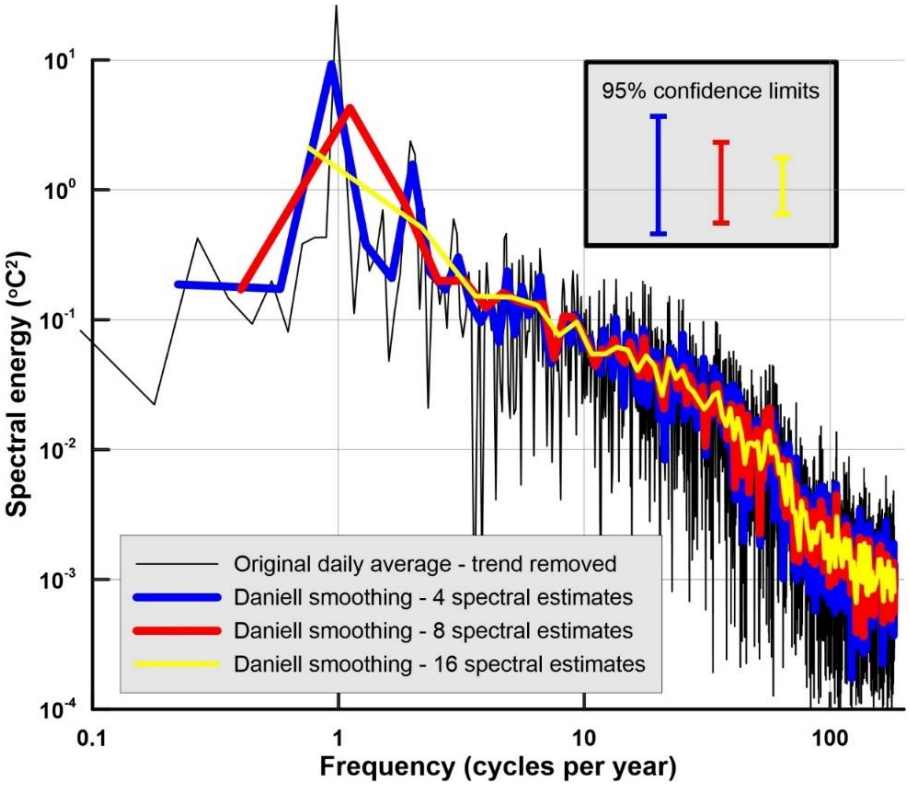


Figure 4–2. Spectral estimates with smoothing using three different values for *m* and the corresponding confidence limits for daily water temperature data – 4096 data points in the FFT.

The annual signal remains in the data. The FFT of the original data (black line) and the three smoothed spectra are shown in Figure 4–2 (blue line – average of 4 spectral estimates, red line – average of 8 spectral estimates and yellow line – average of 16 spectral estimates).

- Averaging over 4 spectral estimates results in a significant peak near 1 cycle per year and a second (not statistically significant) peak at 2 cycles per year. The averaging process also results in a slight frequency shift (towards low frequency) of the dominant peak.
- Averaging over 8 spectral estimates also results in a significant peak near the annual period, although the peak is relatively broad. However, the peak at 2 cycles per year has disappeared. We can also see that the dominant spectral peak has shifted towards the high frequency end of the spectrum. This suggests that averaging 8 spectral estimates in a time series of 11 years in length does not properly interpret the annual signal. To do so, a longer data set would be required.
- Averaging over 16 spectral estimates hides all spectral peaks – none are statistically significant. The annual signal appears as a linear trend. This averaging is too great to effectively interpret the annual signal. Again, a longer data set would be required to effectively interpret the annual signal if averaging of 16 spectral peaks is to be undertaken.

So, what happens when we do have a much longer data set? Let's look at 1,024 data points of monthly average water temperatures (about 85 years of data). Again, three sets of spectral smoothing are undertaken – spectral estimates are been averaged into bins of 4, 8 and 16 frequency units wide. The FFTs are shown in Figure 4–3. The obvious difference between Figure 4–2 and Figure 4–3 is the latter plots (over a much longer time period) shows very clear and large signal at the annual period. Also, the harmonics (at 2, 3, 4, 5 and 6 cycles per year) are also very clear and sufficiently large to show up as statistically significant peaks in the different spectra. The shift in the frequency of the annual signal (in the Figure 4–3 spectra) is small, compared with the shift observed in Figure 4–2. Therefore, with a longer data set, we can average over a greater number of spectral estimates and still resolve important features.

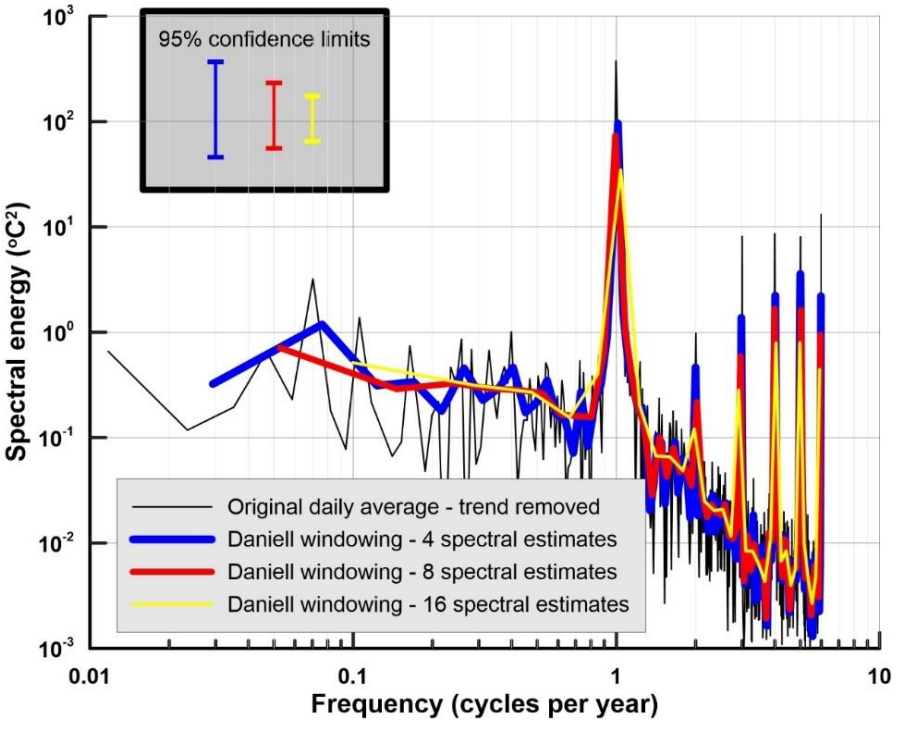


Figure 4–3. Spectral estimates with smoothing using three different values for m and the corresponding confidence limits for monthly water temperature data – 1024 data points in the FFT.

Different sections of the periodogram can be smoothed differently. An observation of the spectra (when plotted on log-log axes) is that there are fewer spectral estimates at low frequency than at high frequency. Therefore, an option is to smooth the low frequency end of the spectrum using only a few spectral estimates (i.e. use a relatively small $'m'$) and smooth the high frequency end of the spectrum over many spectral estimates (i.e. use a relatively large $'m'$). We'll demonstrate using the spectrum from Figure 4–2, which is reproduced in Figure 4–4 (blue line) together with the new smoothed spectrum (in red). At low frequency 2 spectral estimates are averaged, at mid frequencies 4 spectral estimates are averaged while at high frequency 8 spectral estimates are averaged. The upper and lower 95% confidence limits are indicated by the horizontal,

stepped black lines. Within each frequency band, (red) spectral peaks that exceed the distance between the two black lines are statistically significant. The only spectral peak whose size exceeds the distance between the two black lines occurs at the annual period (i.e. at a frequency of 1 cycle per year). The smaller peak at 2 cycles per year is a harmonic of the annual signal and its size does not exceed the distance between the two corresponding black lines.

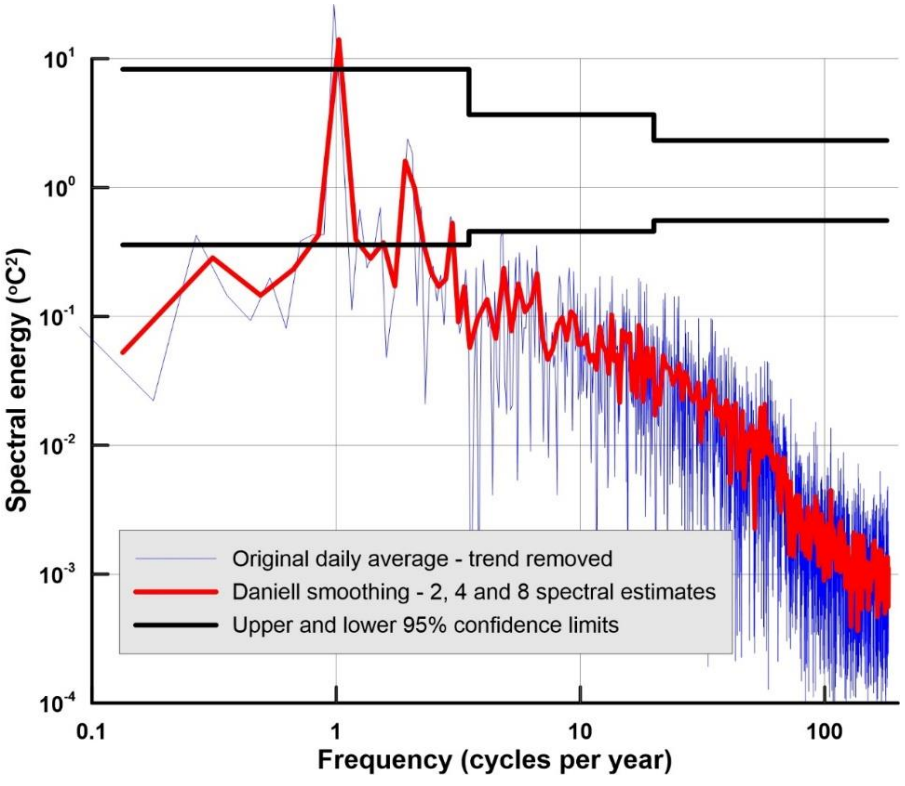


Figure 4–4. Spectral estimates with smoothing in three different regions of the periodogram. The values of *m* in each region are: 2, 4 and 8 – corresponding to low, medium and high frequency regions of the spectrum.

So, we can generate confidence limits for the spectrum by averaging spectral estimates. If we have a very long time series and we're interested

in (relatively) high frequencies, we can generate confidence limits in a slightly different manner.

For example, assume we have 12 years of hourly data and we're interested in what happens on scales of, say, 1 day to 1 week. An option is to split the time series into, say, 12 x 1 year data sets then undertake an FFT on 12 x 4,096 two-hourly data points. We then average the corresponding frequencies of the 12 FFTs. There will be $\nu = 12 - 1 = 11$ degrees of freedom and the upper and lower $100(1-\alpha)\%$ confidence limits are given by: $\frac{\nu}{\chi^2_{\nu,1-\alpha/2}}$ and $\frac{\nu}{\chi^2_{\nu,\alpha/2}}$.

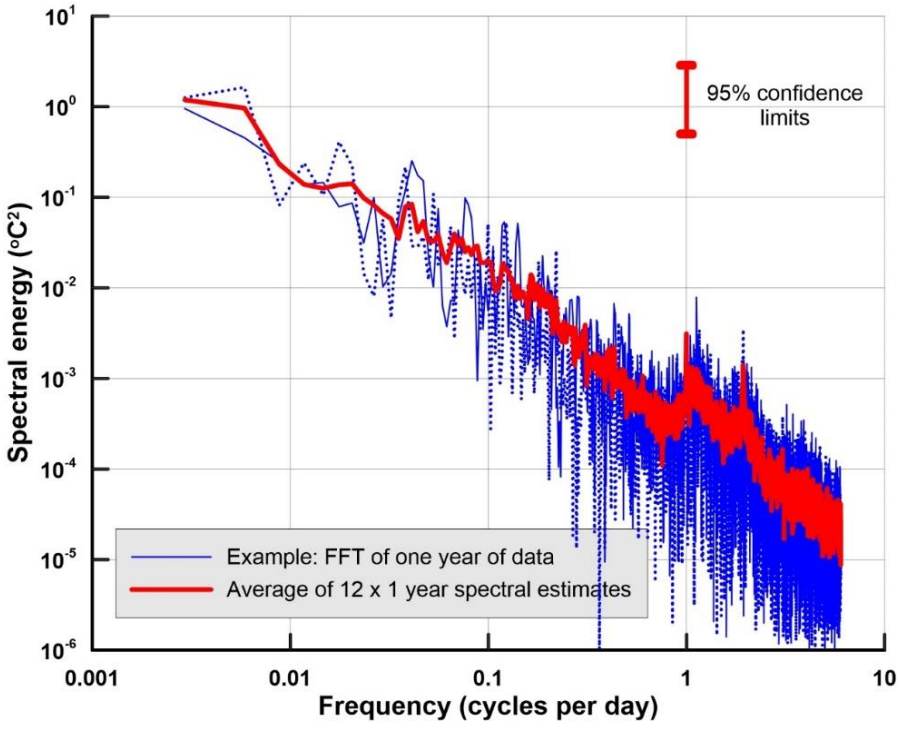


Figure 4–5. Spectral estimates using 12 x 2 hourly FFTs (one such plot is shown – the blue line). At each frequency, the 12 spectral estimates are averaged (thick red line). The 95% confidence limits apply over the whole periodogram.

An example is provided in Figure 4–5. Twelve FFTs were undertaken on 4,096 data points with $\Delta t = 2\ hours$. (Hence each periodogram covers almost one year). For clarity in Figure 4–5, only two years are shown (blue solid line and blue dotted line). The thick red line is the average of all 12 spectra. The 95% confidence limits (generated via the averaging process) are shown and apply across the whole spectrum. Peaks in the red spectrum occur at one and two cycles per day – the size of the former peak is about the same as the distance between the upper and lower confidence limits and may be statistically significant.

4.3. Examples of the auto spectrum

We're now in a position to examine the periodogram for a range of data sets. We'll primarily use water temperature (at various Δt scales) and water level. These examples are provided to highlight the features that we may come across when using spectral analysis. The more examples looked at, the more experience we get and (likely) the better will be our interpretation of the spectrum.

Each example starts with a question that we'll attempt to answer using spectral analysis. Initially, this question is often exploratory, designed to determine the broad features in the spectrum. When the broad features are known, more tailored questions can be asked. This usually means that running the spectral analysis multiple times.

Note that the questions we wish to answer should be clearly identified in the experimental design phase of the data collection process. Too often, data are collected and the questions are defined *post-hoc*, based on what an initial examination of the data allow us to address. Further questions may be raised, which may require additional data collection exercises or, if we're lucky, can also be answered with our original data set. Try to think of all the questions that you might like to answer in the design phase of your project and design your data collection exercise accordingly.

The question(s) is followed by the data used and any processing that is undertaken prior to the spectral analysis (e.g. trend removal, smoothing, removal of known signals). Testing is then undertaken to determine if the

data meet the stationarity condition. It should be made clear here that some conditions may not always be met. Despite this, we continue with the analysis, on the understanding that if we strictly adhere to these conditions, we probably wouldn't use spectral analysis at all. Finally, a description of the spectrum is provided.

Unless explicitly stated otherwise, all data are obtained from locations in the Southern Hemisphere.

4.3.1. Example 4-1: Water temperature – 5 min data

The first example uses a single data set comprising 14 days of 5 minute water temperatures. It has been kept relatively simple so that we can explore some of the main features of the auto spectrum.

Questions: Are there any significant peaks in the water temperature spectrum? If so, at what periods do these peaks occur?

Data used:

- Water temperature at 5 minute intervals, the FFT=4,096 points.
- The Nyquist period is 10 minutes (Nyquist frequency is 6 cycles per hour).
- The data set is about 14 days long, therefore the longest period features we can realistically interpret is about 2-3 days (i.e. we want about 5-10 cycles in the data set before we are confident of interpreting the low frequency end of the spectrum).
- Practically, the highest frequency we can confidently interpret is about 5-10 Nyquists, a period of about 60 minutes.
- So, the practical frequency range for this problem is about 1 cycle per hour and 1/3 cycles per day. (The theoretical values for these limits are 1/6 cycles per hour and 1/14 cycles per day).

Pre-processing:

- The linear trend is removed from the data to give temperature anomalies.
- No other smoothing or signal removal is undertaken.

The time series of anomalies is shown in Figure 4–6. The lagged correlation coefficient is plotted in Figure 4–7. This is the correlation coefficient of the data with itself when the data are lagged by 0, 1, 2 etc records. It provides a quick assessment of the presence of a repeating signal. In this example, the lagged correlation coefficient shows a minimum near a lag of 12 (i.e. a 60 minutes), which suggests that the data set is not random and that a signal exists near a frequency of 1 cycle per

hour. The frequency histogram is shown in Figure 4–8, and the spectrum (or periodogram) in Figure 4–9.

Superimposed on Figure 4–6 is a fourth order polynomial (red line) and this suggests that there may be a low frequency feature (with period of about 10 days) that may dominate the spectral energy. The physical process(es) that caused this feature is/are not obvious (they are neither diurnal nor tidal) and may form an area for subsequent research. As we don't know the cause of this feature, we don't have much justification to remove it from the data set. However, we do expect to see this feature dominate the spectral energy.

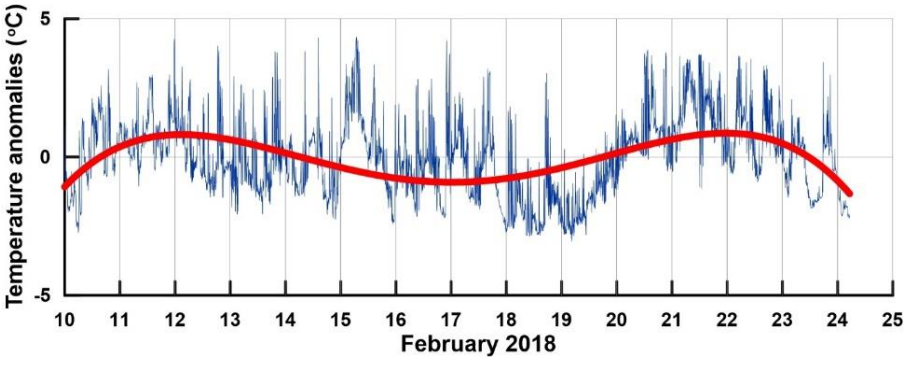


Figure 4–6. Time series of the 5-minute temperature anomalies.

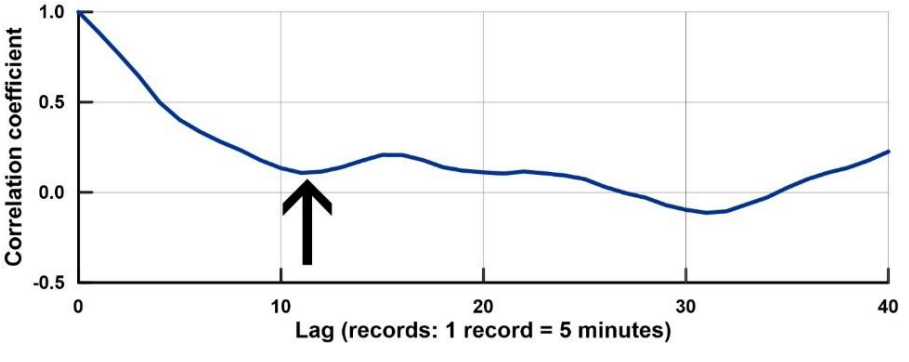


Figure 4–7. Lagged correlation coefficient. The arrow indicates the first local minimum (at lag=12 = 60 minutes).

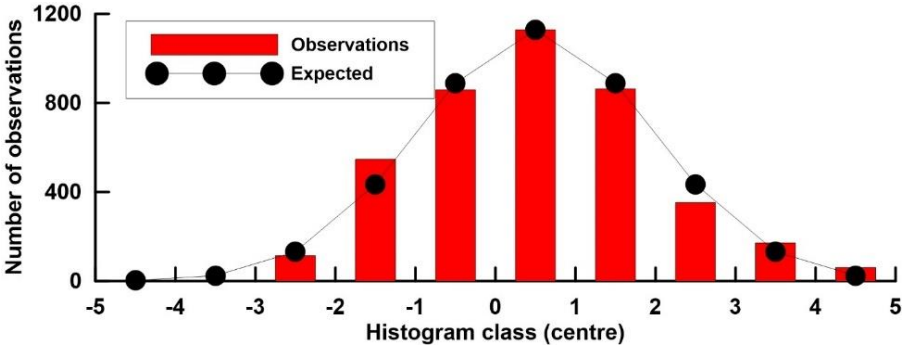


Figure 4–8. Frequency histogram of the 5-minute anomalies (red bars) and the expected number of observations in each histogram class (black circles).

Testing: The following statistical tests were undertaken. The 5% level of significance was used for all tests.

- Stationarity: The Dickey-Fuller test for stationarity results in a calculated Tau-statistic of -17.97, while the critical Tau value is -2.86. As the Tau statistic is less than the critical value for Tau, we can reject the null hypothesis that the data set is not stationary.
- Independence: Von Neumann's test for serial independence resulted in a t-calculated value of 54.63 and a t-critical value of 1.96 and it is concluded that the data are not serially independent at the 5% level.
- Normality: Based on the chi-squared goodness of fit test, the data are not normally distributed at the 5% level of significance. However, the frequency histogram (Figure 4–8) does 'look' normal in that it has a peak in the middle and tails off at each end of the distribution.
- Homogeneity of means and of variances: The anomalies were split into 16 groups of 256 data points each and subjected to Bartlett's test for homogeneity of variance and a 1-way ANOVA to compare the means. Neither test resulted in acceptance of the null hypothesis (i.e. that the variances and means could be from the same population) at the 5% level of significance.

The likely reason for the failure of these statistical tests is that, with 4,096 data points, the statistical power is very high and even small deviations will result in a rejection of the null hypothesis.

So, we have a dilemma – do we forgo the spectral analysis because we fail some of the statistical tests, or do we note this and continue? We note the exceptions and choose to proceed with the spectral analysis.

Smoothing of the periodogram:

- Daniell windowing is used to smooth the periodogram into frequency bands of size 2 (for frequencies lower than 10 cycles per day), 4 (for frequencies between 10 and 50 cycles per day) and 8 (for frequencies greater than 50 cycles per day).
- Smoothing of the periodogram via the Daniell windowing is used to estimate the 95% confidence limits.

In practice we'd probably also look at smoothing over a different set of bands e.g. 4, 8 and 16 bands and possibly a third set of bands, corresponding to low, medium and high frequencies. Then we'd chose the smoothing that best highlights those features important to the question we're asking.

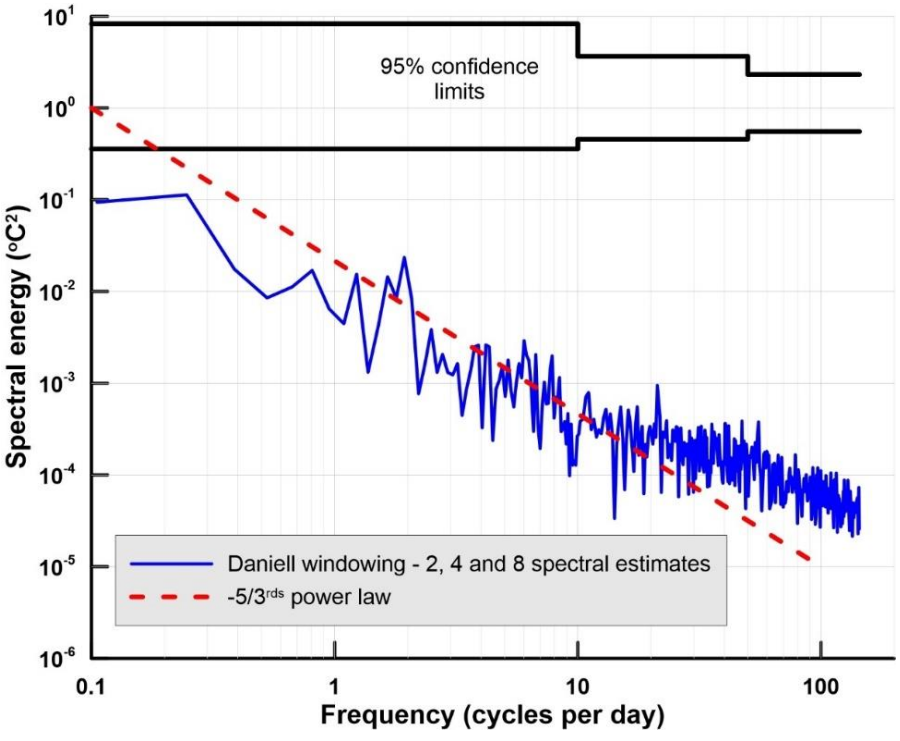


Figure 4–9. Periodogram of the 5-minute temperature anomalies. The 95% confidence limits (solid black lines) are shown together with the "-5/3rds power law" curve (dashed red line).

Description (Figure 4–9):

- The spectral energy near 1 cycle per day contributes less than 1% of the total spectral energy.
- Maximum spectral energy occurs at low frequency, reflecting the low frequency feature observed in the time series plot (Figure 4–6).
- There is a departure from the -5/3rds power law at high frequencies. This increase is spectral energy suggests an external energy source at frequencies higher than about 24 cycles per day. (This may be the result of internal waves, known to occur in this region of the ocean).

Finally, specifically addressing our questions (Figure 4–9:

(a) To answer our question, there are no spectral peaks of significance in the spectrum.
(b) There is one small spectral peak near 2 cycles per day, although the height of this spectral peak does not exceed the height of the 95% confidence limits.

So, what happens when we remove the fourth order (quadratic) polynomial, identified in Figure 4–6? The spectra with and without the quadratic polynomial removed are plotted in Figure 4–10. The same band averaging using the Daniell windows was used. The differences between the two spectra are small, with a slight lowering of the spectral energy at low frequency when the quadratic polynomial is removed. A slight phase shift can also be seen, probably as a result of the smoothing of the data set when the quadratic polynomial was removed. Low frequency still dominates the spectral energy. We could identify (and remove) additional low frequency components, but it is unclear how much we would gain from this exercise. Further, it is unlikely whether these low frequency components could be explained by any physical processes to justify their removal.

In terms of answering our original questions, the result would be unchanged from our Description above. The largest spectral peak (which is less than the 95% confidence limits) is near the semi-diurnal period (2 cycles per day) contributes less than 1% to the overall variance of the data.

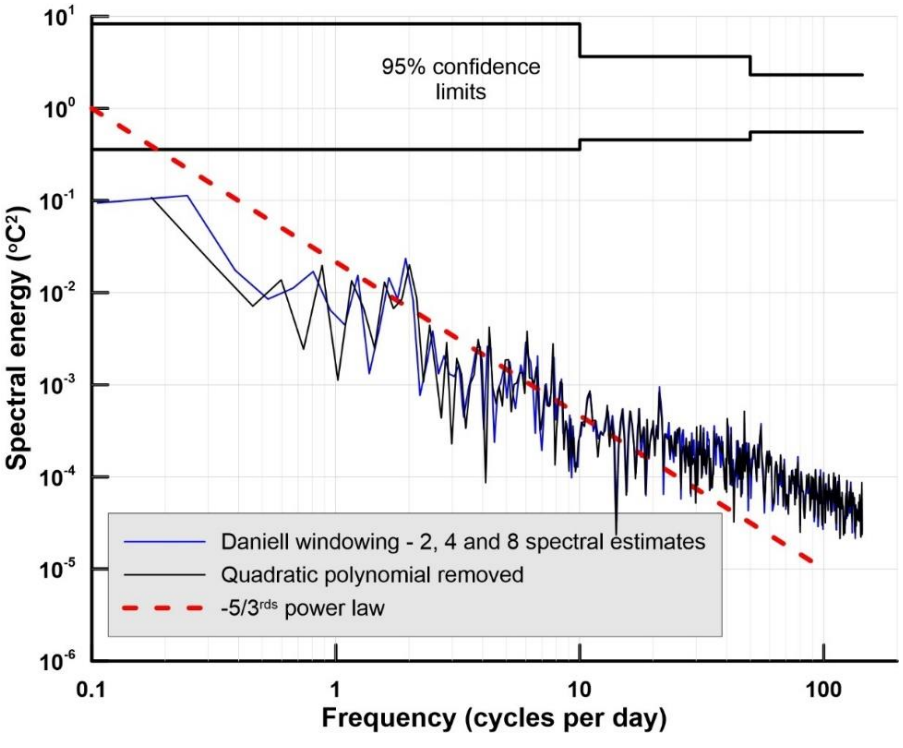


Figure 4–10. Replotting of Figure 4–9 with a new periodogram of the 5-minute temperature anomalies with the quadratic polynomial removed added.

4.3.2. Example 4-2: Water temperature – hourly data

Example 4-2 effectively repeats Example 4-1 using a larger time step and longer data set. The larger time step increases the Nyquist frequency, while the increased length of the data set enables longer period features to be resolved. Focus here is on the diurnal period, rather than general spectral peaks in the previous example.

Questions: Are there peaks in the water temperature spectrum at the diurnal period? If so, do these peaks exceed the 95% confidence limits?

Data used:

- Water temperature at 60 minute intervals, FFT=4,096 points.
- The Nyquist period is 2 hours (Nyquist frequency is ½ cycles per hour).
- The data set is about 170 days long, therefore the longest period features we can realistically interpret is about 20-30 days (i.e. we want about 5-10 cycles in our data set before we are confident of interpreting the low frequency end of the spectrum).

Pre-processing:

- A 7-point Gaussian average was used to smooth the original, 5 minute data in the time domain, then data were subsampled to hourly intervals.
- The linear trend was removed to produce temperature anomalies (Figure 4–11). Also shown in this figure is the frequency histogram of the anomalies.

What is immediately obvious from Figure 4–11 is the difference in high frequency variations in the first and second halves of the data set. Apart from potential stationarity problems, the spectral analysis will try to fit these high frequency fluctuations to the whole data set. So, the fit will likely underestimate the high frequency features in the first half of the data set while introducing spurious high frequency features to the second half of the data set.

Spectral analysis is probably not the best tool to examine these data. One alternate option is to split the data set into two halves and analyse each half separately.

Testing:

- Stationarity: The Dickey-Fuller test was used to check for stationarity of the time series. The calculated Tau value was -11.70, while the critical Tau was -2.86. As the Tau statistic is less than the critical value for Tau, we can reject the null hypothesis that the data set is not stationary. This result is a little surprising as the time series plot seems to indicate a large difference in the variance in the first and second half of the time series data set.
- Independence: Von Neumann's test for serial independence resulted in a t-calculated value of 59.88 and a t-critical value of 1.96 and it is concluded that the data are not serially independent at the 5% level.
- Normality: Based on the chi-squared goodness of fit test, the data are not normally distributed at the 5% level of significance. However, the frequency histogram (Figure 4–11) does 'look' normal in that it has a peak in the middle and tails off at each end of the distribution.
- Homogeneity of means and of variances: The anomalies were split into 16 groups of 256 data points each and subjected to Bartlett's test for homogeneity of variance and a 1-way ANOVA to compare the means. Neither test resulted in acceptance of the null hypothesis (i.e. that the variances and means could be from the same population) at the 5% level of significance.

As with the previous example, the likely reason for the rejection of the null hypotheses is that, with 4,096 data points, the statistical power is very high and even small deviations will result in such a rejection of the null hypothesis.

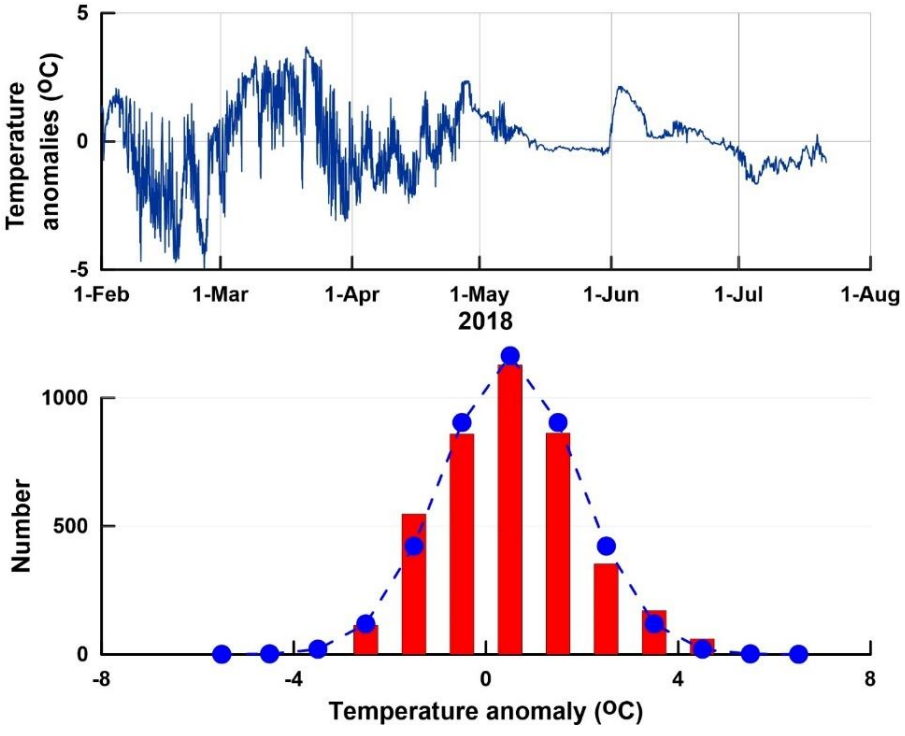


Figure 4–11. Top: Time series of the 60-minute temperature anomalies. Bottom: Frequency histogram of temperature anomalies (red bars are the data, blue circles are the expected values).

Smoothing of the periodogram:

- Daniell windowing is used to smooth the periodogram into frequency bands of size 2 (for frequencies lower than 0.5 cycles per day), 4 (for frequencies between 0.5 and 5 cycles per day) and 8 (for frequencies greater than 5 cycles per day).
- Smoothing over a different set of bands and choosing the smoothing that best highlights those features important to the question we're asking is recommended.
- The 95% confidence limits are presented.

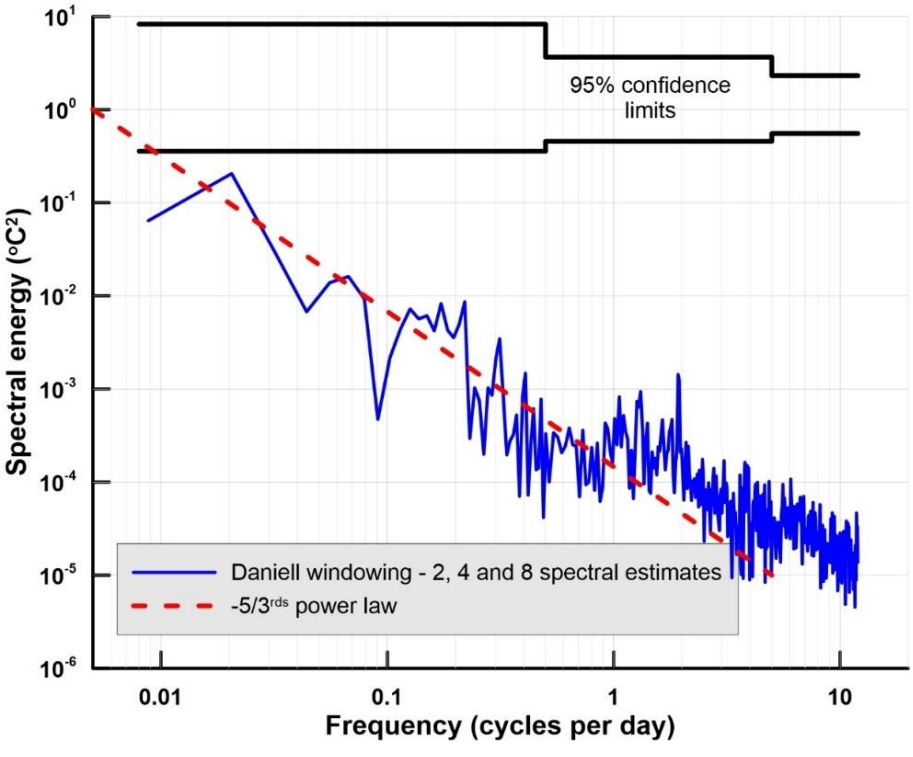


Figure 4–12. Periodogram of the 60-minute temperature anomalies. The 95% confidence limits (solid black lines) and the '-5/3rds power law' curve (dashed red line) are both shown.

Description (Figure 4–12):

- Spectral energy near 1 cycle per day contributes less than 1% to the total. A small peak (that may be significant) is present at 2 cycles per day.
- Maximum spectral energy occurs at low frequency, suggesting low frequency features remain in the time series (i.e. removing the trend did not remove all of the low frequency features).
- There is a departure from the -5/3rds power law at high frequencies. This increase is spectral energy suggests an external energy source at frequencies higher than about 1 cycle per day.

(This may be the result of internal tides and/or high frequency internal waves).

Finally, specifically addressing our questions (Figure 4–12):

(a) There is no spectral peak at the diurnal period (one cycle per day).
(b) The second question does not really apply as no significant spectral peak at the diurnal period was observed.

Above, after visually inspecting the time series plot, we suggested that a better way to analyse this data set would be to split the data into two halves and analyse each separately. The results of such an analysis are presented in Figure 4–13: each spectrum uses 2,048 data points in the FFT. The differences between the two spectra are substantial. The spectrum of the first half of the data set contains about ten times the energy than the spectrum of the second half of the data set. Both spectra contain most of their energy at the low frequency end of the spectrum. Neither spectrum contains much energy at the diurnal period – the question that we were asking. However, the spectrum of the first half of the data does contain a significant peak near 2 cycles per day. This is likely the result of semi-diurnal internal tides.

This example highlights the need to critically examine the data set (i.e. plot your data) before undertaking a spectral analysis.

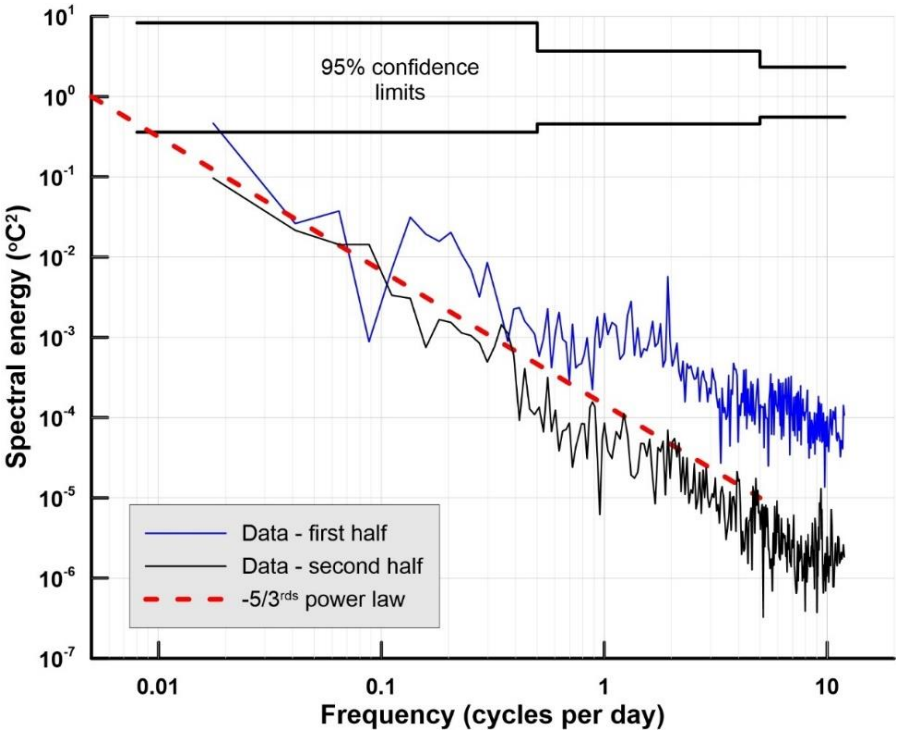


Figure 4–13. Periodograms for each half of the time series (Figure 4–11).

4.3.3. Example 4-3: Water temperature – daily data

Example 4-3 uses a long data set (11 years) and focusses on the annual signal in water temperatures. Specifically, we'll highlight differences in the spectra when the annual and semi-annual signals are, and are not, removed from the data set.

Questions: What is the difference in the spectral signal for daily average water temperature when the annual signal is and is not removed?

Data used:

- Water temperature at daily intervals, FFT=4,096 points.
- The Nyquist period is 2 days (Nyquist frequency of ½ cycles per day).
- The data set is about 11 years long, therefore the longest period features we can realistically interpret is about 1-2 years (i.e. we want about 5-10 cycles in the data set before we can be confident of interpreting the low frequency end of the spectrum).

Pre-processing:

- The original data were sampled at 5-minutes intervals.
- A 7-point Gaussian average was used to smooth the data in the time domain, then data were subsampled to hourly intervals.
- A cosine-Lanczos filter with 24 weights in each wing and half-amplitude response 1 cycle per day was applied to the hourly data.
- The trend was removed from the data to give daily temperature anomalies (Figure 4–14, top plot, blue line).
- The annual and semi-annual curves were determined (Figure 4–14, top plot, red line) and the daily anomalies with the annual and semi-annual curves removed replotted (Figure 4–14, middle plot).

One observation is that there does not appear to be much difference between the two time series.

Testing:

- Stationarity: The Dickey-Fuller test was used to check for stationarity of the time series. The calculated Tau value was -21.97, while the critical Tau was -2.86. As the Tau statistic is less than the critical value for Tau, we can reject the null hypothesis that the data set is not stationary.
- Independence: Von Neumann's test for serial independence resulted in a t-calculated value of 50.42 and a t-critical value of 1.96 and it is concluded that the data are not serially independent at the 5% level.
- Normality: Based on the chi-squared goodness of fit test, the data are not normally distributed at the 5% level of significance. However, the frequency histogram (Figure 4–14) does 'look' normal in that it has a peak in the middle and tails off at each end of the distribution.
- Homogeneity of means and of variances: The anomalies were split into 16 groups of 256 data points each and subjected to Bartlett's test for homogeneity of variance and a 1-way ANOVA to compare the means. Neither test resulted in acceptance of the null hypothesis (i.e. that the variances and means could be from the same population) at the 5% level of significance.

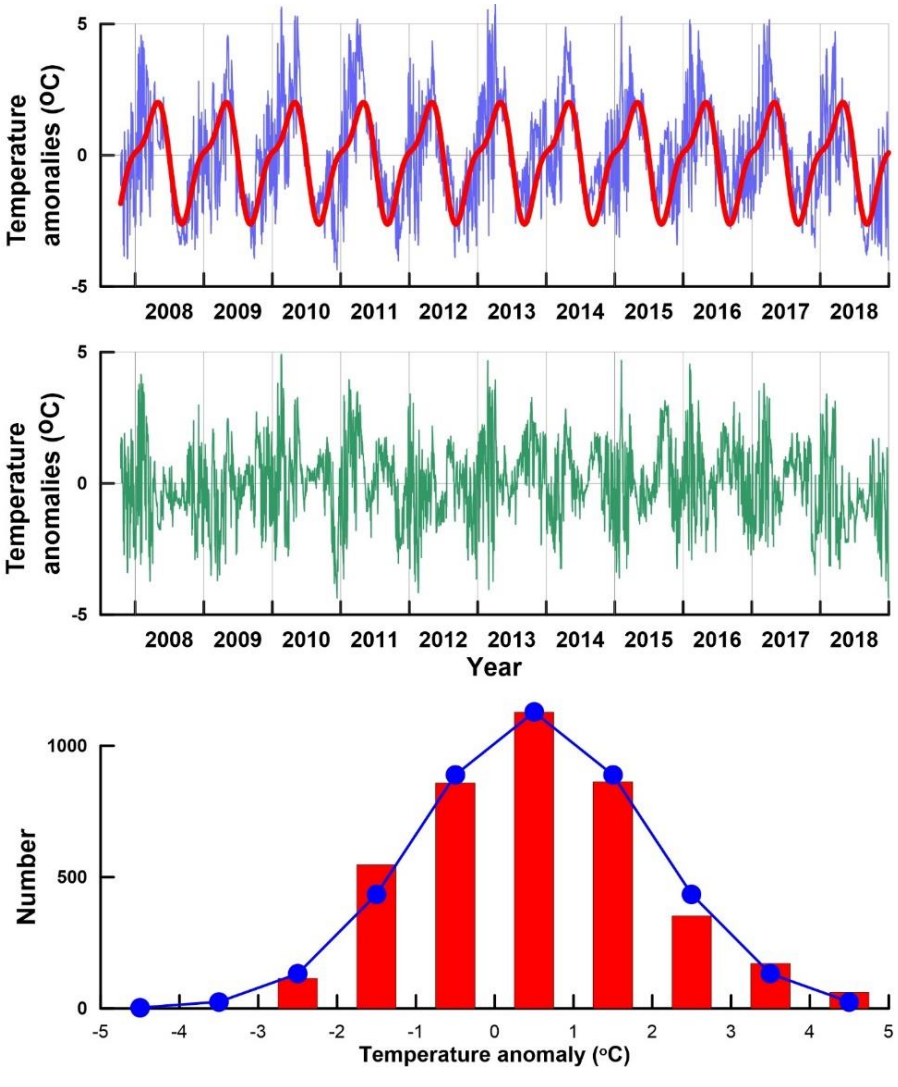


Figure 4–14. Time series of the daily temperature anomalies without the annual curve removed (top, blue) and the annual curve (top, red). Anomalies with the annual curve removed are shown in the middle plot. The frequency histogram of the anomalies with the annual curve removed are shown in the bottom plot (red bars are the data, blue circles are the expected values).

As with the previous two examples, the likely reason for rejection of the null hypotheses is that, with 4,096 data points, the statistical power is very high and even small deviations will result in a failure of many of these statistical tests.

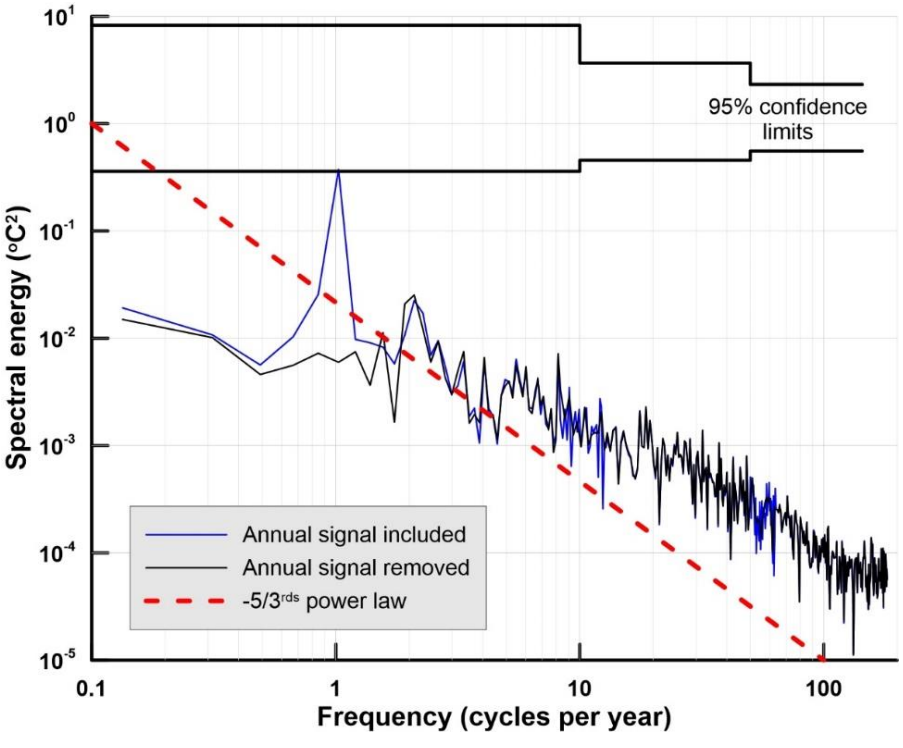


Figure 4–15. Periodogram of the daily temperature anomalies with (black line) and without (blue line) the annual curve removed. The 95% confidence limits (solid black lines) are shown together with the "-5/3rds power law" curve (dashed red line).

Smoothing of the periodogram:

- Daniell windowing is used to smooth the periodogram into frequency bands of size 2 (for frequencies lower than 8 cycles per year), 4 (for frequencies between 8 and 50 cycles per year) and 8 (for frequencies greater than 50 cycles per year).
- The 95% confidence limits are presented.

Description (Figure 4–15):

- The annual signal represents about 45% of the total spectral energy and the height of the spectral peak exceeds the distance between the two confidence limits at the frequency of 1 cycle per year.
- Despite removing both the annual and semi-annual signals, there still appears to be some energy in both plots at 2 cycles per year. However, the heights of these spectral peaks do not exceed the distance between the confidence limits, and the peaks are not statistically significant.
- Visually, these also appears to be some difference in the variability of the two spectral plots near 10 cycles per year (i.e. the monthly period) and again near 50-60 cycles per year (i.e. the weekly period). It is unclear why the removal of the annual and semi-annual curves would cause this difference.
- As the frequency increases beyond about 10 cycles per year, both spectral plots start to deviate from the -5/3rds power law. This suggests an energy source at the monthly and shorter periods – possibly internal tides and high frequency internal waves.

Addressing our question:

(a) The difference in the spectral estimates when the annual signal is, and is not, removed is large at the annual period – as expected (Figure 4–15).

Thus far we've looked at three examples, each with different time spans and with different values for delta-t. In all three examples we find that the tests for stationarity, serial independence, normality, homogeneity of the means and homogeneity of the variances all reject their respective null hypotheses. The likely reason for this is because there are thousands of data points. When a very large number of data points are involved, the statistical power is very large (probably >99%). This indicates that very small departures from the expected values become statistically important, although such differences may not be meaningful. This is likely to occur for many (perhaps most) time series problems. Therefore, from a pragmatic point of view, we note that this is likely to occur for all examples and proceed with our spectral analyses.

Note: we have explicitly omitted using the KPSS test for stationarity. In all cases, the null hypothesis (that the data set is stationary) will be rejected. This result is contrary to the Dickey-Fuller test.

4.3.4. Example 4-4: Water temperature – monthly data

This example examines very long period water temperature data from 1900-2019. The data comprise monthly Sea Surface Temperature (SST) data from the Tasman Sea (the body of water between Australia and New Zealand). The data cover a large area and span almost 120 years. Therefore, the coverage and quality of these data may be variable. The 30 year period 1961-1990 is used as a base and are used to generate SST anomalies. Data are from the Australian Government's Bureau of Meteorology.

Questions: Are there any significant spectral peaks in the frequency 0.1 – 1 cycle per year?

Data used:

- SST anomalies from the Tasman Sea at monthly intervals; 1,431 data points (Figure 4–16).
- The Nyquist period is 2 months (Nyquist frequency is about 1/60 cycles per day or 6 cycles per year).
- The data set is about 120 years long, therefore the longest period features we can realistically interpret is about 10-20 years (i.e. we want about 5-10 cycles in the data before we can be confident of interpreting the low frequency end of the spectrum).

Pre-processing: The original data were based on data from different sources and averaged into monthly bins.

- The 30 year period 1961-1990 was used to estimate an annual curve, which has been removed from the data. (Using the trigonometric functions over the whole data set to remove the annual and semi-annual signals produced almost identical results, Figure 4–17).
- The linear trend was also removed from the data to give the de-trended monthly SST (Figure 4–16).

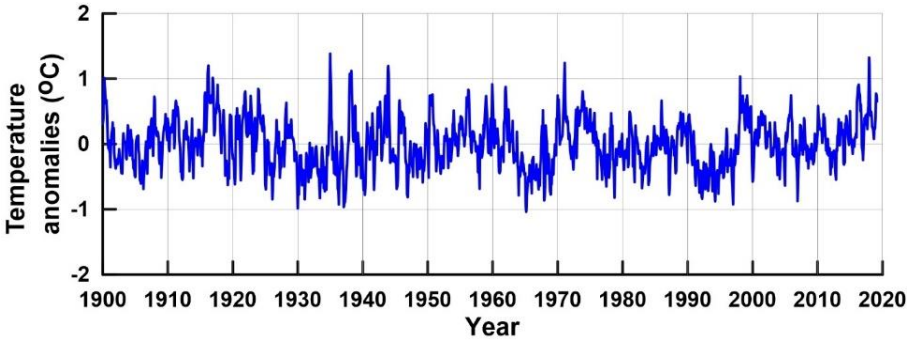


Figure 4–16. Time series of the monthly SST anomalies in the Tasman Sea.

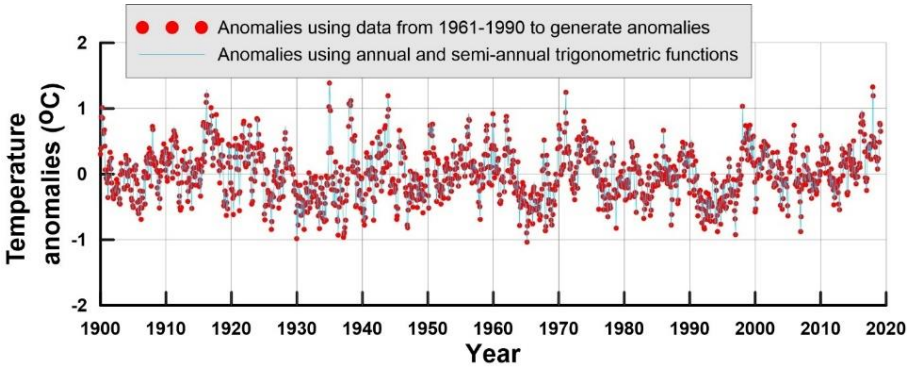


Figure 4–17. Time series of the monthly SST anomalies in the Tasman Sea using both the 1961-1990 time period for the annual signal (red circles) and the trigonometric functions as the annual signal (light blue line).

Testing:

- Stationarity: Results from the Dickey-Fuller test: Tau calculated is -12.15 while the Tau critical value is -2.86. As Tau calculated is less than Tau critical we reject the null hypothesis that the data are not stationary in favour of the alternate hypothesis – the data are stationary.

- Based on the independence, normality, homogeneity of means and homogeneity of variances, the data do not normally meet these conditions at the 5% level of significance. We note this and proceed with the spectral analysis.

We have 1,431 data points and this suggests (at least) three ways of undertaking the spectral analysis:

(a) Removing one data point and using the 'normal' or 'slow' Fourier transform (to distinguish it from the FFT). This will be a bit slower in terms of data execution and will need to be done outside the Excel routine, but we will use all of the data available (less one data point to provide an even number of data points for the analysis).
(b) Padding the data with 617 zeroes to the anomalies and using the FFT with 2,048 data points. This is more than 40% of the length of the original time series and this is likely to impact the results. However, the length of the data set will be increased and lower frequency features will be better resolved. No tapering was undertaken.
(c) Using 3 x 512 point FFTs, with some overlap. The shorter length of each time series means that we won't be able to resolve the low frequency end of the spectrum, which will reduce our ability to answer our question. However, we can use the 3 sets of FFTs to estimate the confidence limits without having the undertake any band averaging (e.g. Daniell windows). Other variations on this theme include: 2 x 1,024 point FFTs with an overlap of about 600 data points and so on.

All three approaches are examined and compared.

Smoothing of the periodogram:

- Daniell windowing is used to smooth the periodogram into frequency bands of size three across the spectrum for each of the Fourier transform methods used.
- The 95% confidence limits are presented.

Description (Figure 4–18):

- The effect of a large amount of padding in the data set is clear – the green periodogram contains considerably less energy than the other two periodograms. However, it does produce spectral estimates further into the low frequency region (our region of interest, based on the question that we have asked), although the reliability of these estimates may be questionable.
- Low frequency spectral energy of the '3 x 512' FFTs is reduced, compared with the 'slow' Fourier transform results. In general, spectral energy of the '3 x 512' FFTs is slightly lower than that resulting from the 'slow' Fourier transform.
- It is also noted that most energy occurs at low frequency, suggesting that the data set still contains a long period, inter-annular feature.

Given that we're interested in the low frequency end of the spectrum, using the 'slow' Fourier transform would be preferred (or waiting until an additional 617 data points are available, i.e. another 51 years).

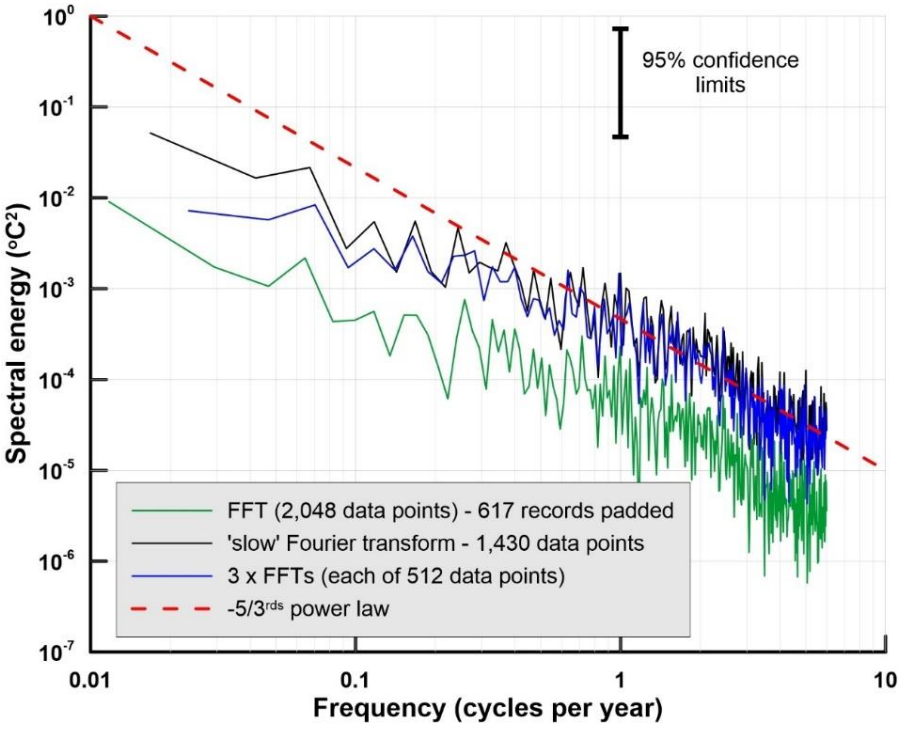


Figure 4–18. Periodograms of the monthly temperature anomalies using three different approaches: the 'slow' Fourier transform with 1,430 data points (black line), padding to produce 2,048 data points then use the FFT (green line) and average of the 3 x FFTs each using 512 data points (with some overlap), blue line. The 95% confidence limits are shown together with the '-5/3rds power law' curve (dashed red line).

To answer our question, there are no significant peaks at the low frequency (0.1 – 1 cycle per year) end of the spectrum in any of the three periodograms (Figure 4–18). We also note that large amounts of padding of the data set (about 40% of the total data set) results in a considerable reduction in spectral energy.

The previous examples have all examined water temperature. Generally, there is only one signal is present in oceanic temperature plots – at the annual period, at least this is the dominant signal. Therefore, it is, perhaps, relatively easy to understand where we should find the greatest spectral energy. The exception is when we look at short time spans of high frequency temperature signals (say, >1 cycle every 30 minutes). The following set of examples will focus on water level and in particular the tides.

Tides are more complex than water temperature as they comprise of many constituents resulting in many spectral peaks that will be of varying degrees of importance depending on the type of tidal signal and the location of the monitoring site.

One question that we may like to ask concerns the separation of peaks associated with different processes. Some tidal constituents have similar periods (or frequencies) and data collected over a long period of time is required for such signals to be effectively interpreted.

4.3.5. Example 4-5: Water level – hourly data

Example 4-5 focuses on water level, particularly those frequencies associated with the main tidal signals (i.e. 1 and 2 cycles per day). The two data sets we've chosen to analyse are about 21, 170 and 512 days long, with sampling at hourly intervals. We'll examine each data set separately, then compare the results.

Questions: Can we identify the diurnal and semi-diurnal spectral peaks in the water level? At what frequencies do these spectral peaks occur? Do these peaks align with the corresponding tidal periods? What happens when we have data sets of different length?

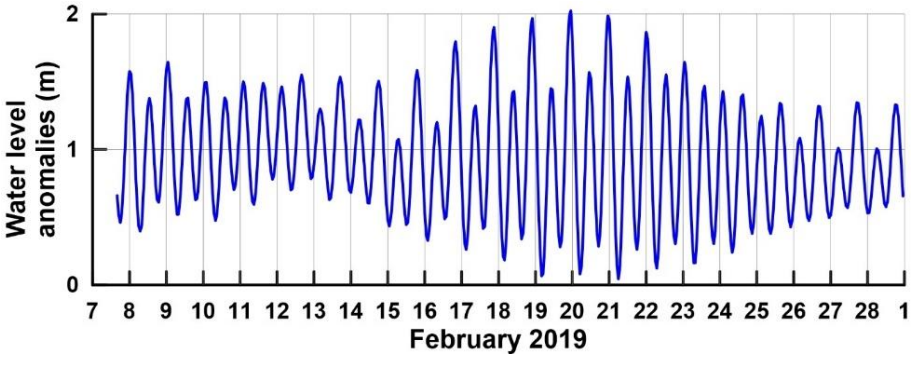


Figure 4–19. Time series of 512 hourly water level anomalies near Sydney, Australia.

Data used:

- Hourly water levels. FFT uses 512 data points (Figure 4–19).
- The Nyquist period is 2 hours (Nyquist frequency of ½ cycles per hour). This sampling frequency should be sufficiently small to resolve the semi-diurnal tidal signal: about 12 hours, in which case we'll have about 6 cycles at the highest frequency we wish to resolve.
- The data set is about 21 days long, therefore the longest period features we can realistically interpret is about 2-3 days (i.e. we want about 5-10 cycles in the data set before we are confident of

interpreting the low frequency end of the spectrum). This should be of sufficient length to resolve the diurnal signal requested by our question.

Pre-processing:

- The linear trend was removed from the data producing a set of water level anomalies.
- No other processing was undertaken.

Testing:

- The Dickey-Fuller test results in a Tau calculated value of -5.57 while the Tau critical value is -2.86. As Tau calculated is less than Tau critical we reject the null hypothesis that the time series is not stationary.
- Based on the relevant statistical tests, the data are not independent, normally distributed, nor are the means and variances homogeneous at the 5% level of significance. We note this and continue with the spectral analyses.

Smoothing of the periodogram:

- No smoothing of the periodogram was undertaken. This would normally be done to determine the confidence limits for the spectrum. However, at this stage, we just want to demonstrate what happens when we have data sets of different length.

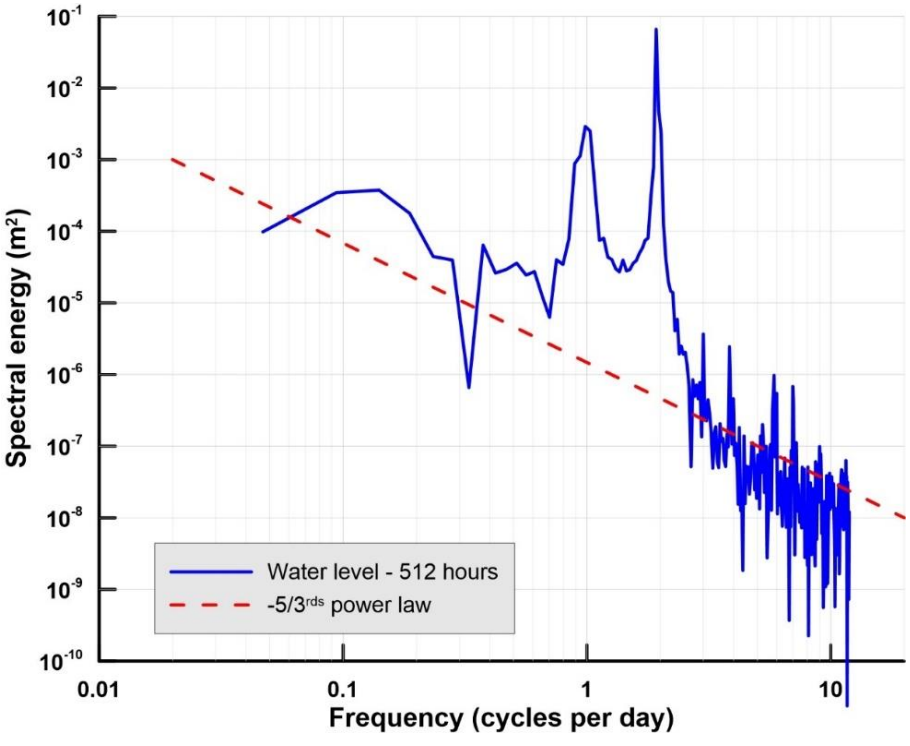


Figure 4–20. Periodogram of the hourly water level anomalies using 512 data points.

Description (Figure 4–20):

- The diurnal and semi-diurnal spectral peaks in the water level can be clearly identified. However, the spectral peaks are relatively broad, and we can't separate the different tidal components within each of the diurnal and semi-diurnal frequency bands.
- The periodogram generally follows the -5/3rds power law except for the regions surrounding 1 and 2 cycles per day. These correspond to the diurnal and semi-diurnal tidal frequencies, respectively.
- There are also several peaks at higher frequencies resulting from the higher harmonics.

- We do know that the semi-diurnal and diurnal tidal components each comprise four major constituents (Table 4–1). The periods of the four semi-diurnal components are all similar as are the periods of the four diurnal components.

Table 4–1. Characteristics of the major diurnal and semi-diurnal tide producing forces.

Tide name	Symbol	Period (solar hours)	Frequency (cycles per day)
Semi-diurnal components			
Principal lunar	M2	12.42	1.932
Principal solar	S2	12.00	2.000
Larger lunar elliptic	N2	12.66	1.896
Luni-solar semi-diurnal	K2	11.97	2.005
Diurnal components			
Luni-solar diurnal	K1	23.93	1.003
Principal lunar diurnal	O1	25.81	0.930
Principal solar diurnal	P1	24.07	0.997
Larger lunar elliptic	Q1	26.87	0.893

With a data set of length 512 hours, it is not possible to separate the four semi-diurnal components nor can we separate the four semi-diurnal components.

As alluded to in Section 4.2, it turns out that to resolve spectral components that are close together we need to sample over a long period of time. In this case, 512 hours is just not sufficient time to separate the different semi-diurnal tidal components (nor is it sufficiently long to separate the different diurnal tidal components).

So, let's now repeat the analysis but using a longer data set. In particular 4,096 hours (about 170 days). The time series is presented in Figure 4–21 (top plot) along with the corresponding periodogram (bottom plot). Using 4,096 hours of water level data the separation of the individual tidal components can be identified (Figure 4–21, bottom plot) although the components provided in Table 4–1 cannot all be clearly identified. To do so, an even longer time series is required. Foreman and Henry (1989) show that a minimum of 188 days of water level data is required to separate the four semi-diurnal and diurnal tidal components given in Table 4–1.

To effectively interpret the tidal components listed in Table 4–1 new have two options: (a) we increase the size of the FFT, or (b) we increase the time between data readings. The question that is being asked (viz. concerning the diurnal and semi-diurnal tidal components) provides some scope to do either (or both). It is noted that if we increase the time between data points, we may encounter problems with the Nyquist frequency.

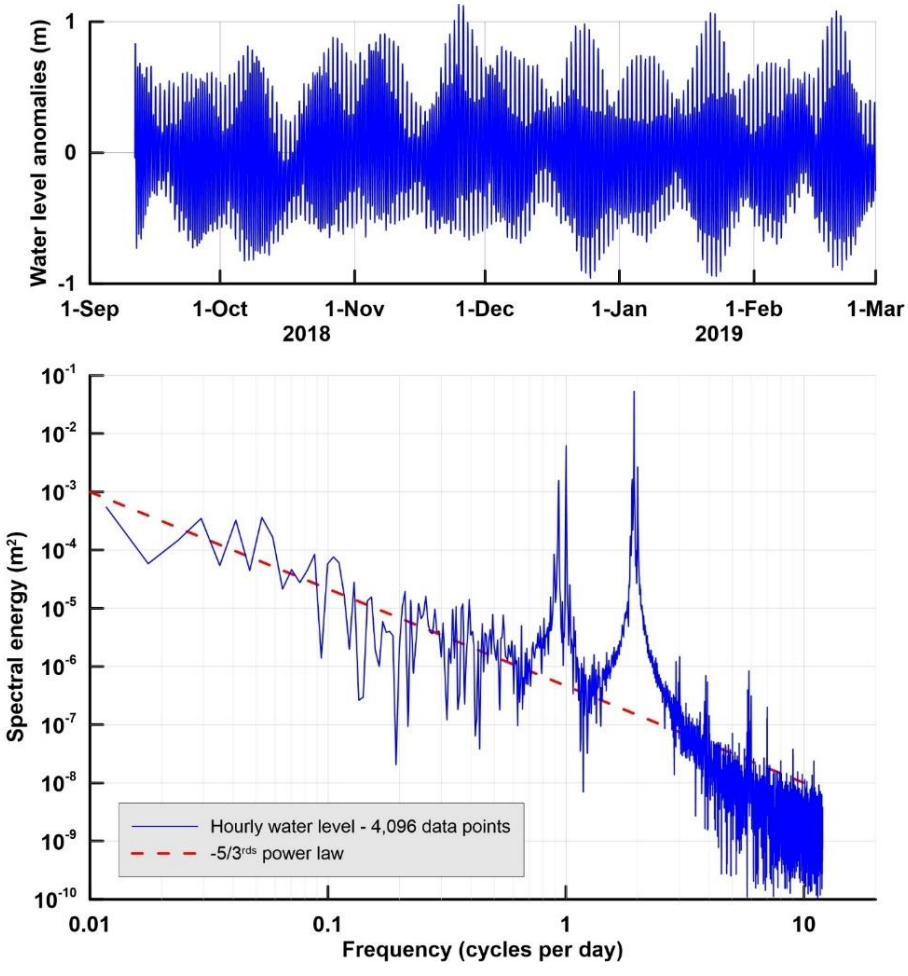


Figure 4–21. Time series of 4,096 hourly water level data (i.e. 170 days) near Sydney, Australia (top) and corresponding periodogram (bottom).

Let's continue to use the Excel FFT routine, which is restricted to 4,096 data points. We can average (or subsample) the time series data into, say, 3 hourly bin and continue with our 4,096 data points in the FFT. Alternatively, we could use the Matlab or Fortran FFT routines, continue using the hourly data but over a much longer time span (e.g. 16,384 hourly records, or slightly less than 2 years).

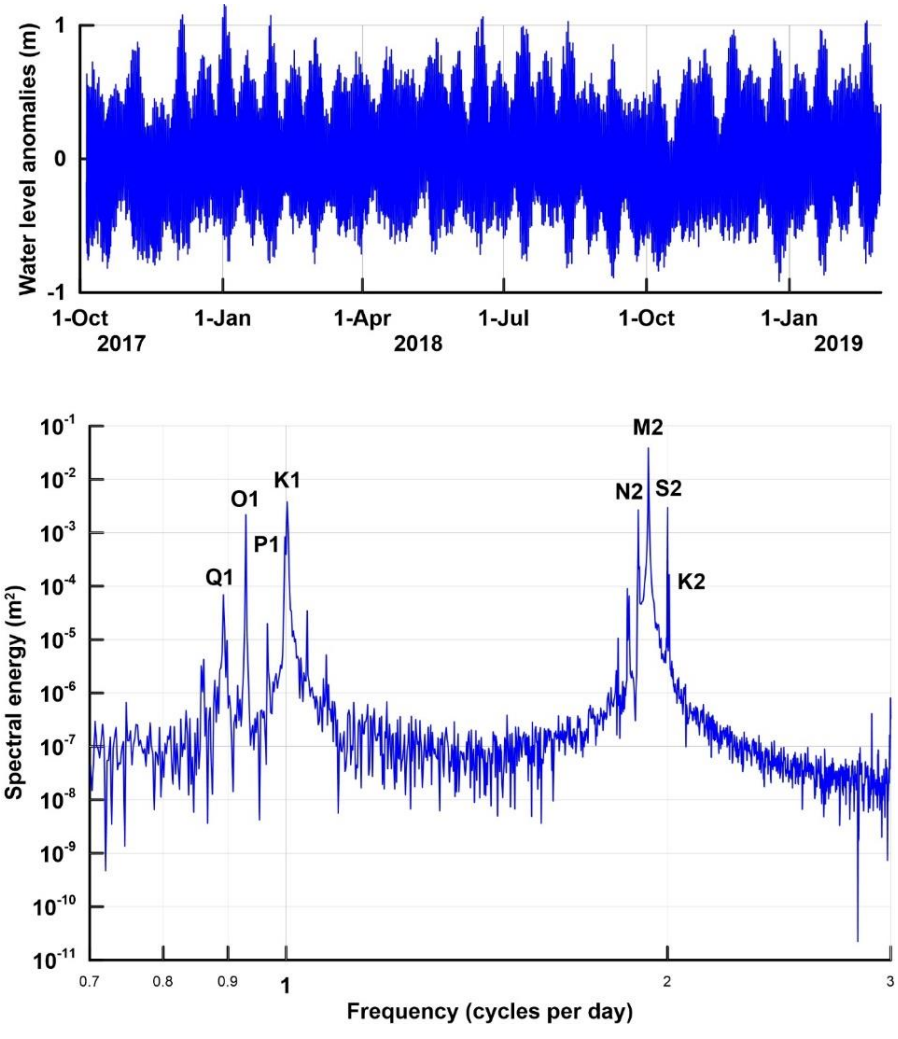


Figure 4–22. Time series of 4,096 three-hourly (i.e. 512 days) water level data near Sydney, Australia (top) and corresponding periodogram (bottom).

A 3-point Gaussian average of the hourly data was undertaken to produce a 3 hourly data set, spanning 512 days and is plotted in Figure 4–22 (top plot). The Nyquist period is now 6 hours and we only have 2 cycles that fall inside the semi-diurnal tidal signal. This violates our rule-of-thumb

which (ideally) requires a minimum of about 5 cycles before the signal can be effectively interpreted (see Section 3.7). The semi-diurnal tidal signal is very strong in this data set and this rule of thumb can be relaxed.

The corresponding periodogram is also plotted (Figure 4–22, bottom plot). As we're interested only in the diurnal and semi-diurnal periods, the frequency range plotted is between 0.7 and 3 cycles per day. The four diurnal and four semi-diurnal tidal components can be identified and have been marked on Figure 4–22 (bottom plot). Even with 512 days of data, it is still difficult to clearly separate the K1 and P1 diurnal tidal components and the S2 and K2 semi-diurnal components.

Returning to the questions that we want to answer:

(a) Yes, we can we identify peaks at the diurnal and semi-diurnal periods in the periodogram of water level from a time series of length 512 hours.
(b) The frequencies at which these peaks occur can be read directly from the periodogram.
(c) Comparing the frequencies at which the spectral peaks occur with the information in Table 4–1, we can confirm that these spectral peaks do indeed correspond to the tidal periods.
(d) The analyses were undertaken for time series of lengths 512 hours (i.e. 21 days), 4,096 hours (i.e. 170 days) and 12,288 hours (i.e. 512 days). While all three spectra allowed us to identify large spectral peaks at the diurnal and semi-diurnal periods, resolution of the tidal features within each of the diurnal and semi-diurnal periods could only be effectively interpreted using the longest of these time spans. Hence, the longer the data set, the better the resolution of spectral features that can be achieved.

4.3.6. Example 4-6: - Water level – weekly data

Example 4-6 again examines the water level. The previous example looked at the relatively high frequency signals of the diurnal and semi-diurnal tides. This example focuses on the low frequency end of the spectrum. If we want to examine low frequency features, we need to use a long data set. The data chosen are weekly samples (with the trend, diurnal, semi-diurnal and spring / neap signals removed). The length of the data set is about 10 years.

Questions: What does the spectrum of the very long period water level show? At what frequencies do these spectral peaks occur?

Data used:

- Weekly water level data from Sydney, Australia.
- The Nyquist period is 2 weeks (Nyquist frequency of 1/14 cycles per day).
- The data set is about 10 years long, therefore the longest period features we can realistically interpret is about 1-2 years (i.e. we want about 5-10 cycles in the data set before we can be confident of interpreting the low frequency end of the spectrum).

Pre-processing:

- Hourly water level data were subject to the 120i913 filter described in Thompson (1983) and subsampled to weekly intervals. This filter is designed to remove features with tidal and inertial periods. It is noted that this filter is about 10 days long, hence we really should use a delta-t of 10 days or longer. Weekly data is sufficiently close to 10 days for this demonstration.
- Time series plot is shown in Figure 4–23.
- The linear trend was removed from the data.

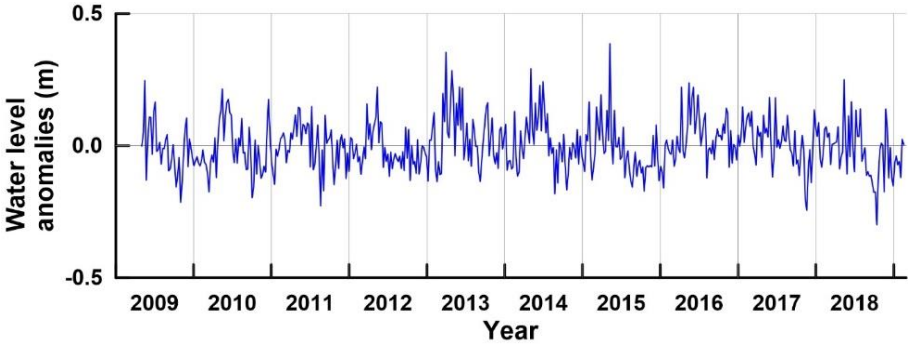


Figure 4–23. Time series of 512 weekly water level data. The diurnal and semi-diurnal tidal components have been removed.

Testing:

- Stationarity: Results from the Dickey-Fuller test indicate Tau calculated value of -15.59, while the Tau critical value is -2.86. As Tau calculated is less than Tau critical, we reject the null hypothesis that the time series is not stationary.
- Based on the relevant statistical tests, the data are not independent, normally distributed, nor are the means and variances homogeneous, at the 5% level of significance. This is noted and we proceed with the spectral analysis.

Smoothing of the periodogram:

- No smoothing of the periodogram was undertaken. This would normally be done to determine the confidence limits for the spectrum. The purpose of this example was to identify long period components in the spectrum rather than to examine spectral confidence limits.

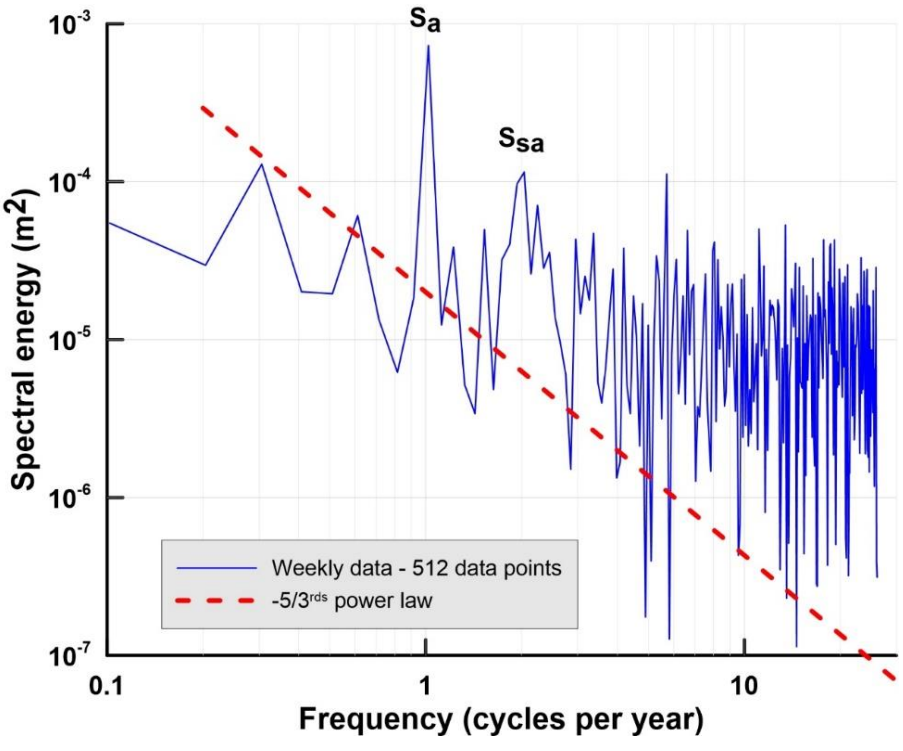


Figure 4–24. Periodogram of 512 weekly water level anomalies.

Description (Figure 4–24):

- The very long period water level data shows a single, large spectral peak together with several smaller peaks at higher frequencies.
- The periodogram shows peaks at the annual (Sa) and semi-annual (Ssa) periods. These correspond to the solar annular tide and the solar semi-annular tide, respectively.
- A spectral peak also appears near 6 cycles per year, the driving force of which is unknown and may warrant further investigation. It is possible that this is a harmonic of the semi-annular tide.

Answering our questions:

(a) The periodogram indicates relatively large energy at a single frequency. The relative energy in this signal is almost one order of magnitude greater than any other signals in the data set.
(b) The dominant spectral peak occurs near one cycle per year, corresponding to the solar annual tidal signal. A second spectral peak occurs near two cycles per year, corresponding to the solar semi-annular tidal signal.

4.4. Spectral analysis in the space domain

Thus far, we've only considered spectral analysis with time as the independent variable. Spectral analysis can also be undertaken in space. While the time variable transforms to frequency (the inverse of time), the spatial variable transforms to wave number (the inverse of space).

In time, the Fourier transform / inverse transform pair are:

$$\mathcal{F}(\omega) = \int_{-\infty}^{\infty} f(t)\, e^{-i\omega t}\ dt \quad \text{and} \quad f(t) = \frac{1}{2\pi}\int_{-\infty}^{\infty} \mathcal{F}(\omega) e^{i\omega t}\ d\omega$$

In space, the Fourier transform / inverse transform pair becomes:

$$\mathcal{F}(\kappa) = \int_{-\infty}^{\infty} f(z)\, e^{-i\kappa z}\ dz \quad \text{and} \quad f(z) = \frac{1}{2\pi}\int_{-\infty}^{\infty} \mathcal{F}(\kappa) e^{iz\kappa}\ d\kappa$$

Where z is the spatial co-ordinate and κ is the wave number.

In all other respects, the spectral analysis is the same whether the analysis is in the space-wave number or time-frequency domains. However, the interpretations of the results may differ. Intuitively, we find the time-frequency domain easier to understand than the space-wave number domain. This may simply be a matter of familiarity. One way in which this can be visualised is using harmonics, which will be discussed in detail in Chapter 7.

4.4.1. Example 4-7: Water temperature – 4 metre data

This example uses temperature data from 14 locations in the water column. The data were obtained from a thermistor string with temperature sensors located every 4m from near the sea floor to near the sea surface. The water depth is about 65 m.

Time series plots of the temperature data over a 3 hour period are shown in Figure 4–25 (bottom plot). Sometimes it is more instructive to plot the depth of each isotherm (Figure 4–25, top plot) rather than the temperature at each depth. These data show an internal wave train, with about individual 4-5 waves in the train. The height of the first internal wave is about 12 m, this height decreasing as we move towards the end of the wave train.

Three FFTs are examined, corresponding to the times of 19:20, 19:25 and 19:30 on 29 November 2019. This is indicated by the shaded box in Figure 4–25. A single spatial FFT is run at each of these three times.

Questions: At what wave number is the spectral energy greatest? Does this differ for the different spectra? Is the spectrum markedly different under the crest or trough of the internal wave?

Data used:

- 5-minute temperature data from Sydney, Australia. There are 14 temperature sensors aligned vertically in the water column, each 4 m apart.
- The Nyquist wave number is 1/8 cycles per metre.
- The data set is 56 metres long, therefore the longest spatial feature that can be resolved is 28 metres.

Pre-processing:

- There were 14 levels from which temperature data were obtained (at every 5 minute time step). One data point was added to each end of the data (at each time). This provided data from 16 levels and enables us to use the radix 2 FFT.

- At each time, the spatial average was removed from the data, resulting in a data set of spatial anomalies. This is equivalent to removing a linear trend from a time series.

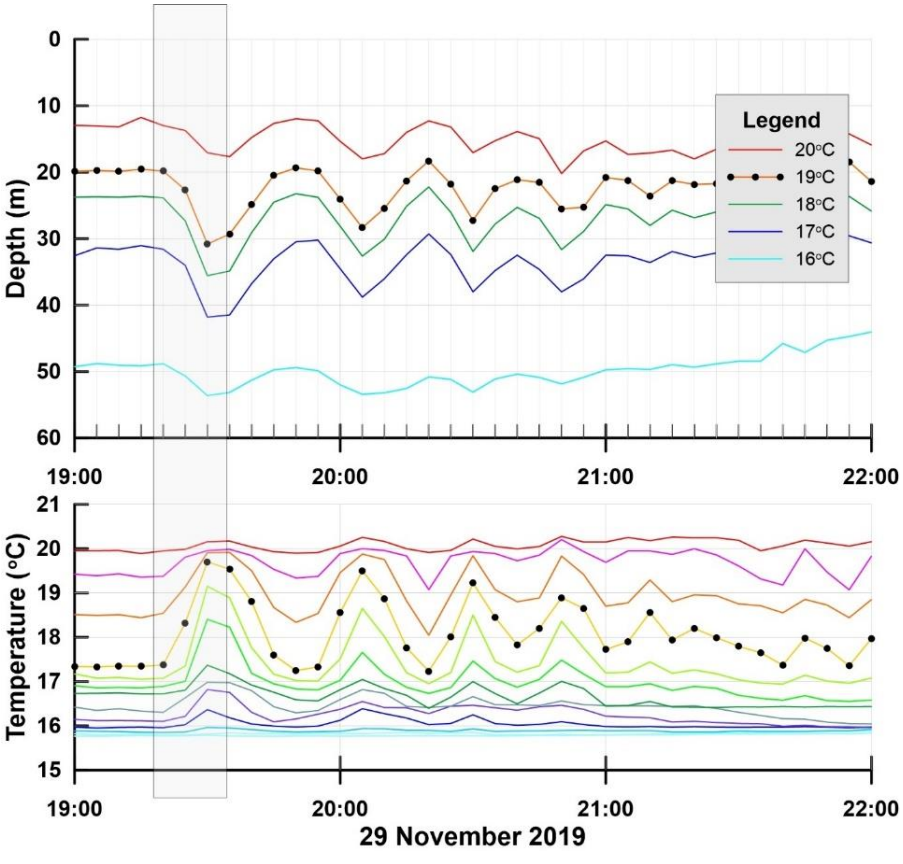


Figure 4–25. Time series of (bottom) water temperature every 4m throughout the water column, and (top) the depth of the 16°C to 20°C isotherms. Black circles indicate the sampling times for one of the isotherms and for one of the thermistors. The shaded box indicates the three times at which the FFT was undertaken.

Testing:

- Stationarity: Results from the Dickey-Fuller test indicate Tau calculated values of -1.14, -1.57 and -1.95 while the Tau critical

value is -2.99. The calculated Tau is greater than Tau critical values, hence we cannot reject the null hypothesis that the spatial data sets are not stationary.

- As each time step only contained 16 data points, we felt that this was insufficient to provide reliable statistics on independence normality and homogeneity of the means and variances. These tests were not undertaken, which is noted and we proceed with the spectral analysis.

Smoothing of the periodogram:

- No smoothing of the periodogram was undertaken, hence there are no confidence limits associated with these spectra.

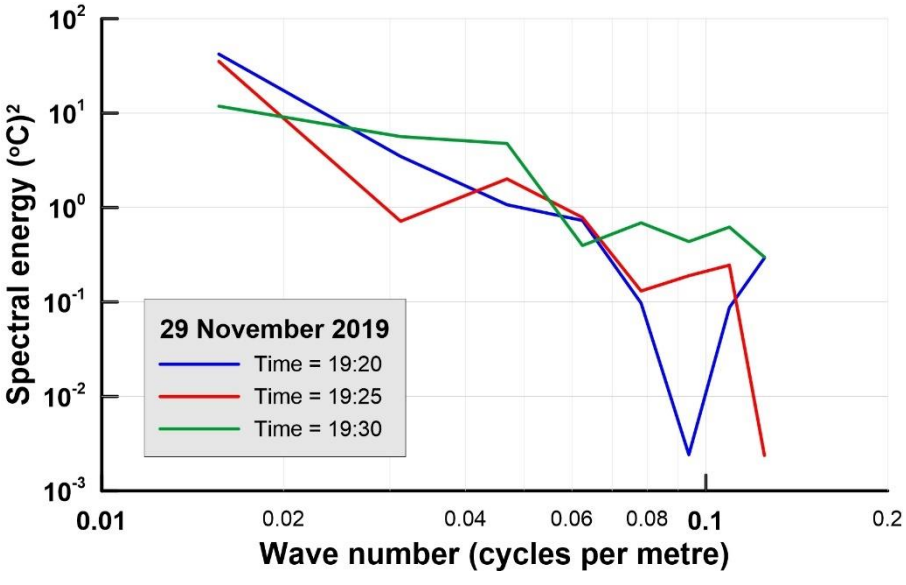


Figure 4–26. Three spatial FFTs, at times 19:20, 19:25 and 19:30 on 29 November 2019.

Description (Figure 4–26):

- The largest spectral energy occurs at low wave number. This may be analogous to the fundamental frequency for a vibrating string.

- There are no spectral peaks of note in any of the three FFTs examined.

Answering our questions from above:

(a) The maximum spectral energy occurs at the smallest wave number (about 0.015 cycles/metre).
(b) The maximum values occur at the same wave number for all three FFTs examined.
(c) There does not appear to be a marked difference in the spectra under the crest (at time 19:20) or the trough (at time 19:30) of the internal wave.

4.4.2. Example 4-8: Across-shelf currents – 2 metre data

This example uses temperature data from 29 depths in the water column. The data were obtained from an Acoustic Doppler Current Profiler, taking current speed and direction readings (in east-west and north-south component form) every 2 m throughout the water column. The times and locations are the same as those used in the previous example. The water depth is about 65 m.

Time series plots of the east-west current component (approximately the across-shelf component) for three time are shown in Figure 4–27. The data indicate flow in opposite directions in the upper and lower parts of the water column. This is indicative of internal wave activity.

Three FFTs are examined, corresponding to the times of 19:20, 19:25 and 19:30 on 29 November 2019. A spatial FFT is run at each of these times.

Questions: At what wave number is the spectral energy greatest? Does this differ for the different spectra? Is the spectrum markedly different under the crest or trough of the internal wave?

Data used:

- 5-minute temperature data from Sydney, Australia. There are 29 east-west components of the currents, each 2 m apart.
- The Nyquist wave number is 1/4 cycles per metre.

- The data set is 58 metres long, therefore the longest spatial feature that can be resolved is 29 metres.

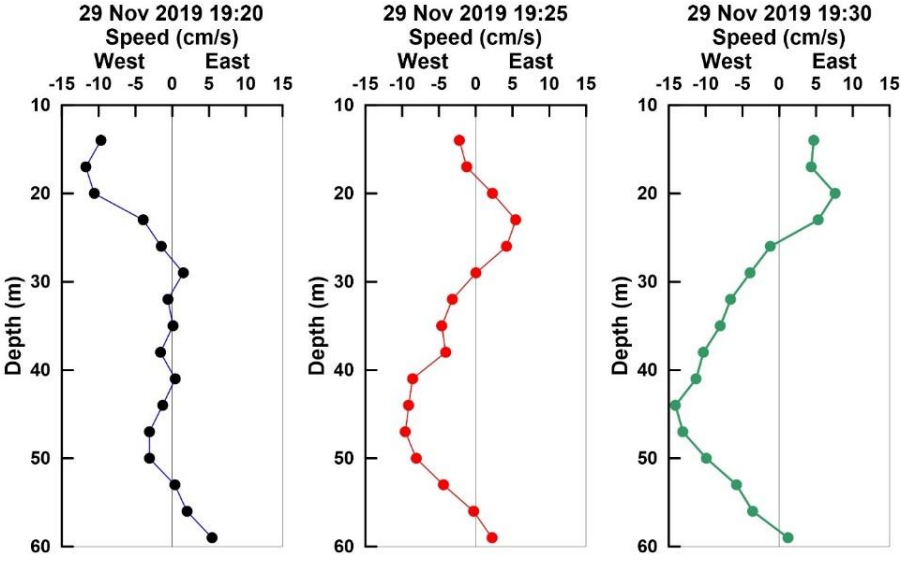


Figure 4–27. Time series of east-west current component as a function of water depth at three times. Negative values indicate currents flowing towards the west, while positive values indicate currents flowing towards the east.

Pre-processing:

- There were 29 levels from which temperature data were obtained (at every 5 minutes). We could subtract one data point to obtain an even number of records, then undertake the spectral analysis. Another option (which we used) subsampled the data to one reading every 3 m, which provides data from 16 levels in the water column. This then allows us to use the radix 2 FFT algorithm.
- However, this processing does change the Nyquist wave number, which is now 1/6 cycles per metre and results in the data set being 48 metres long. Therefore, the longest spatial feature that can be resolved in 24 metres.

- At each time, the spatial average was removed from the data, resulting in a data set of spatial anomalies. This is equivalent to removing a linear trend from a time series.

Testing:

- Stationarity: Results from the Dickey-Fuller test indicate Tau calculated values of -1.03, -0.88 and -1.16 while the Tau critical value is -2.99. The calculated Tau is greater than Tau critical values, hence we cannot reject the null hypothesis that the spatial data sets are not stationary.
- As each time step only contained 16 data points, we felt that this was insufficient to provide reliable statistics on independence normality and homogeneity of the means and variances. These tests were not undertaken, which is noted and we proceed with the spectral analysis.

Smoothing of the periodogram:

- No smoothing of the periodogram was undertaken, hence there are no confidence limits associated with these spectra.

Description (Figure 4–28):

- The largest spectral energy occurs at low wave number.
- There is slightly less energy in the first spectrum (at 19:20 hours) than the other two spectra. This may be expected as the data plot (Figure 4–27) indicates departure from the zero value to be considerably less than the data from the other two times.
- There are no spectral peaks of note in any of the three FFTs examined.

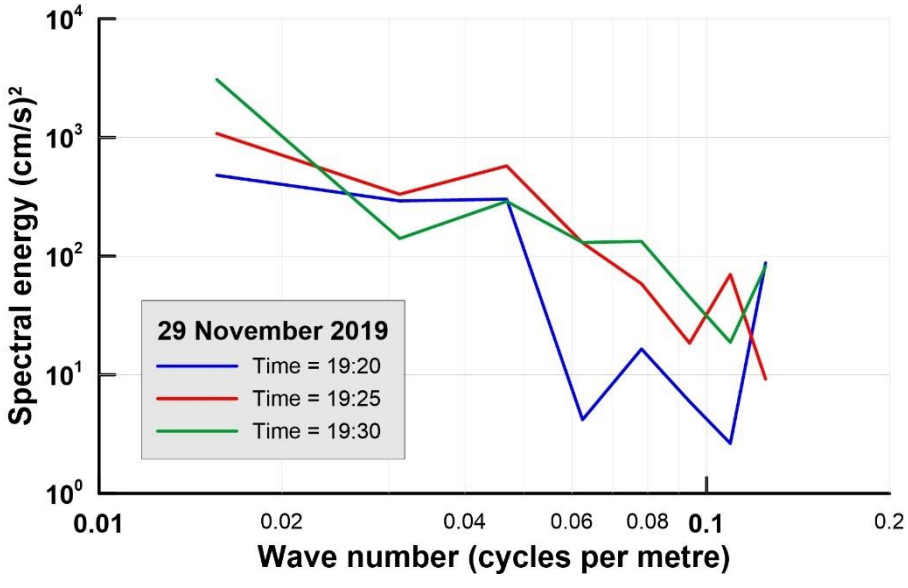


Figure 4–28. Three spatial FFTs for the east-west component of the currents, at times 19:20, 19:25 and 19:30 on 29 November 2019.

Answering our questions from above:

(a) The maximum spectral energy occurs at the smallest wave number (about 0.015 cycles/metre).
(b) The maximum values occur at the same wave number for all three FFTs examined.
(c) There may be some evidence to suggest that there is less spectral energy under the crest of the internal wave (at time 19:20) than under the trough of the internal wave (at time 19:30).

These results are broadly similar to those obtained from the temperature data set (Example 4-7, Section 4.4.1).

4.5. Summary

The examples above focused on water temperature and water level. The former has a single (in general) dominant signal, at the annual frequency. This provides a good introduction to what can be expected from spectral analysis. The latter has multiple signals, at the tidal frequencies. This illustrates the complexity (and ambiguity) that can arise when multiple signals interact. In particular, it is the length of the data set that governs how effectively signals of different frequencies can be identified.

- Clearly define the question that you want to answer. Note that you may need to change this question or it might lead to other questions that need subsequent analyses.
- If you're collecting your own data, think carefully about how you will analyse the data before setting up your instruments. Things to consider are the total length of the data set and the time interval between adjacent data points. Are both of these appropriate to answering your question?
- Plot your data. It is surprising how one can get a 'feel' for the data just by looking at it.
- In theory, the data should be independent. In practice this is unlikely to occur. The fact that there is likely be repetitive features in the data indicates that there likely will be some serial dependence of the data.
- Spectral analysis requires data to be stationary. The Dickey-Fuller test is one that can be used to test for stationarity. Often, we had a large number of data points in our time series. In such circumstances, use of the Dickey-Fuller test will usually result in rejection of the null hypothesis that the data are not stationary.
- Other tests for stationarity (strong or weak stationarity) often assume that the data are independent. Again, the statistical tests usually result in rejection of the null hypothesis when large data numbers are involved. Therefore, stationarity may not be achieved (or even exist) in a real time series. We need to note this, then progress with our analysis.

- Normality and homogeneity. Split the data into (about) 10 uniform sizes. Test for normality (e.g. chi-squared goodness of fit test). Test whether the variances in each of the split sets are from the same population (e.g. use an homogeneity of variance test such as Bartlett's test). Test whether the means in each of the split sets are from the same population (e.g. use an ANOVA test, or similar). Generally, we test at the 5% level of significance, although the reason for this is more historical rather than for anything definitive.
- Always remove the linear trend before you undertake a Fourier analysis. Depending on the question that you wish to answer, you may want to remove any known signals such the annual temperature signal or tidal signals. Removing signals that do not form part of your question will help highlight your features of interest.
- If you need to patch in gaps in your data set, try to keep the patching to less than about 5% of the total length of the data set.
- If you're going to use the FFT, keep any padding of the data set to less than about 5% of its total length.
- If you wish to taper the data set, restrict tapering to less than about 5% at each end of the data set. We're not big fans of tapering as it alters an otherwise perfectly good data set, so we rarely taper a data set.
- Average the spectral estimates; confidence limits can then be generated. Note, there are other ways to generate confidence limits, but we believe that this averaging (Daniell windowing) is the simplest.
- Run the spectral analysis at least three times: one without smoothing, one with much smoothing and one in between.
- We like to smooth differently in three different regions of the spectrum. A small amount of smoothing at the low frequency end of the spectrum, moderate smoothing in the middle of the spectrum and heavy smoothing at the high frequency end of the spectrum.

- We feel that spectral analysis in the time-frequency domain is a little easier to interpret than that from the space-wave number domain. Interpretation in the latter domain may become a little clearer when we examine harmonic analysis in Chapter 7.

Chapter 5. The Cross Spectrum

The previous chapters have dealt with auto-spectra – a single time series resulting in a single spectrum comprising an amplitude spectrum and a phase lag spectrum. But what if we have two time series? Can we use spectral analysis to look at the relationship between two time series? The short answer is 'yes', by using a technique called cross spectral analysis. However, as you can appreciate, the complexity increases substantially.

Our two time series might be two scalars or a pair of vector components. The only restriction we have is that our two time series must be of the same length. Some examples that we'll use include:

- Two sets of water temperature data from the same time period,
- Two sets of water temperature data from different time periods,
- One set of water temperature data and one set of water level data,
- One pair of vectors (2 vector components) representing, for example, the east-west and the north-south components of ocean currents.

One aspect of the cross spectrum that may cause confusion is the 'phase'. When we used the auto spectrum we generated a 'phase lag' spectrum, which basically gives us the starting point of the trigonometric curves at each frequency. With the cross spectrum, a 'phase difference' spectrum is generated. This is the angular difference between the two spectra at each frequency. We talk about one data set 'leading' or 'lagging' the other data set at a particular frequency.

A flowchart for generation of the cross spectrum is shown in Figure 5–1. Embedded in this flowchart are the auto spectra for each of the two data sets comprising the cross spectrum.

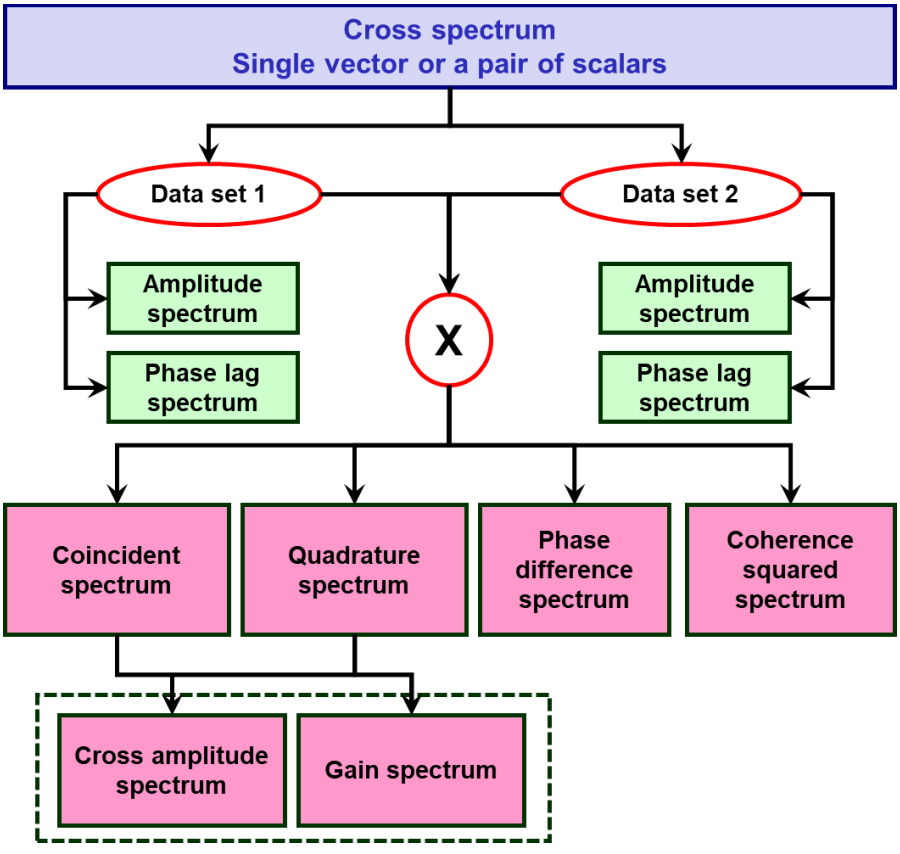


Figure 5–1. Flow chart of the cross spectrum. The 'X' indicates analyses involving both data sets – the 'cross' part of the cross spectrum. Cells shaded green indicate outcomes from the auto spectrum, while cells shaded pink indicate outcomes from the cross spectrum.

The preliminary procedures used for the auto spectrum (including tests for independence and stationarity, removing trends, generating confidence limits etc) also apply to the cross-spectrum.

5.1. Theory behind the cross spectrum

Let's suppose we have two time series x_t and y_t. The two auto spectra (with their linear trends removed) are given by:

$$\mathcal{F}_{x(p)} = \mathcal{F}_x = \frac{2}{N}\sum_{t=1}^{N} x_t \left[e^{i\frac{2\pi pt}{N}}\right] = a_p + ib_p \quad for \quad p = 1, 2, 3, \ldots . \left(\frac{N}{2}\right) - 1$$

$$\mathcal{F}_{y(p)} = \mathcal{F}_y = \frac{2}{N}\sum_{t=1}^{N} y_t \left[e^{i\frac{2\pi pt}{N}}\right] = c_p + id_p \quad for \quad p = 1, 2, 3, \ldots . \left(\frac{N}{2}\right) - 1$$

Where:

$\mathcal{F}_x$ and $\mathcal{F}_y$ are the Fourier transforms of the x and y data sets, respectively.

All of the information provided in the previous chapter also applies to these auto spectra.

We'll now introduce two new terms: the coincident-spectrum (CO spectrum) and the quadrature-spectrum (QU spectrum). Other authors may use different definitions to those given below, but they should only differ by a constant.

$$CO_p = \left[Re_p(\mathcal{F}_x)\, Re_p(\mathcal{F}_y) + Im_p(\mathcal{F}_x)\, Im_p(\mathcal{F}_y)\right] = \left[a_p\, c_p + b_p\, d_p\right]$$

$$for \quad p = 1, 2, 3, \ldots . \left(\frac{N}{2}\right) - 1$$

$$QU_p = \left[Re_p(\mathcal{F}_x)\, Im_p(\mathcal{F}_y) - Im_p(\mathcal{F}_x)\, Re_p(\mathcal{F}_y)\right] = \left[a_p\, d_p - b_p\, c_p\right]$$

$$for \quad p = 1, 2, 3, \ldots . \left(\frac{N}{2}\right) - 1$$

Where:

Re, Im are the real and imaginary components of the respective auto spectra, and

N is the number of data points in the time series.

The cross spectrum is then defined, for each value of p, as:

$$\mathcal{F}_{xy(p)} = CO_p - i\, QU_p$$

Now: $\mathcal{F}_{x(p)} = a_p + ib_p$ and $\mathcal{F}_{y(p)} = c_p + id_p$

Hence: $\mathcal{F}_{x(p)}\mathcal{F}^*_{y(p)} = (a_p + ib_p)(c_p - id_p)$

(where the superscripted asterisk represents the complex conjugate)

$$= (a_p\, c_p + b_p\, d_p) - i(a_p\, d_p - b_p\, c_p)$$

$$= CO_p - i\, QU_p$$

So, the cross spectrum can also be defined as the product of the first spectrum and the complex conjugate of the second spectrum.

We can define the cross amplitude spectrum for each value of p as (again, the superscripted asterisk indicates the complex conjugate):

$$XAmp_p = \sqrt{\mathcal{F}_{xy(p)}\, \mathcal{F}^*_{xy(p)}}$$

$$= \sqrt{(CO_p - i\, QU_p)(CO_p + i\, QU_p)}$$

$$\sqrt{CO_p^2 + QU_p^2}$$

Which can also be expressed as the 'gain' spectrum, defined as:

$$G_p = \sqrt{\frac{\mathcal{F}_{y(p)}\ CO_p}{\mathcal{F}_{x(p)}}}$$

One of the main outcomes of cross spectral analysis is to determine the correlation between the two data sets at each frequency. This 'correlation' is known as the coherence squared (Coh_p^2) spectrum and is given by:

$$Coh_p^2 = \frac{CO_p^2 + QU_p^2}{\mathcal{F}_{x(p)}\, \mathcal{F}^*_{x(p)} \times \mathcal{F}_{y(p)}\, \mathcal{F}^*_{y(p)}}$$

(where the superscripted asterisk indicates the complex conjugate).

$$= \frac{CO_p^2 + QU_p^2}{(a_p + ib_p)(a_p - ib_p) \times (c_p + id_p)(c_p - id_p)}$$

The coherence squared ranges between 0 and 1.

The phase difference is given by:

$$\phi_p = tan^{-1}\left(\frac{-QU_p}{CO_p}\right)$$

The cross-spectrum system can be uniquely specified by any of the following sets:

- Co, quadrature and coherence squared spectra,
- Cross-amplitude, coherence squared and phase difference spectra,
- Gain, coherence squared and phase difference spectra,
- X, Y and coherence squared spectra.

Our preference is to plot the X, Y, coherence squared and phase difference spectra. There is a level of redundancy in these sets of four spectra (the X and Y spectra can be combined to form the cross-amplitude spectrum or gain spectrum), but it enables us to study each auto spectrum as well as the correlation and phase relations between the two data sets.

There is an ambiguity in the sign of the quadrature spectrum. This is purely a definitional problem and is of no major concern. However, this is one reason that we do not like using the 'co' and 'quad' spectra for plotting.

If we expand the definition for the coherence squared spectrum we find:

$$Coh_p^2 = \frac{CO_p^2 + QU_p^2}{(a_p + ib_p)(a_p - ib_p) \times (c_p + id_p)(c_p - id_p)}$$

$$= \frac{(a_p c_p + b_p d_p)^2 + (a_p d_p - b_p c_p)^2}{(a_p + ib_p)(a_p - ib_p) \times (c_p + id_p)(c_p - id_p)}$$

$$= \frac{(a_p c_p + b_p d_p)^2 + (a_p d_p - b_p c_p)^2}{(a_p^2 + b_p^2) \times (c_p^2 + d_p^2)}$$

$$= \frac{a_p^2 c_p^2 + b_p^2 d_p^2 + 2a_p c_p b_p d_p + a_p^2 d_p^2 + b_p^2 c_p^2 - 2a_p d_p b_p c_p}{(a_p^2 + b_p^2) \times (c_p^2 + d_p^2)}$$

$$= \frac{a_p^2 c_p^2 + b_p^2 d_p^2 + a_p^2 d_p^2 + b_p^2 c_p^2}{a_p^2 c_p^2 + b_p^2 d_p^2 + a_p^2 d_p^2 + b_p^2 c_p^2}$$

$$= 1 \qquad for\ all \quad p$$

This is a trivial result. For the coherence squared spectrum to produce non-trivial results, we need to average several spectral estimates.

When we averaged m auto spectral results we could generate confidence limits. Averaging m cross spectral estimates allows us to determine the coherence squared spectrum and the phase difference spectrum as well as enabling the generation of confidence limits. So, the average co and quadrature spectra over m spectral bands become:

$$\langle CO_p \rangle = \frac{1}{m}\left\{\sum_{k=1}^{m} Re_{pk}(x) \sum_{k=1}^{m} Re_{pk}(y) + \sum_{k=1}^{m} Im_{pk}(x) \sum_{k=1}^{m} Im_{pk}(y)\right\}$$

$$= \frac{1}{m}\left\{\sum_{k=1}^{m} a_{pk} \sum_{k=1}^{m} c_{pk} + \sum_{k=1}^{m} b_{pk} \sum_{k=1}^{m} d_{pk}\right\}$$

$$\langle QU_p \rangle = \frac{1}{m}\left\{\sum_{k=1}^{m} Re_{pk}(x) \sum_{k=1}^{m} Im_{pk}(y) - \sum_{k=1}^{m} Im_{pk}(x) \sum_{k=1}^{m} Re_{pk}(y)\right\}$$

$$= \frac{1}{m}\left\{\sum_{k=1}^{m} a_{pk} \sum_{k=1}^{m} d_{pk} - \sum_{k=1}^{m} b_{pk} \sum_{k=1}^{m} c_{pk}\right\}$$

Where the angled brackets represent an average in the frequency domain.

The coherence squared spectrum now becomes:

$$\langle Coh_p \rangle^2 = \frac{\langle CO_p \rangle^2 + \langle QU_p \rangle^2}{\langle \mathcal{F}_x \mathcal{F}_x^* \rangle \times \langle \mathcal{F}_y \mathcal{F}_y^* \rangle}$$

And the phase difference spectrum becomes:

$$\langle \phi_p \rangle = tan^{-1}\left(\frac{-\langle QU_p \rangle}{\langle CO_p \rangle}\right)$$

If the coherence squared value at a particular value for $\langle p \rangle$ is small, then the phase difference spectrum at that frequency loses meaning. A small value for the coherence squared may occur when both the co and quadrature spectra are small. When this occurs, the argument in the expression for the phase difference

$$\phi_p = tan^{-1}\left(\frac{-QU_p}{CO_p}\right)$$

becomes 'small number divided by a small number' and the phase difference spectrum may become highly variable.

The average frequency over the m spectral bands is given by:

$$\frac{\langle p \rangle}{N\,\Delta t} = \frac{1}{N\,\Delta t}\ \frac{1}{m}\sum_{k=1}^{m} p_k \qquad for \quad p = 1, 2, 3, \ldots . \left(\frac{N}{2}\right) - 1$$



The $(1-\alpha)100\%$ confidence limit for the coherence squared spectrum is given by:

$$Conf\ Lim\ =\ 1-\alpha^{\left(\frac{1}{m-1}\right)}$$

If we choose $\alpha = 0.05$ (this is a generally accepted, but arbitrary, value) and average over $m = 4$ spectral bands, then the 95% confidence limit is given by:

$$Conf\ Lim\ =\ 1-0.05^{\left(\frac{1}{4-1}\right)} = 1-0.05^{\left(\frac{1}{3}\right)} = 0.632$$

Confidence limits can also be generated for the phase difference estimates, although they are not often quoted. The $(1-\alpha)100\%$ confidence limit for the phase difference spectrum (at a specific frequency) is given by:

$$|Conf\ Lim|\ =\ sin^{-1}\left[\left(\frac{1-Coh^2}{\{2m-2\}\,Coh^2}\right)t_{\alpha,2m-2}\right]$$

Where:

m is the number of spectral bands that are averaged ($m > 1$), and

$t_{\alpha,2m-2}$ is student's t distribution, with $2m-2$ degrees of freedom.

It is important to note that the confidence limit for the phase difference depends on the coherence squared. As the coherence squared appears in the denominator, the confidence limit for the phase difference becomes ambiguous when the coherence squared is small.

If we choose $\alpha = 0.05$ and average over $m = 4$ spectral bands, and the coherence squared is 0.9, then the 95% confidence limit for the phase difference is given by:

$$|Conf\ Lim|\ =\ sin^{-1}\left[\left(\frac{1-0.9}{\{8-2\}\,0.9}\right)t_{0.05,6}\right] = sin^{-1}\left[\left(\frac{0.1}{5.4}\right)1.94\right] = \pm 2^o$$

If the coherence squared falls to 0.5, then the confidence limits are: $\pm 20^o$. For the phase difference confidence limits to be in the range $\pm 10^o$ (which is a reasonable value for uncertainty in the phase) then the coherence squared needs to exceed 0.67 when 4 spectral estimates are averaged.

The confidence limits for the phase difference spectra in the examples below are not presented, mainly because their inclusion makes the phase difference spectrum plot difficult to visualise. Instead, we use the rule of thumb that, when 2, 4, 8 and 12 spectral estimates are averaged, we are 95% confident that the phase difference lies in the range $\pm 10^o$, provided the coherence squared values exceed 0.90, 0.65, 0.40 and 0.30, respectively.

5.2. Examples of the cross spectrum

The examples provided in this chapter follow the same format as in the previous chapter.

5.2.1. Example 5-1: Water temperature at two levels – 5 min data

In this example we're interested in the high frequency relationship between two sets of water temperature data. As we're interested in the high frequency end of the spectrum, we don't need a long data set, but we do need one that is sampled frequently. In this case we have chosen 5 minute data.

Questions: What does the cross spectrum of high frequency water temperature collected from two locations separated vertically by 4m show? At what frequencies are the two data sets most highly correlated? Given the spatial separation of the two temperature sensors is relatively small, we might expect the correlation (coherence squared) to be high and the phase difference to be close to zero.

Data used:

- 5 minute water temperature collected from two locations separated vertically by 4m.

- The Nyquist period is 10 minutes (Nyquist frequency is 1/10 cycles per minute, or 6 cycles per hour).
- The data set is about 14 days long (4,096 data points), therefore the longest period features we can realistically interpret is about 1-2 days (i.e. we want about 5-10 cycles in the data set before we are confident of interpreting the low frequency end of the spectrum).

Pre-processing:

- No pre-processing of the data was undertaking.
- The linear trend was removed from the data producing two sets of temperature anomalies.
- The time series plots of the temperature anomalies are shown in Figure 5–2 (note, data in red have been offset by 2°C on the plot). As might be expected (as the data are only separated vertically by 4m), many of the features in the two plots are similar i.e. the same features appear in both plots and they occur at the same time.

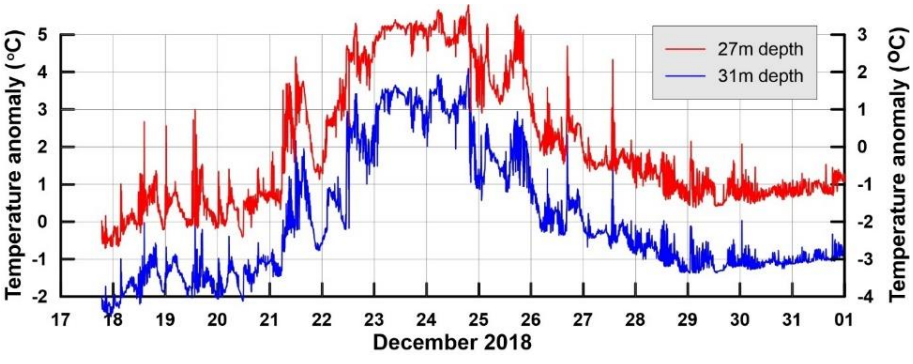


Figure 5–2. Time series of water temperature anomalies at two levels in the water column (at 27m depth and 31m depth). Note the vertical axis on the right has been offset by 2°C to separate and clarify the two plots.

Testing:

- Stationarity: Results from the Dickey-Fuller test indicate Tau calculated values of -5.53 for both sets of data, the Tau critical value is -2.86. The calculated Tau is less than Tau critical values, hence we reject the null hypothesis that the data sets are not stationary.
- Based on the relevant statistical tests, the data are not normally distributed at the 5% level of significance, nor are the (approximate) stationarity conditions (homogeneity of means and variances) satisfied. Noting this, we proceed.

Smoothing of the periodogram:

- Daniell windowing is used to smooth the periodogram into frequency bands of size 2 (for frequencies lower than 7 cycles per day), 4 (for frequencies between 7 and 20 cycles per day) and 8 (for frequencies greater than 20 cycles per day).
- The 95% confidence limits are presented for the energy spectra and the coherence squared spectrum.

Description: The resulting energy, coherence squared and phase difference spectra are presented in Figure 5–3 (the bottom, middle and top plots, respectively). The 95% confidence limits for the spectrum and coherence squared are superimposed on the respective plots, as is the '-5/3^{rds} power law'. No confidence limits for the phase difference spectrum are presented.

- No discernible difference between the two spectra (Figure 5–3, bottom plot).
- Most of the energy occurs at low frequency, despite the trend being removed. However, the time series shows a major low frequency signal (high temperature anomaly in the middle of the time series and low temperature anomalies at either end). This is the likely cause of most of the variance in the data sets (or energy in the periodograms).

- There are no significant peaks in the spectra (recall that statistically significant peaks are those whose heights exceed the distance between the two confidence limits).
- However, small spectral peaks can be observed at frequencies of 1 and 2 cycles per day. These peaks may correspond to elevated energy from the diurnal and semi-diurnal internal tides.
- The middle of the energy spectra closely follows the '$-5/3^{rds}$ power law'.
- At frequencies exceeding about 30 cycles per day, there is a slight flattening of the spectrum i.e. it departs from the '$-5/3^{rds}$ power law'. This suggests a high frequency source of energy, possibly from high frequency internal waves.

The coherence squared spectrum is shown in Figure 5–3, middle plot.

- Much of the coherence squared values lie above the 95% confidence limit – particularly at higher frequencies. This is, perhaps, a bit unusual as it is more common that the coherence squared is greater at low frequencies than at high frequencies. The two sets of temperature data are only separated by 4 m, so we might expect the coherence between them to be relatively high across the whole spectrum.
- Between frequencies of 7 and 20 cycles per day the coherence squared is, generally, 0.63 or greater. In this region the confidence limits for the phase difference spectrum lie between $\pm 10^{o}$.

The phase difference spectrum is shown in Figure 5–3, top plot.

- As anticipated (because of the geographical proximity of the two data sets), the phase difference spectrum is generally close to zero. The exceptions below are minor and, perhaps, not of great importance.

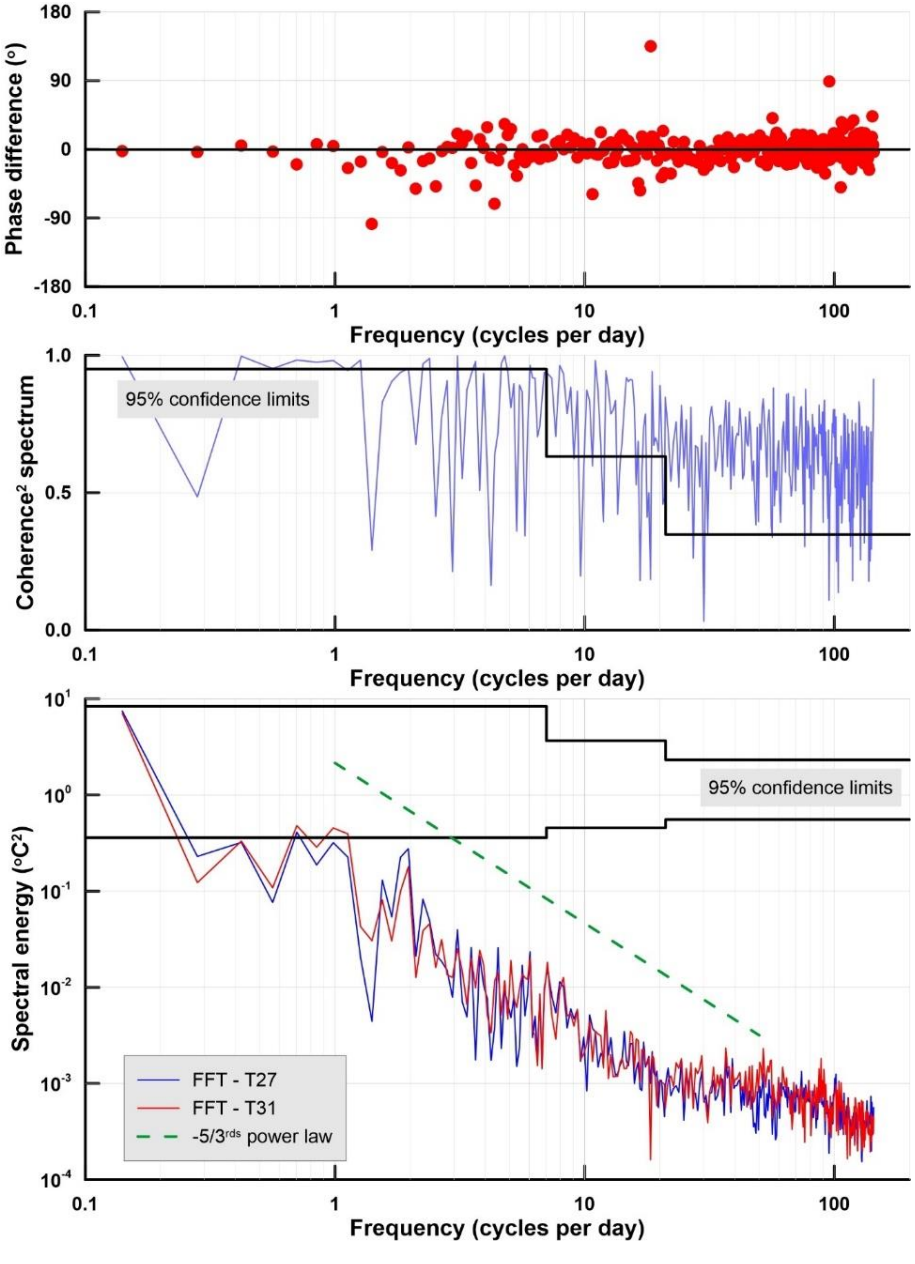


Figure 5–3. Periodogram, coherence squared and phase difference spectra for water temperature at two levels in the water column (at 27m depth and 31m depth).

- One exception is near a frequency of 1.5 cycles per day, where the phase difference is about -90^o. This means that the first data set (i.e. the data from 27m depth) lags behind the second data set (i.e. the data from 31m depth) by about one quarter of its cycle. A frequency of 1.5 cycles per day translates to a period of 16 hours. One quarter of 16 hours is 4 hours, so for temperature features that repeat themselves every 16 hours, the data at 27m depth appear (on average) 4 hours after they are seen in the data at 31m depth. This is a bit unusual for two data sets that are so close.
- However, if we refer to the coherence squared spectrum at the same frequency, we find that it is very low, about 0.25. Low coherence squared means the phase difference value is questionable. In this case the confidence limits on the phase difference are about $\pm 80^o$.
- The coherence squared is also small for other values of the phase difference spectrum that are substantial departures from zero. Such values of the phase difference spectrum may be unreliable.

Answering our questions:

(a) Visually, there is little difference in the two energy spectra of the two data sets. The coherence between the two data sets is relatively high across the whole spectrum. The phase difference is close to zero across the whole spectrum.
(b) The two data sets are highly correlated across much of the spectrum. The coherence squared appears to be relatively low (i.e. below the 95% confidence limit) in the frequency band 2-7 cycles per day.

These results are expected and understandable, as the two sets of temperature data are only separated by 4 m. Intuitively, we'd expect data from nearby locations to be closely related.

5.2.2. Example 5-2: Water temperature at two levels – monthly data

This example examines the very low frequency end of the spectrum. We're interested in the temperature relationships between two large scale water bodies, viz. the Coral Sea and the Tasman Sea. As our interest is in the very low frequency end of the spectrum, we need to have data sets over a long period of time, but not necessarily sampled frequently.

Questions: How strongly correlated are the temperature signals from long term water data collected from the Coral and Tasman Seas? At what frequencies do these strong correlations occur?

Data used:

- Monthly sea surface temperature data from the Coral and Tasman Seas.
- The Nyquist period is 2 months (Nyquist frequency of ½ cycles per month or 6 cycles per year).
- The data set is about 85 years long (1,024 data points), therefore the longest period features we can realistically interpret is about 10-20 years (i.e. we want about 5-10 cycles in the data set before we are confident of interpreting the low frequency end of the spectrum).

Pre-processing:

- The linear trend, together with the annual and semi-annular signals were removed from both data sets, results in sets of temperature anomalies.
- The time series plots are shown in Figure 5–4. Visually, a strong correlation between the two data sets is not obvious. However, it does appear that the data from the Tasman Sea are more variable than that from the Coral Sea.

- As we're examining water temperatures, we intuitively know that there will be a strong correlation between the two data sets at the annual period. We remove that signal as its strength is likely mast any other low frequency features in the data set.

Testing:

- Stationarity: Results from the Dickey-Fuller test indicate Tau calculated values of -10.81 and -10.35, the Tau critical value is -2.86. The calculated Tau is less than Tau critical values, hence we reject the null hypothesis that the data sets are not stationary.

Although the data sets are from two different seas (i.e. different water bodies) they are adjacent and the time step (one month) is relatively large. Therefore, it may be expected that the calculated Tau values are close.

- Based on the relevant statistical tests, the both data sets are normally distributed at the 5% level of significance. However, the conditions for weak stationarity (homogeneity of means and variances) are not satisfied. Noting this, we proceed.

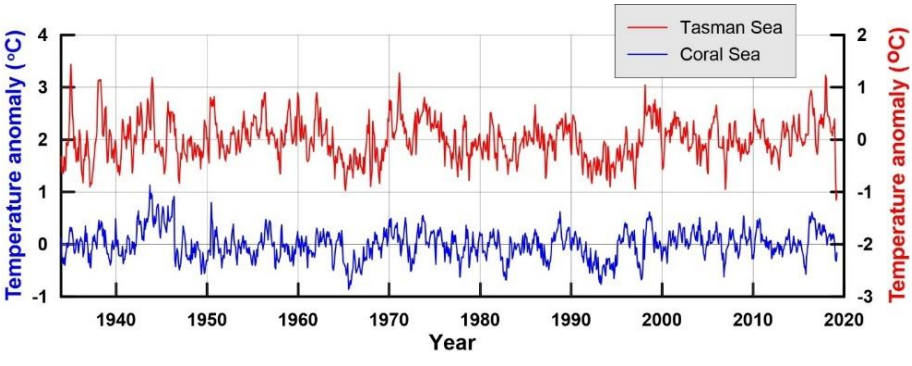


Figure 5–4. Time series of SST anomalies in the Coral and Tasman Seas. Note the vertical axis on the right has been offset by 2°C to separate and clarify the two plots.

Smoothing of the periodogram:

- Daniell windowing is used to smooth the periodogram into frequency bands of size 2 (for frequencies lower than 0.5 cycles per year), 4 (for frequencies between 0.5 and 3 cycles per year) and 8 (for frequencies greater than 3 cycles per year).

- The averaging (or smoothing) of the periodogram is used to generate the 95% confidence limits, which are presented for the energy spectra and the coherence squared spectrum. Again, no confidence limits for the phase difference spectrum have been plotted.

Description: The resulting energy, coherence and phase difference spectra are presented in Figure 5–5 (the bottom, middle and top plots, respectively). The 95% confidence limits for the spectrum and coherence squared are superimposed on the respective plots, as is the '$-5/3^{rds}$ power law'.

The spectral energies for the two data sets are similar (Figure 5–5, bottom plot).

- Both energy spectra are relatively flat for frequencies less than about 0.4 cycles per year. These frequencies are outside the inertial subrange region, hence the $-5/3^{rds}$ power law does not apply.
- Most of the energy occurs at low frequency, despite the trend, annual and semi-annual signals being removed. Low frequency signals still remain in the data sets.
- There are no significant peaks in either spectra.
- There is greater spectral energy in the Tasman Sea data set.
- Frequencies exceeding 0.4 cycles per year follow the '$-5/3^{rds}$ power law'.
- The Nyquist frequency is 6 cycles per year. Even at this highest frequency, there is no substantial departure from the '$-5/3^{rds}$ power law'.

The coherence squared spectrum is shown in Figure 5–5, middle plot.

- There are only a few regions in the spectrum where the coherence squared values lie above the 95% confidence limit. This suggests that, after removing the trend, annual and semi-annual signals, the correlation between the two data sets is relatively weak.

The phase difference spectrum is shown in Figure 5–5, top plot.

- The phase different spectrum is highly variable, even at low frequencies. As the coherence squared values are generally below the 95% confidence limits, the phase difference spectrum loses meaning and it provides only limited information.

Returning to the questions we have asked:

(a) When the linear trend, annual and semi-annual signals are removed from both data sets, there is limited coherence between the two data sets.
(b) High coherence occurs near 1.2, 0.5, 0.75, 1.8 cycles per year (and at a few higher frequencies). It is unclear whether these frequencies correspond to important processes. Their appearance in the spectral may suggest further investigation.

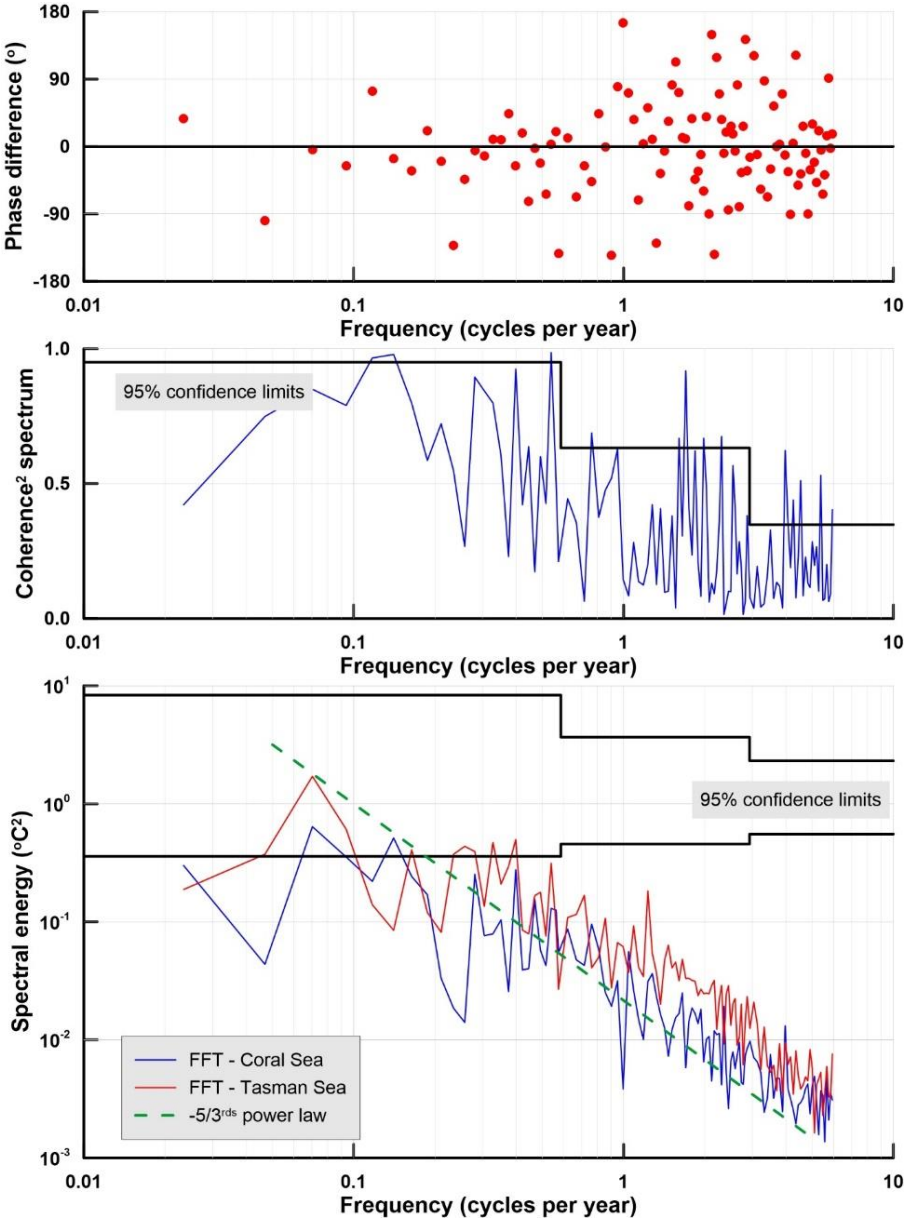


Figure 5–5. Periodogram, coherence squared and phase difference spectra for Coral and Tasman Sea temperature anomalies.

5.2.3. Example 5-3: Water temperature at two levels – hourly data

Here we examine the relationship in water temperature between two sets of data separated vertically by 50 m. Let's assume that we're only interested in the frequency range 1 cycle per day to 1 cycle per month. Therefore, to effectively interpret the low frequency end of the spectrum we need a data set that is at least 6-12 months long (to ensure we have 5-10 cycles in the data) and sampled at least twice per day (but preferably about every 2 hours to ensure we 5-10 cycles at the highest frequency we wish to resolve).

The two data sets that we'll use are 12 years long and sampled hourly. This more than satisfies the limiting frequencies of our interest and we could analyse the whole data set, as was done for the previous two examples. However, we could also split the data sets into 12 groups each of 12 months duration and average the 12 results to produce the cross spectrum components. One advantage in this approach is that we don't lose any frequency resolution during the averaging process. (Recall that the previous two examples averaged across 2, 4 and 8 spectral estimates in different frequency regions and in so doing, the frequency resolution in each of those regions was reduced).

Questions: How strongly correlated are the temperature signals between 1 cycle per day and 1 cycle per month using data collected from locations that are vertically separated by 50 m?

Data used:

- Hourly water temperature data offshore collected over 12 years (>105,000 data points).
- The Nyquist period is 2 hours (Nyquist frequency of ½ cycles per hour or 12 cycles per day).
- The data set is about 12 years long, therefore the longest period features we can realistically interpret is about 1-2 years.

To generate the confidence limits, we could undertake a FFT on the whole 12 years of hourly data then band average. However, our interest in this example concerns temperature signals in the range 1 cycle per day

to 1 cycle per month. This then provides an option to split the data into 12 x 12 monthly sets and average the corresponding spectral estimates to generate confidence limits. In so doing, the frequency resolution in the periodogram remains unchanged, however, the features with the longest period that can be effectively interpreted have frequencies of about 1-2 months. This still allows us to answer the question posed above.

Pre-processing:

- The linear trends were removed from both data sets to produce two sets of temperature anomalies.
- The time series plots of the temperature anomalies (for one year only) are shown in Figure 5–6. Visually, there is little correlation between the two data sets, although some features manifest themselves in both plots. The annual signal is clear at 51m (T51) above the sea floor, perhaps less so for the temperature data collected from 1m (T01) above the sea floor.

Testing:

- Stationarity: Results from the Dickey-Fuller test indicate Tau calculated values of -2.71 (for the temperature anomalies T01) and -7.10 (for the temperature anomalies T51), the Tau critical value is - 2.86. The calculated Tau is less than Tau critical for T51 but not for T01. In this case the T01 data set is 'close' to being stationary and we'll assume that it is stationary for this example.
- Based on the relevant statistical tests, the data are not normally distributed at the 5% level of significance. The conditions for weak stationarity conditions (homogeneity of means and variances) are not satisfied. Noting this, we proceed.

Smoothing of the periodogram:

- Twelve periodograms were generated, one for each year of available data.
- At each frequency, the 12 data sets were averaged enabling the confidence limits to be generated without reducing the resolution

of the frequency. Such averaging results in the same confidence limits across the entire spectrum.

- The 95% confidence limits are used for both the energy spectra and the coherence squared spectrum.
- No confidence limits for the phase difference spectrum are presented.

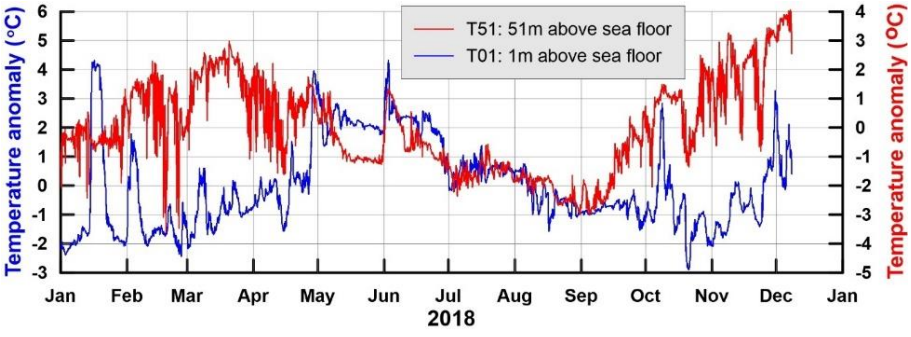


Figure 5–6. Time series of water temperature anomalies (for 2018 only) separated vertically by 50m. Note the vertical axis on the right has been offset by 2°C to separate and clarify the two sets of temperature anomalies.

Description: The resulting energy, coherence and phase difference spectra are presented in Figure 5–7 (the bottom, middle and top plots, respectively). The 95% confidence limits for the spectrum and coherence squared are superimposed on the respective plots, as is the '-5/3^{rds} power law'.

The spectral energies for the two data sets are similar (Figure 5–7, bottom plot).

- Both energy spectra closely follow the -5/3rds power law across most of the spectrum.
- Most of the energy is contained at the low frequency end of the spectrum. This should be expected as the annual signal was not removed from any of the data sets. At low frequencies, there is little difference between the two spectra.
- The spectrum at 51m above the sea floor contains peaks near the diurnal and semi-diurnal periods. The size of the former peak

exceeds the distance between the two 95% confidence limits. These sharp peaks are not apparent in the spectrum of the temperature anomaly at 1m above the sea floor.

- There are no other significant peaks in either spectra.
- At frequencies exceeding about 0.2 cycles per day, there is a separation between the two spectra. This separation distance becomes greater as the frequency increases. Solar thermal energy enters the water at the surface and becomes less effective as the depth increases. This may be an area of work for further investigation.

The coherence squared spectrum is shown in Figure 5–7, middle plot.

- There is high coherence between the two data sets at the lowest frequencies. This may be the result of part of the annual signal remaining in the data sets. We split the data into 12 annual sets (actually each data set was about 11.3 months long). As there is only one annual curve in each data set, the annual signal cannot be properly interpreted and this strong signal will appear as a long period feature – high energy at the low frequency end of the spectrum.
- At higher frequencies, the coherence squared spectrum is highly variable, generally falling below the 95% confidence limit. However, there are many spikes in the coherence squared spectrum that lie above this confidence limit. This makes interpreting the coherence squared spectrum quite difficult. Generally, at frequencies greater than about 0.01 cycles per day (i.e. for periods less than about 100 days) the coherence squared is not significant.
- This spiking in the coherence squared spectrum may suggest that spectral analysis may not be an appropriate tool to answer this question. It may be better suited to wavelet analysis that does not rely on stationarity.

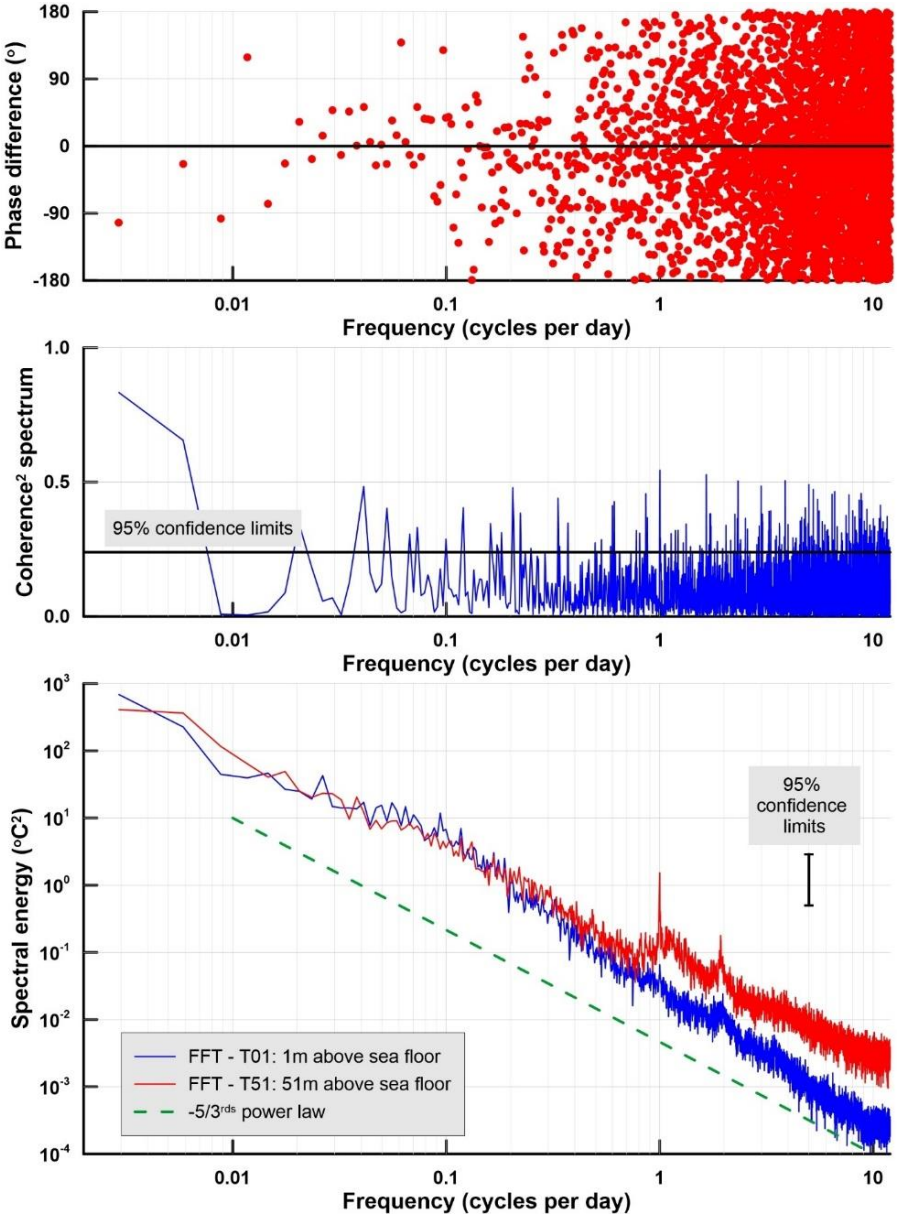


Figure 5–7. Periodogram, coherence squared and phase difference spectra for Coral and Tasman Sea temperature anomalies.

The phase difference spectrum is shown in Figure 5–7, top plot.

- The phase difference spectrum is quite variable, even at low frequencies. As the coherence squared values are generally below the 95% confidence limits, the phase difference spectrum loses meaning and it provides only limited information.
- At the lowest frequency (the annual period) the phase difference spectrum read about -90^o. That is, at the annual period, the signal in the data at T51 leads the signal in the data at T01 by about 3 months. (So, where did the '3 months' come from? 90^o is one quarter of a circle, and the frequency we're dealing with is 1 cycle per year or 0.003 cycles per day. One quarter of a year is 3 months. The negative sign indicates that the first set we analysed, the T01 data, lags behind the second set we analysed. If we'd analysed the data sets the other way around, the sign would have been positive).
- At a frequency of 0.006 cycles per day (period of about 6 months), there is little difference in the phase. Features with a period of about 6 months appear in the data at about the same time.

Returning to the questions we asked.

(a) Our question of interest concerned the correlation of the two data sets at frequencies between 1 cycle per day and 1 cycle per month. Overall, there appears to be little correlation between the two data sets between these two frequencies, although there are particular frequencies in this band that do show high correlation. These episodic features may trigger further investigations.

5.2.4. Example 5-4: Water level in two oceans – hourly data (Part 1)

In Example 5-4 we examine the relationship in water level collected from two locations – one from the South Pacific Ocean and one from the Indian Ocean (Southern Hemisphere). Given the large geographical separation of the two locations we might expect the results to substantially differ in both amplitude and phase.

We're interested in the frequency range 2 cycles per day to 1 cycle per month. Therefore, to effectively interpret the low frequency end of the spectrum we need a data set that is at least 6-12 months (to ensure we have 5-10 cycles in the data) and sampled at least 4 times per day (the Nyquist frequency, but preferable once per hour to ensure we have good resolution at our high frequency region of interest). The two data sets that we're using are about 5½ months long and sampled hourly. Hourly sampling (24 samples per day) will satisfy the Nyquist frequency limitations. However, with only 5½ months of data we may be at the limit of effectively resolving the signal at the monthly period.

Questions: How strongly correlated are the water levels at two locations – one in the Pacific Ocean and one in the Indian Ocean – at frequencies between 2 cycles per day and 1 cycle per month?

Data used:

- Hourly water level data collected over 170 days (4,096 data points).
- The Nyquist period is 2 hours (Nyquist frequency ½ cycle per hour).
- The data set is about 170 days long, therefore the longest period features we can realistically interpret is about 2-4 weeks. Reiterating, this may be at the limit for resolving signals at the monthly period. Nevertheless, we'll proceed and see what eventuates.

Pre-processing:

- The linear trends were removed from both data sets.

- The time series plots are shown in Figure 5–8. Visually, data from Pacific Ocean (blue plot) have a range of about 2m, while the range of the Indian Ocean data (red plot) is about 1m.
- Both plots show clear signals near the diurnal / semi diurnal periods.
- A clear spring / neap tidal signal is also clear in the data from both locations.

Testing:

- Stationarity: Results from the Dickey-Fuller test indicate Tau calculated values of -15.50 (for the Pacific Ocean water level data) and -8.88 (for the Indian Ocean water level data); the Tau critical value is - 2.86. As the calculated Tau is less than Tau critical for both data sets, we can reject the null hypothesis that the data sets are not stationary.
- Based on the relevant statistical tests, the data are not normally distributed at the 5% level of significance, nor are the conditions for homogeneity of means or variances met. The conditions for weak stationarity conditions are not satisfied. Noting this, we proceed.

Smoothing of the periodogram:

- Both periodograms were averaged: 2 spectral bands were averaged over the first 100 spectral estimates, 4 spectral bands over the next 200 spectral estimates and 8 spectral bands were averaged for the remainder of both spectra.
- The frequency separating the regions where 2 and 4 spectral bands were averaged was 0.59 cycles per day and the frequency separating the regions where 4 and 8 spectral bands were averaged occurred at 1.76 cycles per day.
- Spectral averaging enabled confidence limits to be generated. The 95% confidence limits are used for both the energy spectra and the coherence squared spectrum.
- No confidence limits for the phase difference spectrum are presented.

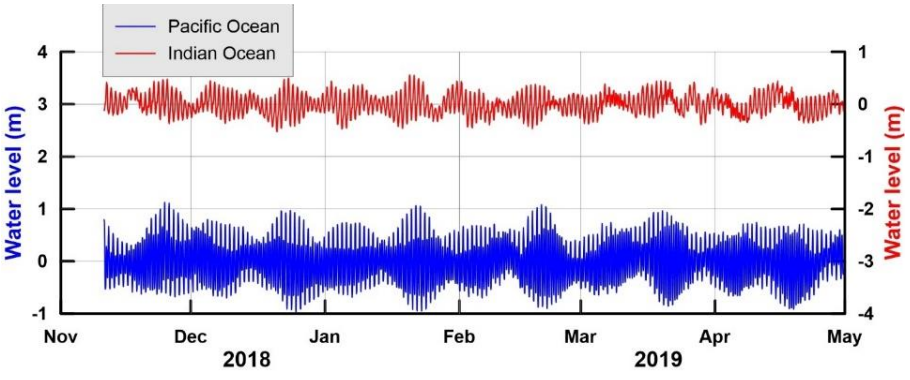


Figure 5–8. Time series of water levels (with trends removed) in the Pacific Ocean (blue) and Indian Ocean (red). Note the vertical axis on the right has been offset by 3m to separate and clarify the two plots.

Description: The resulting spectral energy, coherence and phase difference spectra are presented in Figure 5–9 (the bottom, middle and top plots, respectively). The 95% confidence limits for the spectrum and coherence squared are superimposed on the respective plots, as is the '-5/3^{rds} power law'.

The spectral energy for the two data sets are similar (Figure 5–9, bottom plot).

- Both energy spectra closely follow the -5/3^{rds} power law across much of the frequency range.
- Spectral energy between the frequencies of 1 cycle per month and 2 cycles per day represent about 98% and 96% of the total spectral energy for the Pacific Ocean and Indian Ocean water level data, respectively.
- The spectral energy showed statistically significant peaks near frequencies of 1, 2, 3 and 4 cycles per day.
- The dominant signal in the Pacific Ocean data occurred near 2 cycles per day (i.e. the semi-diurnal signal dominates). This accounts for about 83% of the total spectral energy. The diurnal signal contributes about 11% to the total spectral energy.

- In contrast, the dominant signal in the Indian Ocean data occurs near 1 cycle per day (the diurnal signal dominates). This represents about 64% of the total spectral energy, while the semi-diurnal spectral energy only accounted for about 7% of the total spectral energy.
- There were no significant spectral peaks at frequencies less than about 1 cycle per day. This is, perhaps, a little surprising as the spring / neap signal is quite clear in the time series plots (Figure 5–8). It is possible that the data set is not sufficiently long to effectively interpret this signal, or that this signal is non-stationary – it changes over time.
- There is slightly higher spectral energy in the Pacific Ocean water level data. This is reflected in the corresponding variances.

The coherence squared spectrum is shown in Figure 5–9, middle plot.

- There is high coherence squared between the two data sets in broad spectral bands near frequencies of 1 and 2 cycles per day, corresponding to the diurnal and semi-diurnal tides.
- Areas of high coherence squared are also observed at frequencies higher than 2 cycles per day. (Although this area of the spectrum does not form part of our original question).
- There is also high coherence squared near 1.4 cycles per day. The reason for this is unclear and it may warrant further investigation.

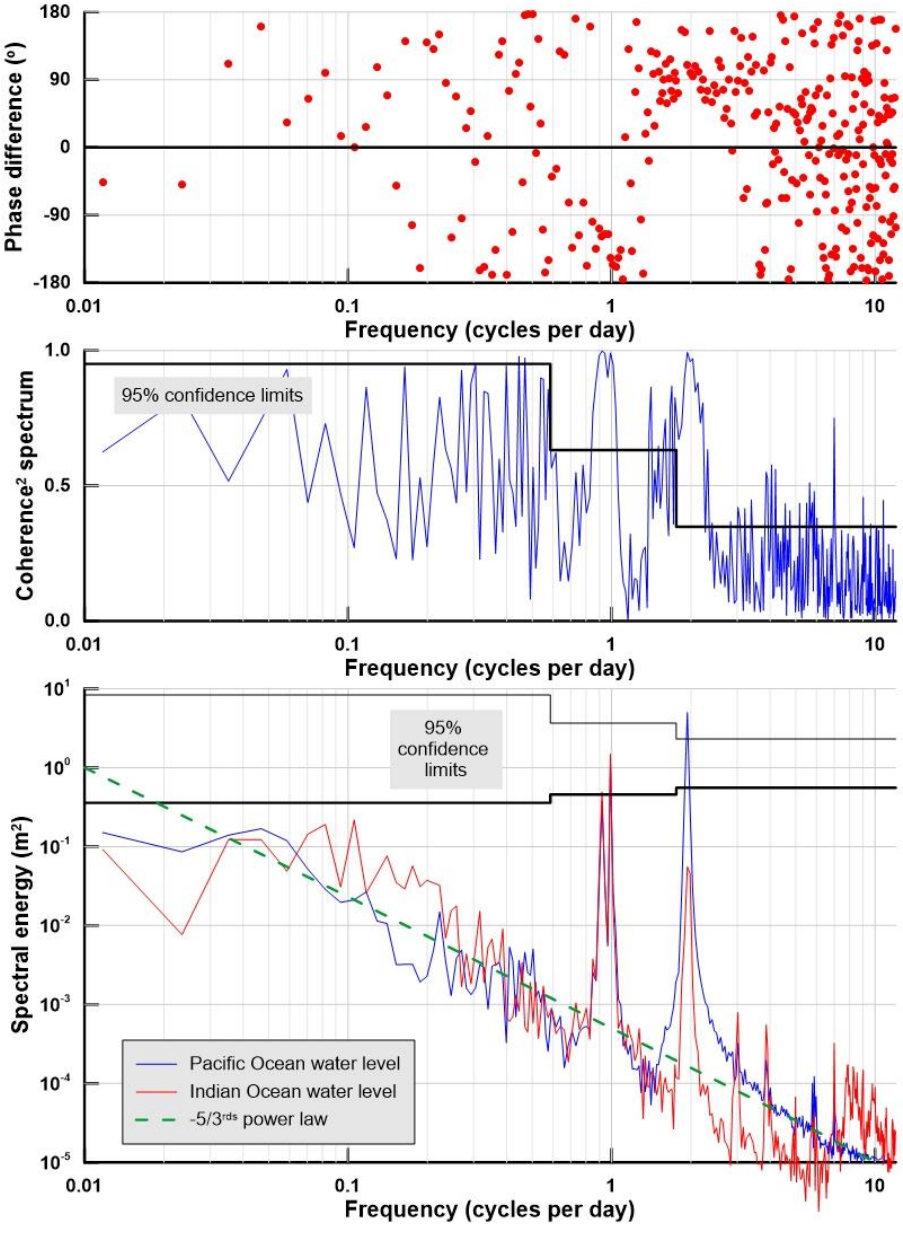


Figure 5–9. Periodogram, coherence squared and phase difference spectra for Pacific Ocean and Indian Ocean water levels.

The phase difference spectrum is shown in Figure 5–9, top plot.

- The phase difference spectrum near between the diurnal and semi-diurnal frequencies contains a cluster of points near -135°. At the diurnal frequency, this represents a lag of about 9 hours ($\frac{135^o}{360^o} \times 24\ hours = 9\ hours$) The negative sign indicates that the Pacific Ocean data lags the Indian Ocean data. However, given that trigonometric functions are cyclic, this may also translate to the Pacific Ocean water level leading the Indian Ocean water level by $(24 - 9 =)$ 15 hours. A more detailed analysis would need to be undertaken to confirm this assertion.
- Near the semi-diurnal frequency, the phase difference is about +90°. That is, the Pacific Ocean data leads the Indian Ocean data by about ($\frac{90^o}{360^o} \times 12\ hours = 3\ hours$).
- At the lower frequencies, the coherence squared spectrum was below the 95% confidence limit, hence the phase difference spectrum lacks meaning.

Returning to our question of interest:

(a) Our question of interest concerned the correlation of the two data sets at frequencies between 2 cycles per day and 1 cycle per month. There are significant spectral peaks near 1 and 2 cycles per day. Between 1 cycle per day and 1 cycle per month, both spectra follow the -5/3rds power law, suggesting no energy inputs to the data between these two sets of water level anomalies. There is high coherence between the two data sets at 1 and 2 cycles per day. At frequencies less than about 1 cycle per day, the coherence is generally below the 95% confidence limit. At the diurnal period the Pacific Ocean data leads the Indian Ocean data by about 225° or 15 hours. At the semi-diurnal period the Pacific Ocean data leads the Indian Ocean data by about 90° or 3 hours.

5.2.5. Example 5-5: Water level in two oceans – hourly data (Part 2)

Example 5-5 is a repeat of Example 5-4, except that we use a different Indian Ocean data set, one in a macro-tidal environment i.e. the tidal range is of the order 10m. The expectations from the previous example remain, although we also expect substantially more energy in the macro-tidal data set of the Indian Ocean than is observed in the Pacific Ocean data.

Focus will lie in the frequency range 1 to 6 cycles per day. This is a slightly different question to the one posed for Example 5-4. To effectively interpret the low frequency end of the spectrum we need a data set that is at least 5-10 days in duration (to ensure we have 5-10 cycles in the data) and sampled at least 12 times per day. The two data sets that we're using are about 5½ months long and sampled hourly. The hourly sampling (24 samples per day) will satisfy the Nyquist frequency limitations. With 5½ months of data we can easily resolve the low frequency end of our region of interest.

Questions: How strongly correlated are the water levels at two locations – one in the Pacific Ocean and one in the Indian Ocean – at frequencies between 1 and 6 cycles per day.

Data used:

- Hourly water level data collected over 170 days (4,096 data points).
- The Nyquist period 2 hours (Nyquist frequency of ½ cycle per hour or 12 cycles per day), about double that required to resolve the highest frequency in which we're interested (viz. 6 cycles per day).
- The data set is about 170 days long, therefore the longest period features we can realistically interpret is about 2-4 weeks – ample time to effectively interpret features at 1 cycle per day.

Pre-processing:

- The linear trends were removed from both data sets. The time series anomalies are shown in Figure 5–10.
- Visually, data from Pacific Ocean (blue plot) have a range of about 2m, while the range of the Indian Ocean data (red plot) is about 10m.
- Both plots show clear signals near the diurnal / semi diurnal periods.
- A clear spring / neap tidal signal is also clear in the data from both locations.

Testing:

- Stationarity: Results from the Dickey-Fuller test indicate Tau calculated values of -15.50 (for the Pacific Ocean water level data) and -16.78 (for the Indian Ocean water level data); the Tau critical value is - 2.86. As the calculated Tau is less than Tau critical for both data sets, we can reject the null hypothesis that the data sets are not stationary.
- Based on the relevant statistical tests, the data are not normally distributed at the 5% level of significance, nor were the conditions for homogeneity of means and variances satisfied. Hence, the conditions for weak stationarity are not satisfied. Again, we note this and proceed.

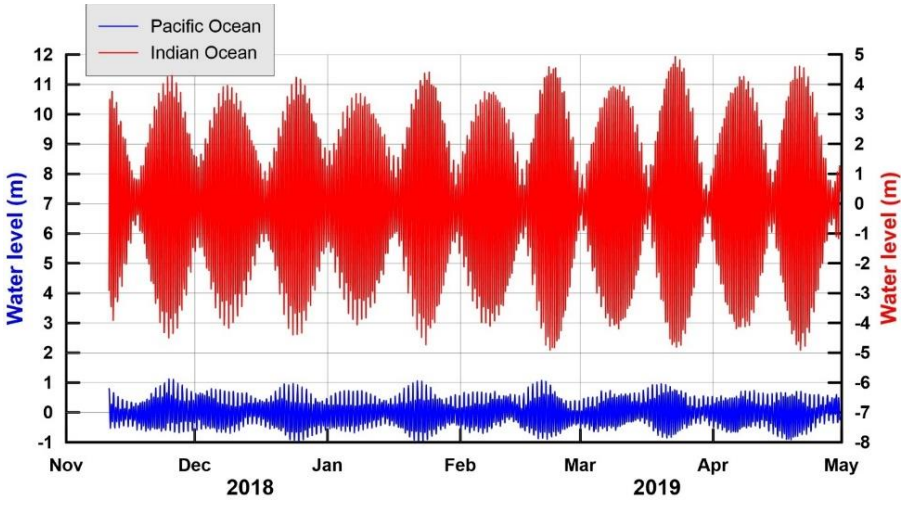


Figure 5–10. Time series of water levels (with trends removed) in the Pacific Ocean (blue) and Indian Ocean (red). Note the vertical axis on the right has been offset by 7m to separate and clarify the two plots.

Smoothing of the periodogram:

- Both periodograms were averaged as follows: 2 spectral bands were averaged over the first 100 spectral estimates, 4 spectral bands over the next 200 spectral estimates and 8 spectral bands were averaged for the remainder of the spectra.
- The frequency separating the regions where 2 and 4 spectral bands were averaged was 0.59 cycles per day and the frequency separating the regions where 4 and 8 spectral bands were averaged occurred at 1.76 cycles per day.
- Spectral averaging enabled confidence limits to be generated.
- The 95% confidence limits are used for both the energy spectra and the coherence squared spectrum.
- No confidence limits for the phase difference spectrum are presented.

Description: The resulting spectral energy, coherence and phase difference spectra are presented in Figure 5–11 (the bottom, middle and

top plots, respectively). The 95% confidence limits for the spectrum and coherence squared are superimposed on the respective plots, as is the '-$5/3^{rds}$ power law'.

The spectral energy for the two data sets are similar (Figure 5–11, bottom plot).

- Indian Ocean spectral energy (red plot) between the frequencies of 1 and 6 cycles per day are dominated by the tides at the diurnal and semi-diurnal frequencies, and their harmonics (3, 4, 6 and 8 cycles per day). Spectral peaks at 4, 6, 8 and 10 cycles per day are also statistically significant.
- Pacific Ocean spectral energy (blue plot) is dominated by the diurnal and semi-diurnal tides. While spectral peaks are also apparent at the harmonics (3, 4, 6 and 8 cycles per day) their size is less than the distance between the upper and lower confidence limits (i.e. these peaks are not statistically significant). This does not necessarily indicate that these peaks are unimportant, they simply don't meet our criteria for significance.
- The diurnal and semi-diurnal signals contribute 10% and 83%, respectively, to the total spectral energy (or variance) at the Pacific Ocean location and 1% and 96%, respectively, for the Indian Ocean location.
- Both spectra follow the '-$5/3^{rds}$ power law' for frequencies between about 0.04 and 0.5 cycles per day.
- At frequencies exceeding about 3 cycles per day (with the exception of the narrow spectral peaks at 3, 4, 6 and 8 cycles per day) both energy spectra follow the '-$5/3^{rds}$ power law'. However, the spectrum of the Indian Ocean at consistently about 1 order of magnitude greater than that of the Pacific Ocean spectrum. The very large energy input at low frequencies cascades into the higher frequencies with an overall net increase in energy.

The coherence squared spectrum is shown in Figure 5–11, middle plot.

- In general, there is high coherence squared between the two data sets for frequencies higher than about 1 cycle per day.

- At lower frequencies there are several areas of high coherence squared (some of which are statistically significant). These peaks occur at 0.2, 0.3, 0.35 and 0.4 cycles per day.

The phase difference spectrum is shown in Figure 5–11, top plot.

- The phase spectrum near the diurnal frequency contains a cluster of points near -135°, as observed in the previous example.
- At the semi-diurnal and higher frequencies, the phase difference is about 180°. That is, when a (tidal wave) peak occurs in the Pacific Ocean data, a trough appears in the Indian Ocean data.
- At the lower frequencies, where the coherence squared spectrum was close to the 95% confidence limit, the phase difference is close to zero. At those frequencies, high (and low) water levels at the two locations are observed at the same time in both data sets.

Returning to our original question:

(a) For frequencies less than about 1 cycle per day, the coherence between the two data sets is, generally, less than the 95% confidence limit. There is significant coherence (above the 95% confidence limit) near 1 cycle per day, and, in general, for frequencies greater than about 1.5 cycles per day. It is, perhaps, a little unusual that signals at high frequencies lie above the 95% confidence limit while those at low frequencies do not. This would likely trigger further investigation to uncover the reason for this.

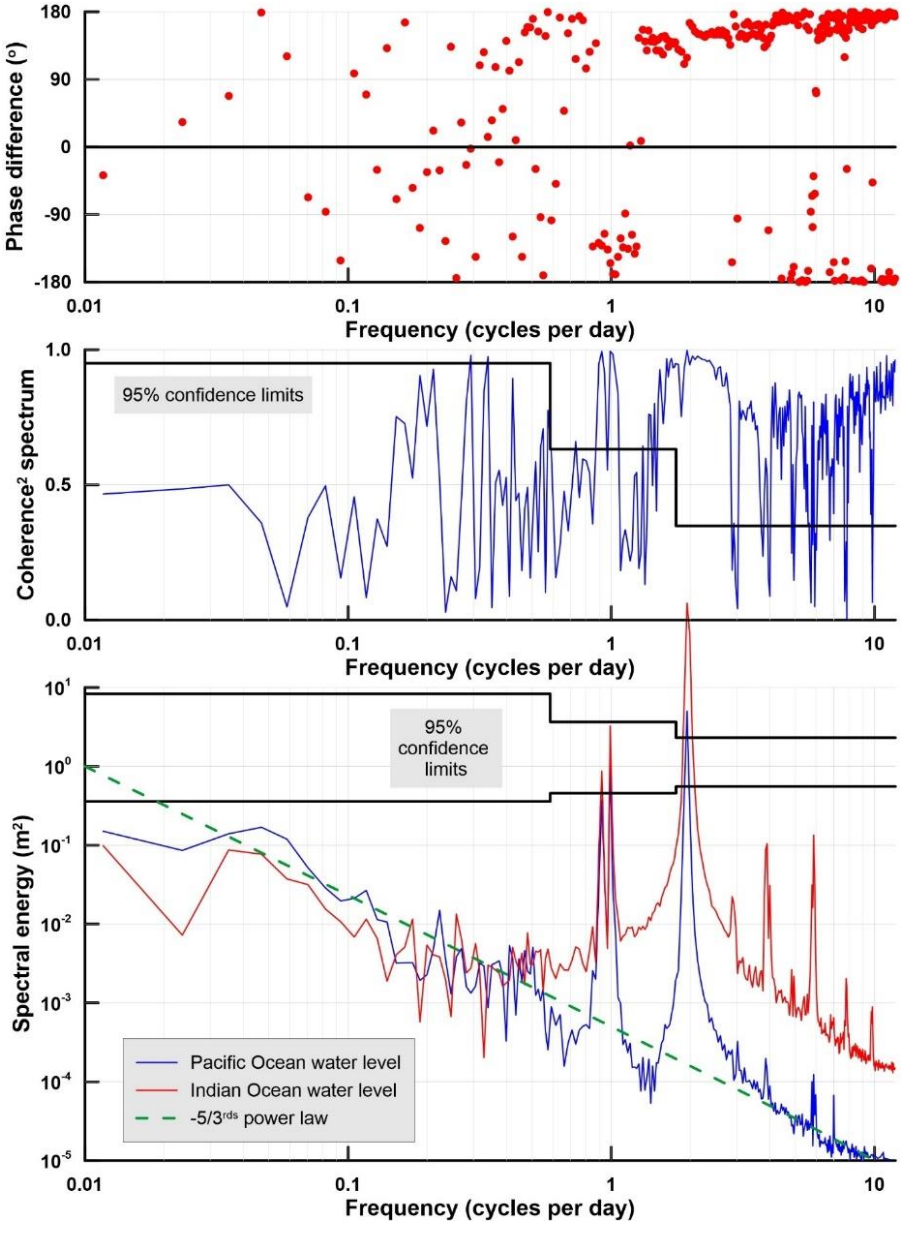


Figure 5–11. Periodogram, coherence squared and phase difference spectra for Pacific Ocean and Indian Ocean water levels.

5.2.6. Example 5-6: Temperature and water level – daily data

The cross-spectrum examples above have all used the same type of variables. There is nothing to stop us from undertaking a cross-spectrum using two different types of variables. In this example the two variables we'll use are water temperature and water level.

If two different types of variables are being use in any analysis, it is sometimes recommended to normalise the data – basically converting the data to non-dimensional form. One common way to do this is to divide the data by the standard deviation. There is nothing to prevent us from normalising the data here. However, dividing the data by a constant (e.g. the standard deviation) has no effect on the coherence squared or phase difference spectra. This is because these two spectra are only interested in the relationship (i.e. correlation) between the two data sets at each specific frequency. Further, the spectral energy will change – by definition dividing by the standard deviation will result in a variance of unity. Having said this, the spectral energy axis becomes effectively unitless, which is perhaps convenient. In practical terms, there is little to gain by normalising the data sets prior to undertaking a cross spectral analysis.

Questions: How strongly correlated is water temperature and water level at the same location at frequencies less than 12 cycles per year?

Data used:

- The original data were sampled at hourly intervals over 11 years. These data were filtered and subsampled at daily intervals, producing 4,096 data points (or days).
- The Nyquist period is 2 days (Nyquist frequency is ½ cycle per day), substantially greater than the highest frequency we wish to resolve (viz. 12 cycles per year or 0.03 cycles per day).
- The data set is 11 years long, therefore the longest period features we can realistically interpret is about 1-2 years – sufficient to effectively interpret features at 1 cycle per year.

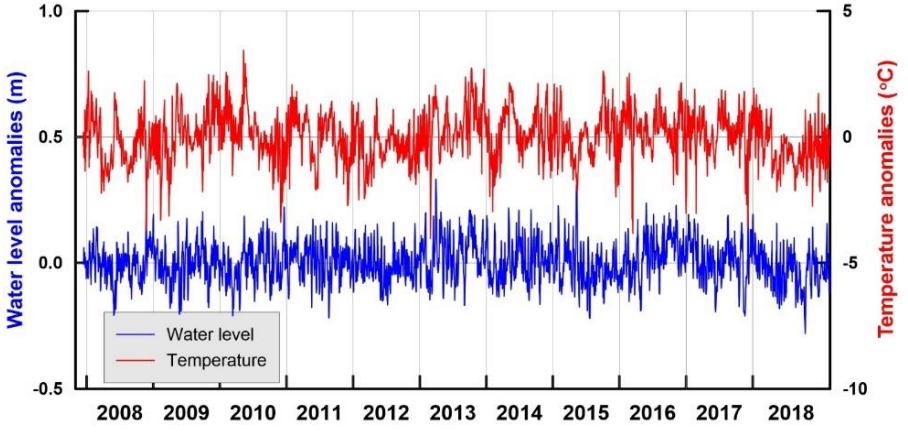


Figure 5–12. Time series of temperature and water levels anomalies. Note the vertical axis on the right has been offset to separate and clarify the two plots and its scale is ten times that of the scale on the left.

Pre-processing:

- Both data sets were subject to the same filtering.
- From previous examples we know that water temperature data contains a strong annual signal and water level data contain strong tidal signals (particularly at the diurnal, semi-diurnal and spring / neap periods). These strong signals are filtered out prior to the cross-spectrum analysis.
- The annual signal was removed using the filter described in Section 3.3.6, while the tides were removed using Munk's tide killing filter described in Section 3.3.7.
- The linear trend was also removed from the data.
- The time series plots of the resulting water level and temperature anomalies are shown in Figure 5–12. Visually, there appears to be little in the plots other than random data.

Testing:

- Stationarity: Results from the Dickey-Fuller test give Tau calculated values of -15.61 (for the temperature anomalies) and -16.13 (for the water level anomalies); Tau critical is - 2.86. As the calculated Tau is less than Tau critical for both data sets, we can reject the null hypothesis that the data sets are not stationary.
- Recapping the relevant statistical tests. A Chi-squared goodness of fit test to check for normal distributions. Both data sets failed this test at the 5% level of significance and we conclude that the data were not normally distributed. To check for weak stationarity, we split each of the temperature and water level data into 16 sets each of 256 data points. Bartlett's test was used to check for homogeneity of variance. ANOVA was carried out on the 16 data sets to check whether the means could have originated from the same data. We rejected the null hypotheses for both data sets and concluded that the data were not obtained from the same population (i.e. the data sets are not weakly stationary). Noting this, we proceeded with the cross spectral analysis.

Smoothing of the periodogram:

- Both periodograms were averaged as follows: 4 spectral bands were averaged over the first 200 spectral estimates, 8 spectral bands over the next 800 spectral estimates and 16 spectral bands were averaged for the remainder of both spectra. The frequency separating the regions where 4 and 8 spectral bands were averaged was 17.8 cycles per year and the frequency separating the regions where 8 and 16 spectral bands were averaged occurred at 53.5 cycles per year.
- Spectral averaging enabled confidence limits to be generated. The 95% confidence limits are used for both the energy spectra and the coherence squared spectrum.
- No confidence limits for the phase difference spectrum are presented.
- Note; given that our interest lies in frequencies less than about 12 cycles per year (i.e. a monthly period), we could analyse the data

as 11 separate data sets each of duration 12 months. The spectral estimates could then be averaged to produce confidence limits as per Example 5-3.

Description: The spectral energy, coherence and phase difference spectra are presented in Figure 5–13 (the bottom, middle and top plots, respectively). The 95% confidence limits for the spectrum and coherence squared are superimposed on the respective plots, as is the '-5/3^{rds} power law'. Note, as we did not normalise the data prior to the analysis, the units for the spectral energy differ for the temperature and water level spectra.

- Neither energy spectra show any clear peaks whose heights exceed the distance between the upper and lower confidence limits.
- Both spectra flatten at lower frequencies (less than about 20 cycles per year).
- At frequencies greater than about 50 cycles per year, both energy spectra approximately follow the '-5/3^{rds} power law'. However, it is noted that the energy cascade in these frequencies does appear to fall off a little steeper than the '-5/3^{rds} power law' – perhaps more like '-2/1'. The reason for this is unclear and may warrant further investigation. (There may be a low energy input across a broad frequency band – about 10-50 cycles per year, possibly generated by weather systems).
- Spectral energy at frequencies greater than about 20 cycles per year is highly variable, spanning about one-to-two orders of magnitude.

The coherence squared spectrum is shown in Figure 5–13, middle plot.

- There is high coherence squared between the two data sets for frequencies between 1 and 2 cycles per year and again at about 3 cycles per year.
- Some specific areas of high coherence squared occur between 10 and 100 cycles per year.
- There is a broad band (frequencies exceeding about 150 cycles per year) where the coherence squared lies above the 95% confidence limit.

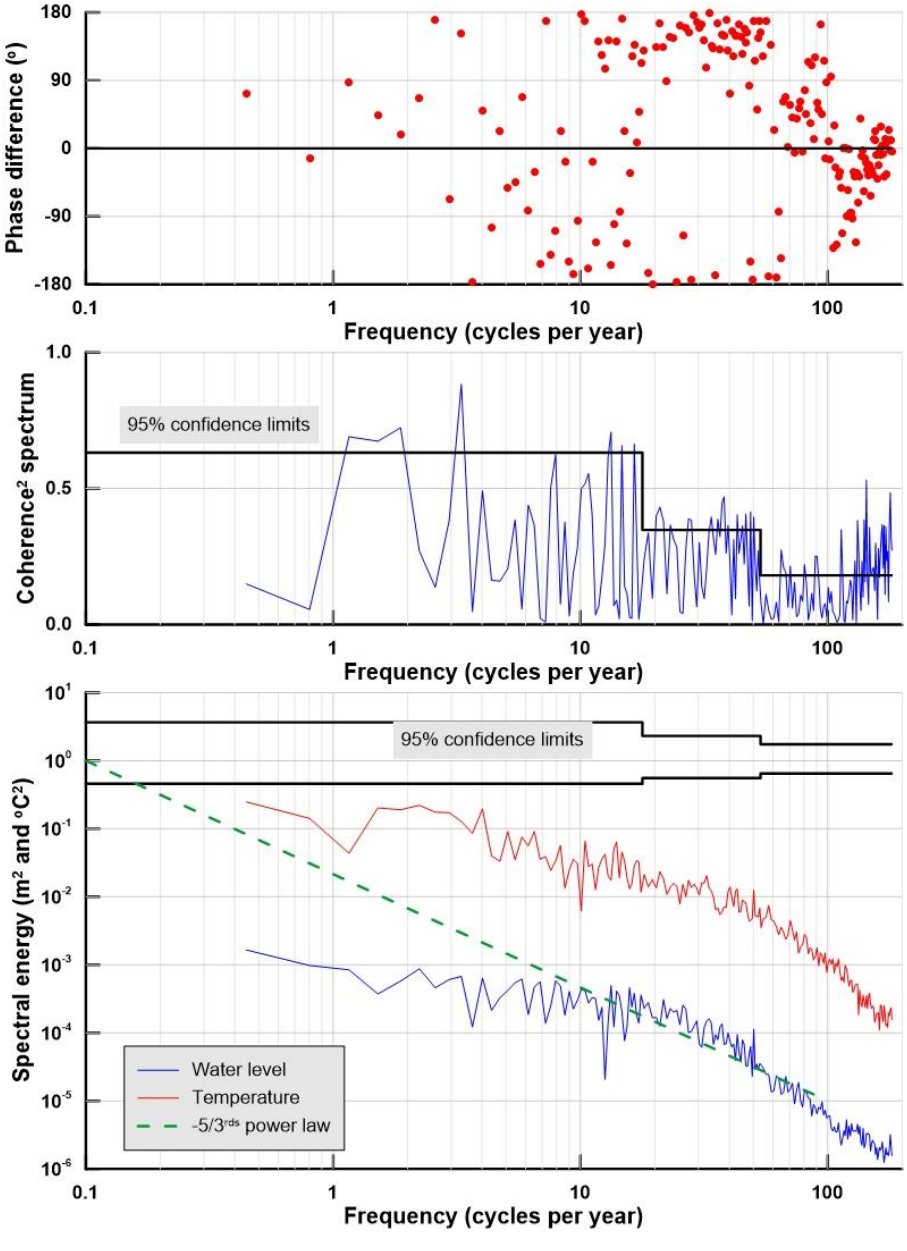


Figure 5–13. Periodogram, coherence squared and phase difference spectra for the water temperature and water level anomalies.

The phase difference spectrum is shown in Figure 5–13, top plot.

- The phase difference between 1 and 2 cycles per year averages at about +45°. That is the temperature leads the water level by about $1/8^{th}$ of a cycle.
- The phase difference at 3 cycles per year is about 180° i.e. high temperatures correspond with low water levels and vice versa.
- Between 10 and 100 cycles per year, the phase difference is also about 180°.

Returning to our original question:

(a) The linear trend and the major signals have been removed from both data sets. The resulting coherence squared is high between about 1 and 2 cycles per year and again at about 3.5 cycles per year. Apart from these two frequency bands, there appears to be little significant correlation between the temperature and water level anomalies.

Above, it was asserted that normalising of the data had no meaningful effect on the results. The data were normalised by dividing by their respective standard deviations and the resulting spectra shown in Figure 5–14. Close examination of Figure 5–13 and Figure 5–14 shows that the coherence squared and phase difference spectra are the same. The only difference occurs in the energy spectra, which have shifted vertically and reduced in magnitude compared with the corresponding plots of the non-normalised spectra of Figure 5–13.

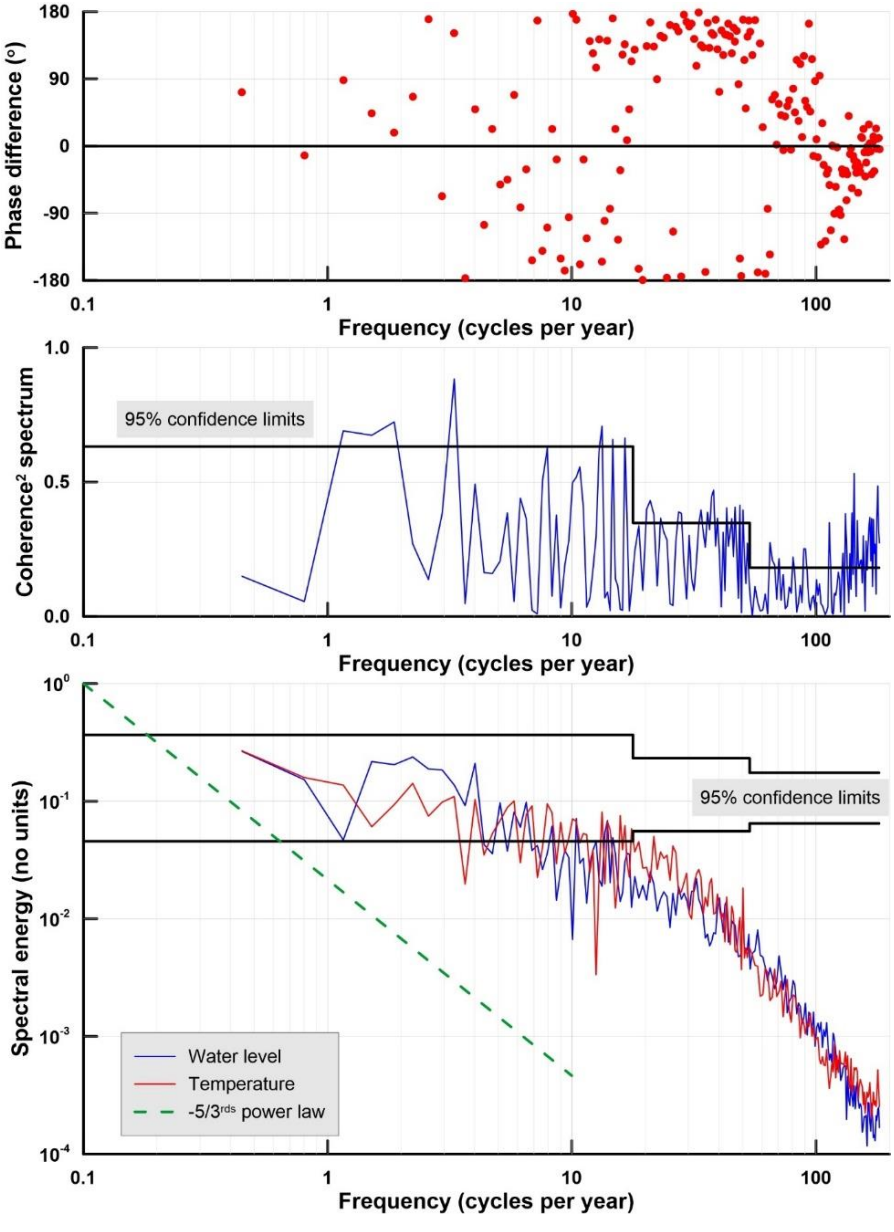


Figure 5–14. Periodogram, coherence squared and phase difference spectra for temperature and water level using normalised data.

5.2.7. Example 5-7: Salinity and east-west current component – 10 min data

Thus far, most of the work has focused on water temperature and water level. This example examines salinity and currents. Data are obtained from a location in an estuary that runs east-west. Like many estuaries, it has a strong tidal signal. A large river discharges water of low salinity to the coastal ocean (high salinity). The magnitude of the salinity will increase or decrease in response to the up stream / down stream motion of the tidal currents. Therefore, we might expect a strong correlation between the salinity and the currents – in this case the dominant (east-west) current component.

Again, the two different units of measurement give us the option of normalising the data and comparing results. As noted in the previous example, this does not affect the results and the dimensional form of the data is used. Note also that salinity is a unitless quantity. However, we've used the old fashioned units of 'psu' (practical salinity units).

Questions: At what frequencies is the correlation strongest between salinity and the east-west current component?

Data used:

- Data were sampled at 10 minute intervals over about one month, producing 4,096 data points.
- The Nyquist period is 20 minutes (a Nyquist frequency of 3 cycles per hour).
- The data set is about 28 days, indicating the longest period features we can realistically interpret is about 3-6 days. This restricts any answer to our question to the frequency band 3 cycles per hour to about 0.2 cycles per day.

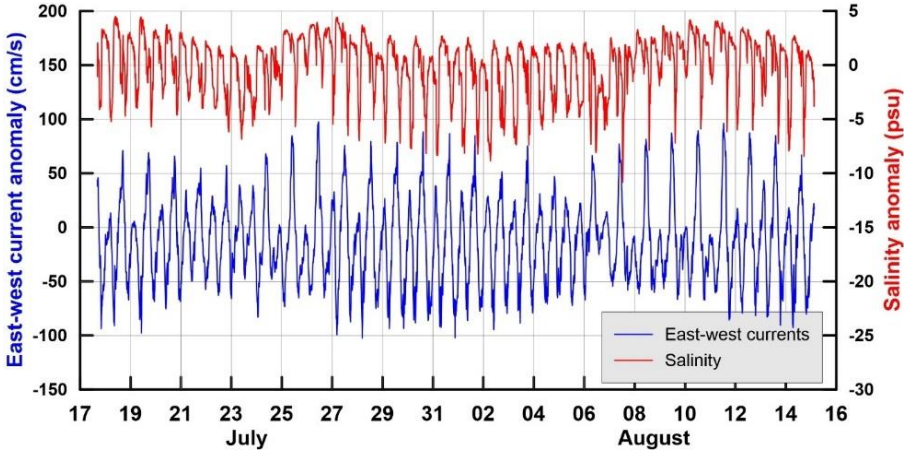


Figure 5–15. Time series of salinity and the east-west current component anomalies. Note the vertical axis on the right has been offset to separate and clarify the two plots and its scale is one tenth that of the scale on the left.

Pre-processing:

- The linear trend was removed from both sets of data.
- The time series anomalies are plotted in Figure 5–15. Visually, both data sets have a strong signal near the semi-diurnal period.

Testing:

- Stationarity: Results from the Dickey-Fuller test give Tau calculated values of -8.86 (for the salinity anomalies) and -5.56 (for the east-west current component anomalies); the Tau critical value is - 2.86. As the calculated Tau is less than Tau critical for both data sets, we can reject the null hypothesis that the data sets are not stationary.
- As we've seen with most of our previous examples, results from the alternate statistical tests for stationarity indicate that the data were not obtained from the same, respective, populations. Noting this, we proceeded with the cross spectral analysis.

Smoothing of the periodogram:

- Both periodograms were averaged as follows: 4 spectral bands were averaged over the first 200 spectral estimates, 8 spectral bands over the next 800 spectral estimates and 16 spectral bands were averaged for the remainder of both spectra. The frequency separating the regions where 4 and 8 spectral bands were averaged was 17.8 cycles per year and the frequency separating the regions where 8 and 16 spectral bands were averaged occurred at 53.5 cycles per year.
- Spectral averaging enabled confidence limits to be generated. The 95% confidence limits are used for both the energy spectra and the coherence squared spectrum.
- No confidence limits for the phase difference spectrum are presented.

Description: The resulting spectral energy, coherence and phase difference spectra are presented in Figure 5–16 (the bottom, middle and top plots, respectively). The 95% confidence limits for the spectrum and coherence squared are superimposed on the respective plots, as is the '-5/3^{rds} power law'. Note, the units for the spectral energy differ for salinity and the east-west current component.

- Both spectra show large peaks at frequencies of 1, 2, 3 and 4 cycles per day, corresponding to the diurnal, semi-diurnal tides and their harmonics. Most of the spectral energy lies in the semi-diurnal band, representing about 40% and 70% of the total spectral energy in the salinity and currents, respectively.
- Large spectral peaks at frequencies of 5, 6, 7 and 8 cycles per day can also be observed in the salinity spectrum.
- At frequencies greater than about 6 cycles per day, both energy spectra approximately follow the '-5/3^{rds} power law'.
- Spectral energy at frequencies greater than about 10 cycles per day is highly variable, spanning about one-to-two orders of magnitude.
- The current spectrum flattens at lower frequencies (less than about 0.8 cycles per day). This flatten of the spectrum at low

frequencies in the salinity spectrum is not as marked, despite having removed the linear trend from the data.

The coherence squared spectrum is shown in Figure 5–16, middle plot.

- There is high coherence squared between the two data sets for selected frequencies across the spectrum. Specifically, at frequencies near 0.5, 1, 2, 3, 4, 5 and 7 cycles per day.

The phase difference spectrum is shown in Figure 5–16, top plot.

- The phase difference at 1 and 2 cycles per year is about $+90^{o}$. That is the salinity variations lead the east-west currents by about $1/4^{th}$ of a cycle.
- The phase difference at 3 and 4 cycles per year is close to zero i.e. high salinity and large current speeds occur at about the same time for these frequencies.
- At frequencies of 5 and 7 cycles per day, the phase difference is about -90^{o} i.e. at these frequencies, salinity lags the east-west current component by about $1/4^{th}$ of a cycle.

Answering the question we asked:

(a) The correlation between salinity and the east-west component of the currents at frequencies near 0.45, 1, 2, 3, 4, 5 and 7 cycles per day. Most of these frequencies correlate with a component of the tides (or the harmonics).

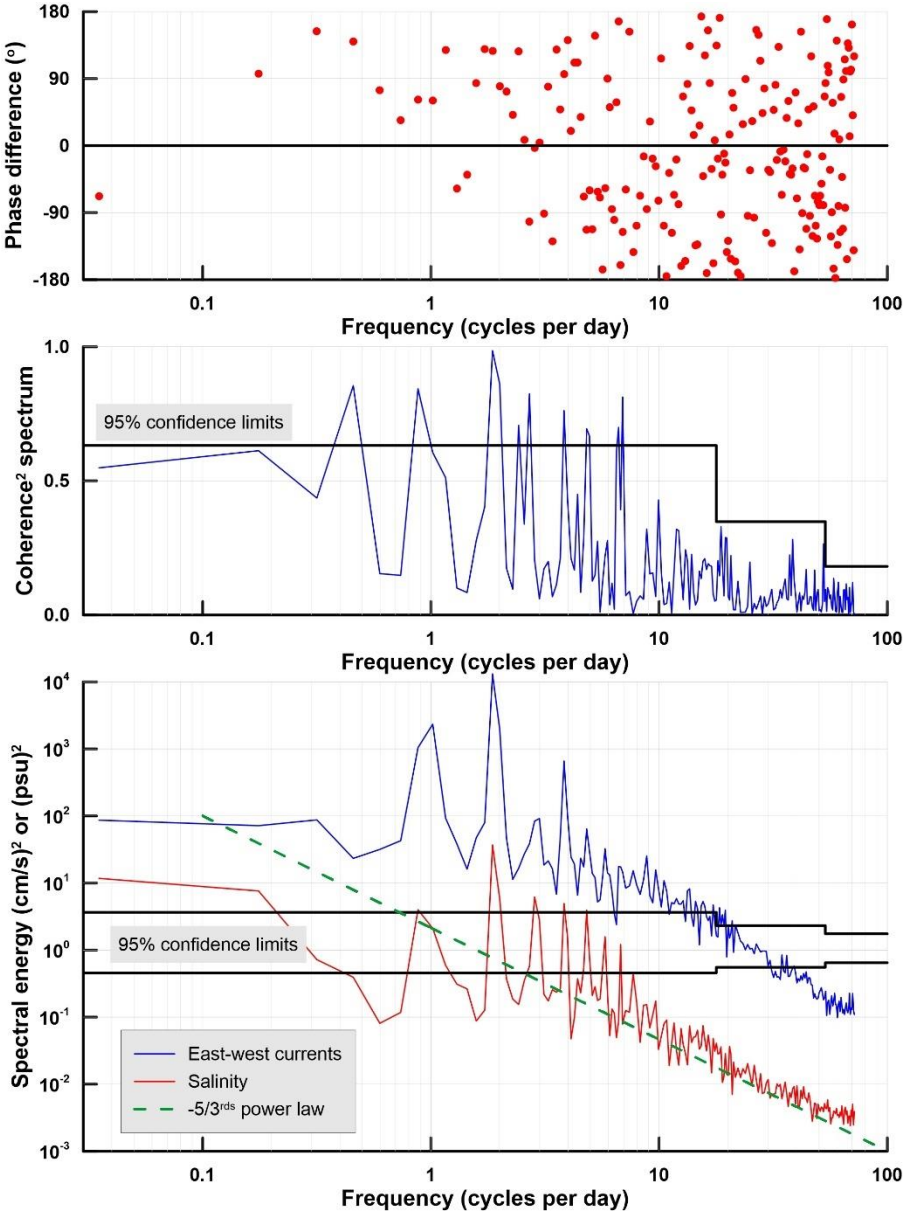


Figure 5–16. Periodogram, coherence squared and phase difference spectra for salinity and east-west current component.

5.2.8. Example 5-8: East-west and north-south current components – 5 min data

The above examples have all been applied to scalars (or a single component of a vector). Cross-spectral analysis can be also applied to vectors e.g. as obtained from wind or current data. Vectors can be expressed in terms of magnitude and direction or in component form. Direction is a circular function, which does not work well in spectral analysis. Therefore, we use the component form of a vector.

Direction convention.

Atmospheric physicists (and wind recording instruments) use a direction convention such that direction is defined as the direction FROM which the wind is blowing. Oceanographers (and current recording instruments) use a direction convention such that direction is defined as the direction TOWARDS which the water is flowing. Our background is in oceanography, so we change any wind data that we might use to the oceanographic direction convention by adding 180° to the wind direction. Speed and direction data are then converted to component form – initially as north-south and east-west components, but resolution along any pair of (usually) orthogonal axes is acceptable.

A compass (as used in an atmospheric or oceanographic instrument) defines positive directions as clockwise from north, while mathematical operations define positive directions as anti-clockwise from the x-axis (usually the horizontal axis or towards the east). Using the oceanographic direction convention and the direction (θ) as recorded by the instrument, the north-south and east-west current components are derived from the speed and direction as follows:

$$North/south\ component = speed\ \times cos(\theta)$$

$$East/west\ component = speed\ \times sin(\theta)$$

Positive values are defined as currents (or winds) flowing (blowing) towards the north (and east), while negative values indicate the currents (or winds) are flowing (blowing) towards the south (or west).

A compass as used in an instrument will read direction with respect to 'magnetic north'. Often this is changed to 'true north' by adding or subtracting the 'magnetic variation' to/from the direction provided by the instrument. The magnetic variation varies at different locations on the earth and over time. For many temperate regions, this value is relatively small (perhaps $<15^{o}$). However, in polar regions, the magnetic variation many be substantially greater than this value.

We select hourly current data for the first half of 2018. Data are converted from speed and direction to the north-south and east-west current components.

Questions: At what frequencies is the correlation strongest between the north-south and east-west current components?

Data used:

- The original data were sampled at 5 minute intervals. A 7-point Gaussian average was used to smooth the data, which were then subsampled to hourly intervals.
- The total length of the data set is 4,096 hours, or about 6 months.
- The Nyquist period is 2 hours (Nyquist frequency is ½ cycle per hour.
- The data set is about 170 days long, indicating the longest period features we can realistically interpret is about 20-40 days. This restricts any answer to our question to the frequency band bounded by 12 and (about) 0.05 cycles per day.

Pre-processing:

- The linear trend was removed from both data sets creating a set of data anomalies.
- The time series plots of the anomalies are shown in Figure 5–17. Visually, the currents in the north-south direction are substantially greater than those in the east-west direction. Despite this, many of the large features in the north-south current component can

also be seen in the east-west current component. To understand the processes in the data set, it may be preferable to maximise the energy in one component and minimise the energy in its orthogonal component. (We'll explore this later in this example and in the following example).

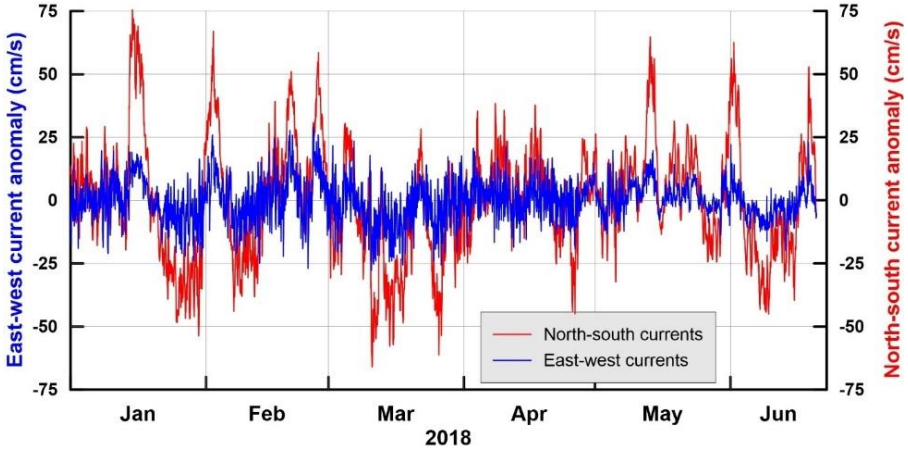


Figure 5–17. Time series of north-south and east-west current components.

Testing:

- Stationarity: Results from the Dickey-Fuller test give Tau calculated values of -21.15 (for the east-west current anomalies) and -6.29 (for the north-south current component anomalies); the Tau critical value is - 2.86. As the calculated Tau is less than Tau critical for both data sets, we can reject the null hypothesis that the data sets are not stationary.
- Results from the alternate statistical tests for stationarity (homogeneity of means and variances) indicate that the data were not obtained from the same, respective, populations. Noting this, we proceeded with the cross spectral analysis.

Smoothing of the periodogram:

- Both periodograms were averaged as follows: 4 spectral bands were averaged over the first 200 spectral estimates, 8 spectral bands over the next 800 spectral estimates and 16 spectral bands were averaged for the remainder of both spectra. The frequency separating the regions where 4 and 8 spectral bands were averaged was 17.8 cycles per year and the frequency separating the regions where 8 and 16 spectral bands were averaged occurred at 53.5 cycles per year.
- Spectral averaging enabled confidence limits to be generated.
- The 95% confidence limits are used for both the energy spectra and the coherence squared spectrum.
- No confidence limits for the phase difference spectrum are presented.

Description: The resulting spectral energy, coherence and phase difference spectra are presented in Figure 5–18 (the bottom, middle and top plots, respectively). The 95% confidence limits for the spectrum and coherence squared are superimposed on the respective plots, as is the '-$5/3^{rds}$ power law'.

- Despite having removed the linear trend, the region of greatest spectral energy lies at the low frequency end of both spectra.
- The frequencies of greatest spectral energy occur near 0.02 and 0.06 cycles per day. The latter has a period of about 16 days, close to the spring / neap tidal period (14 days). The reason for the spectral peak near 0.02 cycles per day is unclear, and may warrant a more detailed analysis.
- There is a slight flattening of both spectra at frequencies less than about 0.1 cycles per day.
- Large spectral peaks are observed at the diurnal and semi-diurnal periods (near 1 and 2 cycles per day), although the spectral energy at these frequencies is substantially lower than that at low frequencies.
- Both spectra generally follow the '-$5/3^{rds}$ power law'. However, at frequencies greater than about 1-2 cycles per day, the spectral energy appears to be offset by about one order of magnitude.

- The high frequency regions of both spectra are highly variable with spectral energy spanning about two orders of magnitude.

The coherence squared spectrum is shown in Figure 5–18, middle plot, together with the phase difference spectrum (Figure 5–18, top plot).

- There is high coherence squared between the two data sets for frequencies less than about 0.2 cycles per day. The phase difference is close to zero.
- At frequencies greater than 0.2 cycles per day, there are many peaks in the coherence squared spectrum that exceed the 95% confidence limits.
- Between frequencies of about 1 and 10 cycles per day (and where the coherence squared lies above the 95% confidence limits) the phase difference varies (approximately) linearly, increasing from 45° to about 135°. Although it is noted that there is considerable variability in the phase difference spectrum.

Returning to the question we have asked.

(a) The correlation (coherence) between the two data sets is greatest at frequencies less than about 0.2 cycles per day and in the frequency band 1 to 4 cycles per day.

Information from the coherence squared and phase difference spectra (and the time series plots) suggests that the same processes are represented in both data sets. So, we may ask if it is possible to maximise the effect of these processes in one direction while minimising the effects in the orthogonal direction? The simple answer is, yes, this can be achieved in many cases by rotating the axes. However, the questions remain: 'by how much do we rotate the axes and how do we justify rotating the axes in the first place'?

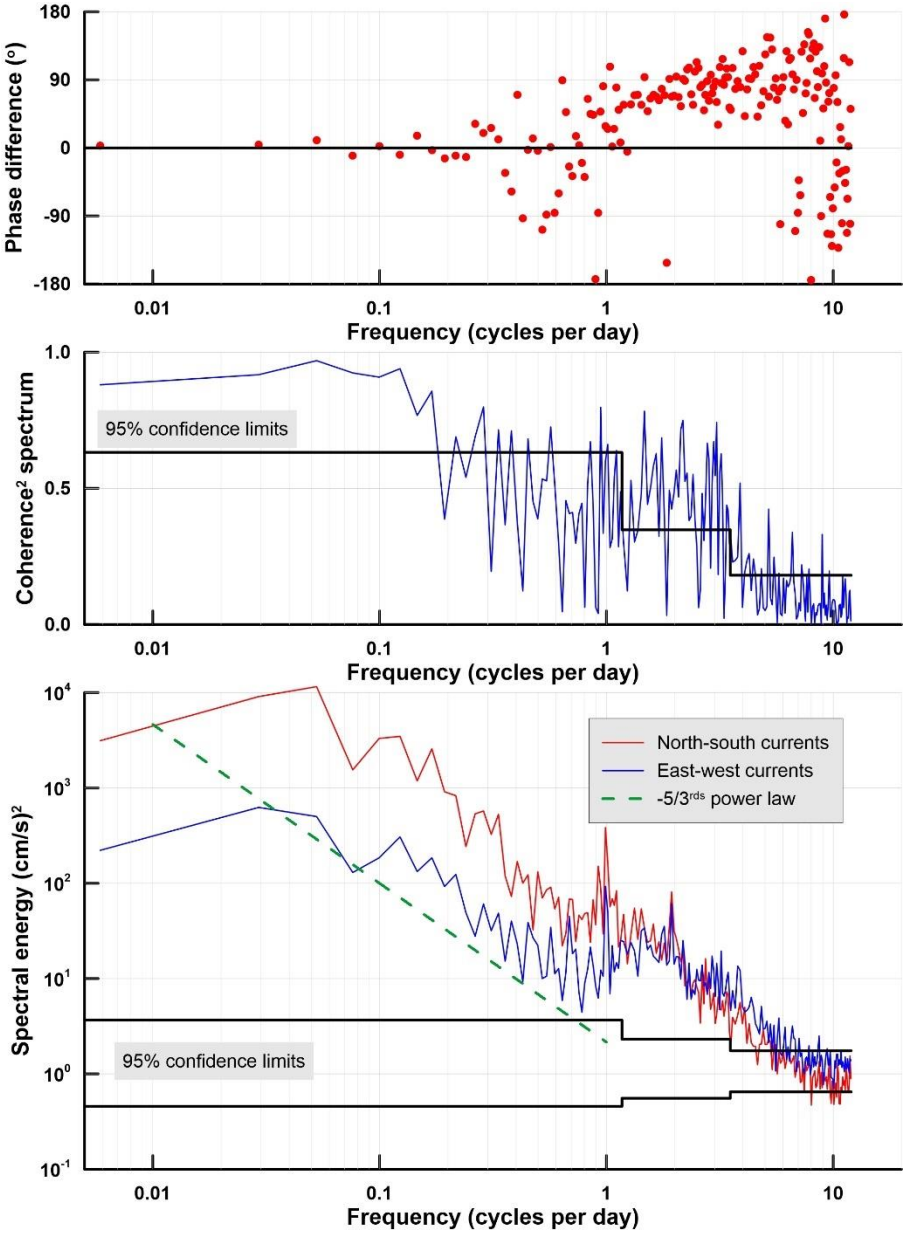


Figure 5–18. Periodogram, coherence squared and phase difference spectra for the north-south and east-west current components.

5.2.9. Example 5-9: Current components – rotated

The axes about which the current components can be rotated are arbitrary. Often, we use the north-south and east-west components of the currents – as in the previous example. However, the spectral energy in each component will differ depending on the orientation of the axes used. The question is: 'what is the most appropriate axis rotation to use'?

Some of the options available to help us make this decision are:

- Randomly choose a direction and see what results (not the most elegant of approaches and certainly not one we'd recommend).
- Undertake a scatter plot of the data and rotate the axes according to the slope of the resulting regression line.
- Use a navigational chart to estimate the orientation of the local bathymetry and rotate the axes such that the axes are parallel to, and perpendicular to, the local bathymetry.
- Use Principal Component Analysis, PCA (or Empirical Orthogonal Function analysis, EOF), or specifically in this case, principal axis analysis. This is an objective approach to determine the orientation of the principal axes i.e. the axes about which the variance in the data is maximised (the first principal axis) and minimised (the second principal axis).

Our preference is the last of these options. PCA provides a consistent approach to choosing an appropriate rotation of the axes. It also turns out that this direction is close to that of the orientation of the local bathymetry. While it sounds a bit daunting, it is a relatively simple approach to undertake. We'll use the same data set as used in the previous example. (PCA can also be used to reduce the dimension of the data i.e. reduce the number of variables that need to be analysed. Details of PCA are presented in Chapter 13).

The direction of the first principal axis is the direction along which the variance in the data is maximised. The direction of the second principal axis is the direction along which the variance in the data is minimised. It is the main direction along which the currents flow. (And, yes, this does

sound counter-intuitive). The direction of the first principal axis is given by:

$$\theta = 0.5 \times tan^{-1}\left[\frac{2 \times covariance(x,y)}{variance(x) - variance(y)}\right]$$

In this example PCA reveals that: $\theta = -14^o$. That is: the direction of the first principal axis is 14° clockwise from east (or 104° clockwise from north). Recall, mathematically, directions are defined as positive being anti-clockwise from the 'east' (or horizontal) axis. The orthogonal direction to this is 14° clockwise from north.

The data selected are the same as those used in Section 5.2.8. Scatter plots of the data anomalies (trends removed) are given in Figure 5–19 (original or unrotated data on the left and rotated data on the right). The unrotated data are the east-west and north-south components, while the rotated data now become the along-shore and across-shore components.

Superimposed on the unrotated scatter plots are two lines – the red line representing the direction of the linear regression line and the blue line representing the approximate alignment of the local bathymetry. The slope of the regression line is about 12° clockwise from north while the local bathymetry is about 10° clockwise from north. These directions are essentially the same, and about the same as the direction obtained from the principal axis analysis (14°). The scatter plot on the right-hand side of Figure 5–19 shows the data rotated anti-clockwise by 14°. The data are now such that the dominant direction of water flow is aligned with the grid (i.e. in the along-shore direction).

Questions: At what frequencies is the correlation strongest between the along-shore and across-shore current components? What are the differences between unrotated and rotated spectra?

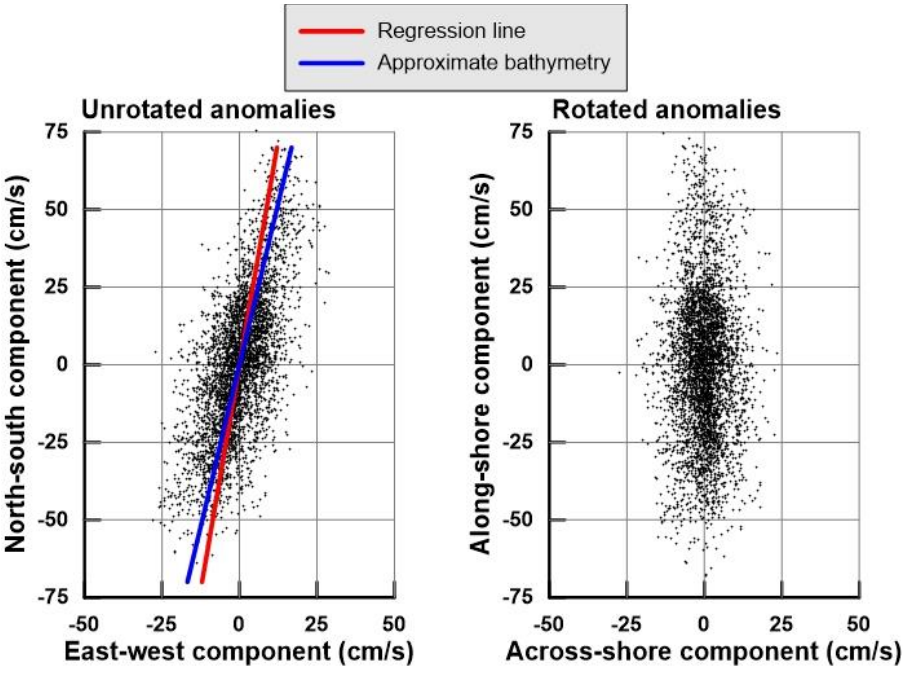


Figure 5–19. Scatter plot of the current component anomalies (left) and data rotated anti-clockwise by 14º (right).

Data used:

- The original data were sampled at 5 minute intervals and a 7-point Gaussian average used to smooth the data, which were then subsampled to hourly intervals.
- The total length of the data set is 4,096 hours, or about 6 months.
- The Nyquist period is 2 hours (Nyquist frequency is 0.5 cycles per hour).
- The data set is about 170 days long, indicating the longest period features we can realistically interpret is about 20-40 days. This restricts any answer to our question to the frequency band bounded by 12 and (about) 0.05 cycles per day.

Pre-processing:

- The linear trend was removed from both data sets to produce anomalies.
- The time series anomaly plots are shown in Figure 5–20. Visually, the currents in the along-shore direction are substantially greater than those in the across-shore direction.
- The along-shore component contains the large signals apparent in the north-south direction (Figure 5–18). However, those same features cannot be seen in the across-shore component, in contrast to the east-west component (Figure 5–18).
- The across-shore component appears random-like.
- From early May onwards, the variations in the across-shore component are considerably less than the variations earlier in the data set.

Testing:

- Stationarity: Results from the Dickey-Fuller test give Tau calculated values of -29.08 (for the east-west current anomalies) and -6.25 (for the north-south current component anomalies); the Tau critical value is - 2.86. As the calculated Tau is less than Tau critical for both data sets, we can reject the null hypothesis that the data sets are not stationary.
- Results from the alternate statistical tests for stationarity (homogeneity of means and variances) indicate that the data were not obtained from the same, respective, populations. Noting this, we proceeded with the cross spectral analysis.

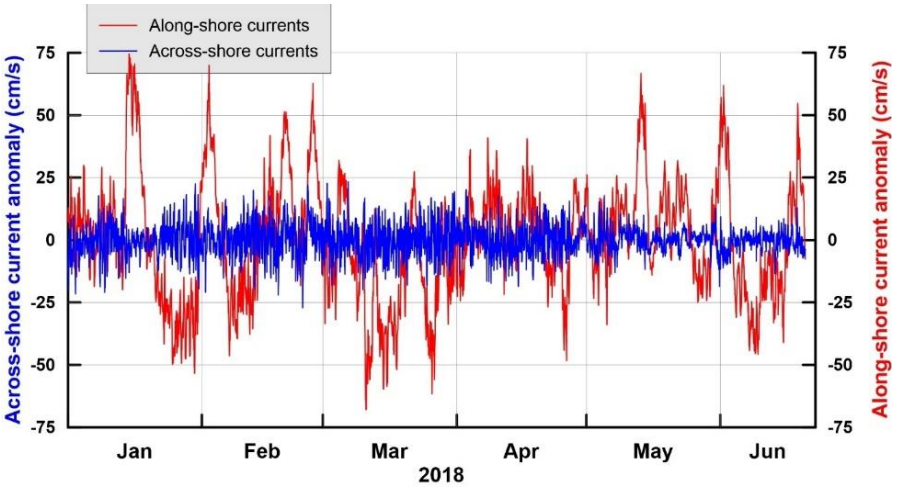


Figure 5–20. Time series of along-shore and across-shore current components.

Smoothing of the periodogram:

- Both periodograms were averaged as follows: 4 spectral bands were averaged over the first 200 spectral estimates, 8 spectral bands over the next 800 spectral estimates and 16 spectral bands were averaged for the remainder of both spectra. The frequency separating the regions where 4 and 8 spectral bands were averaged was 17.8 cycles per year and the frequency separating the regions where 8 and 16 spectral bands were averaged occurred at 53.5 cycles per year.
- Spectral averaging enabled confidence limits to be generated. The 95% confidence limits are used for both the energy spectra and the coherence squared spectrum.
- No confidence limits for the phase difference spectrum are presented.

Description: The resulting spectral energy, coherence and phase difference spectra are presented in Figure 5–21 (the bottom, middle and top plots, respectively). The 95% confidence limits for the spectrum and

coherence squared are superimposed on the respective plots, as is the '-5/3^{rds} power law'.

This example compares the results between the north-south and east-west current components (from the previous example) with the along-shore and across-shore current component. The data used means that we are restricted to the frequency bands 12 to (about) 0.05 cycles per day (i.e. periods between 2 hours and 20 days).

The spectral energy for the two data sets is compared: Figure 5–18 (north-south and east-west components) and Figure 5–21 (along shore and across shore components).

- There is little difference in the north-south and along-shore spectral energy. Rotation of the axes by a relatively small amount (14^{o}) makes little difference to the spectral energy along the dominant axis.
- The major difference in the energy spectrum occurs in the across-shore component at frequencies less than about 0.5 cycles per day. At these low frequencies the spectral energy in the across-shore component is considerably less than that in the east-west component. This 'energy deficit' has been transferred to the along-shore component.
- The low frequency end of the across-shelf energy spectrum is relatively flat, reminiscent of the 'random' energy spectrum shown in Figure 2–2.
- Other features of the energy spectra remain similar to those described in the previous example.

The coherence squared spectrum is shown in Figure 5–21, middle plot, together with the phase difference spectrum (Figure 5–21, top plot).

- The high coherence seen in the unrotated current components (Figure 5–18, middle plot) has now disappeared. All of the low frequency variations have been moved into the along-shore component. At low frequencies, the across-shore component of the current is random-like (the energy spectrum is relatively flat).

- Again, higher frequency features remain largely unchanged from those observed in Example 5.2.8.

Returning to the question we have asked.

(a) The correlation (coherence) between the two data sets is greatest in the frequency band 1 to 4 cycles per day.
(b) The major changes between the unrotated and rotated current components occur at the low frequency end of the spectra (in this case, at frequencies less than about 0.5 cycles per day). The data were obtained from a location about 3 km offshore, in an area where no major estuaries exist. Therefore, there is little opportunity for substantial currents to exist in the across-shore direction at low frequencies (there is simply nowhere for the water to go). Hence the low frequency part of the energy spectrum is relatively flat. However, there remains the opportunity for across-shore currents to still occur for short periods of time. Therefore, the high-frequency part of the across-shore energy spectrum (and the coherence squared spectrum) remains similar to that of the east-west spectrum.

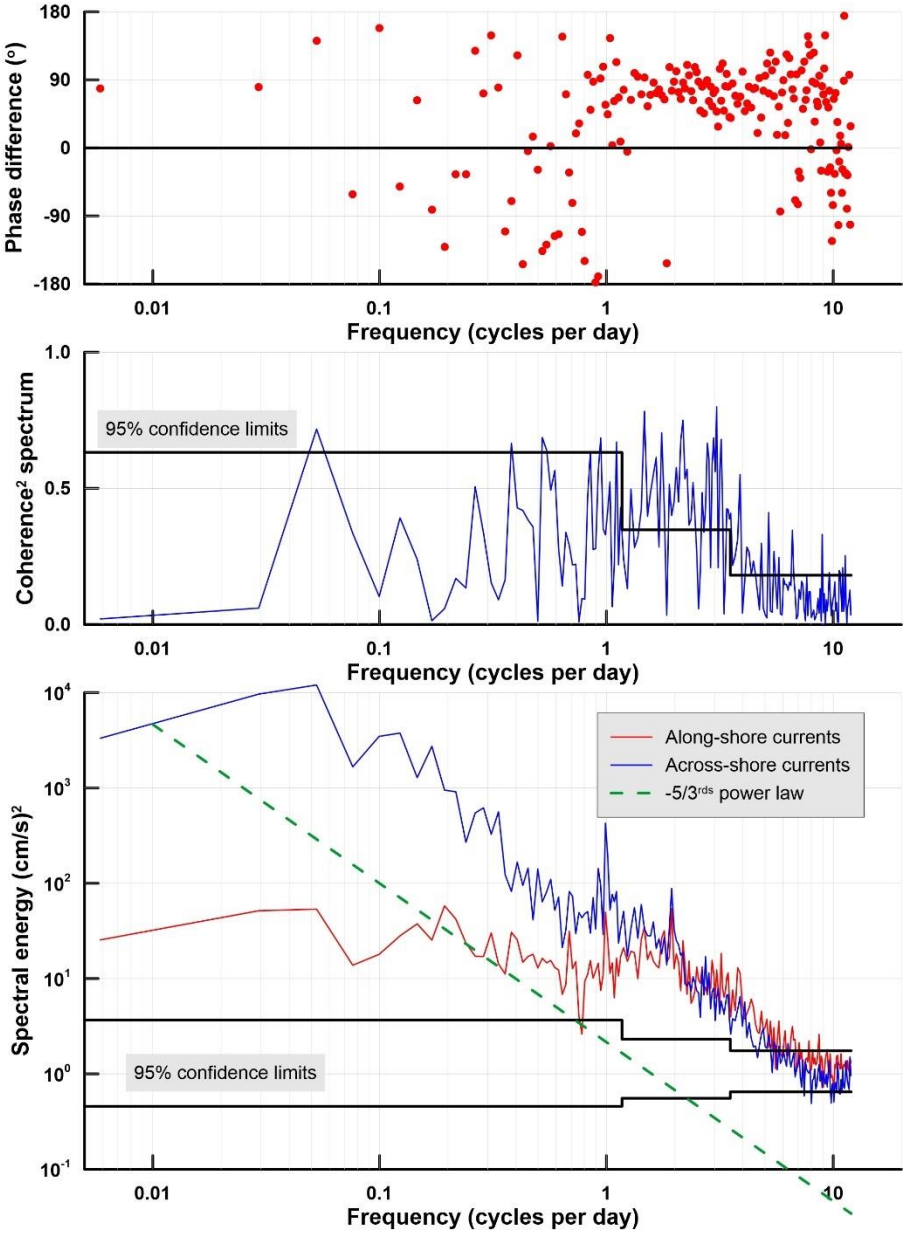


Figure 5–21. Periodogram, coherence squared and phase difference spectra for the along-shore and across-shore current components.

5.2.10. Example 5-10: Some 'what ifs'

We complete this section by examining a few scenarios or 'what ifs'. We'll choose to demonstrate some 'what ifs' using four years of hourly water level data collected from Fort Denison, Australia (although any variable could be used).

In Example 5-1 (Section 5.2.1) we showed the results when two temperature data sets were close – separated vertically by only 4 m. The results were spectral energy plots that were close in magnitude, has high coherence squared and the phase difference was close to zero. What happens if the two data sets are identical?

We would expect to see the following:

(a) The spectral energy of the two data sets to be identical over all frequencies.
(b) Correlation between the two data sets should be perfect i.e. equal to unity. This should translate to the coherence squared likewise being unity.
(c) The phase lag of both data sets should be the same, hence the phase difference should be zero.

This is exactly what is seen in Figure 5–22.

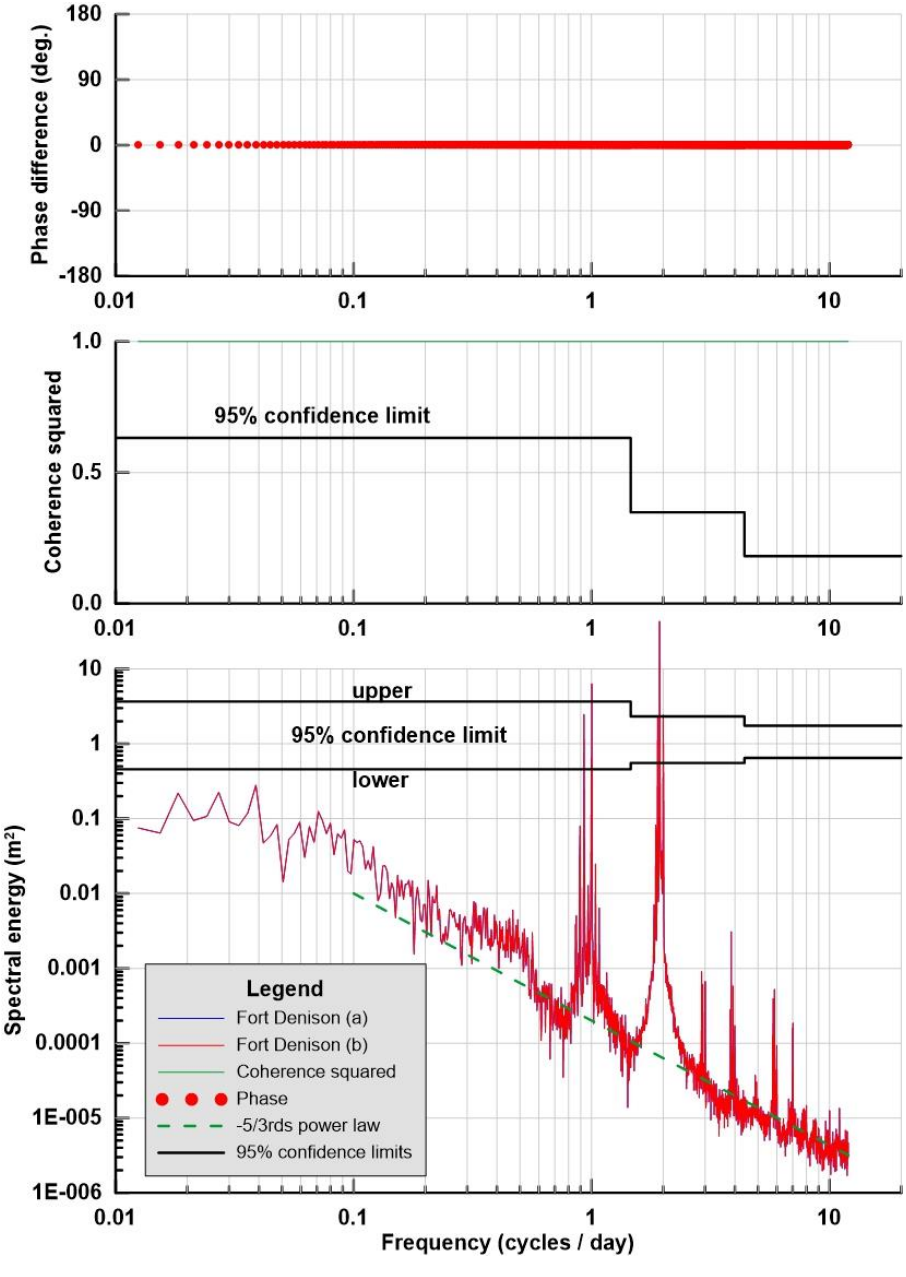


Figure 5–22. Periodogram, coherence squared and phase difference spectra for two identical water level data sets.

We repeat this exercise using identical data sets, however one data set uses 3 decimal places which the second ('identical') data set only uses 2 decimal places. Generally, we'd expect the data sets to produce identical results, but with some slight variation at high frequency. (After all, the only difference in the two data sets occurs in the third decimal place).

This is what we see in Figure 5–23. At low frequencies, there is no difference between the two data sets. At frequencies exceeding about 1 cycle per day small differences in the energy spectra start to appear. These differences translate to slightly lower coherence squared values (although the remain above the 95% confidence limit) and slightly greater variation in the phase difference spectrum.

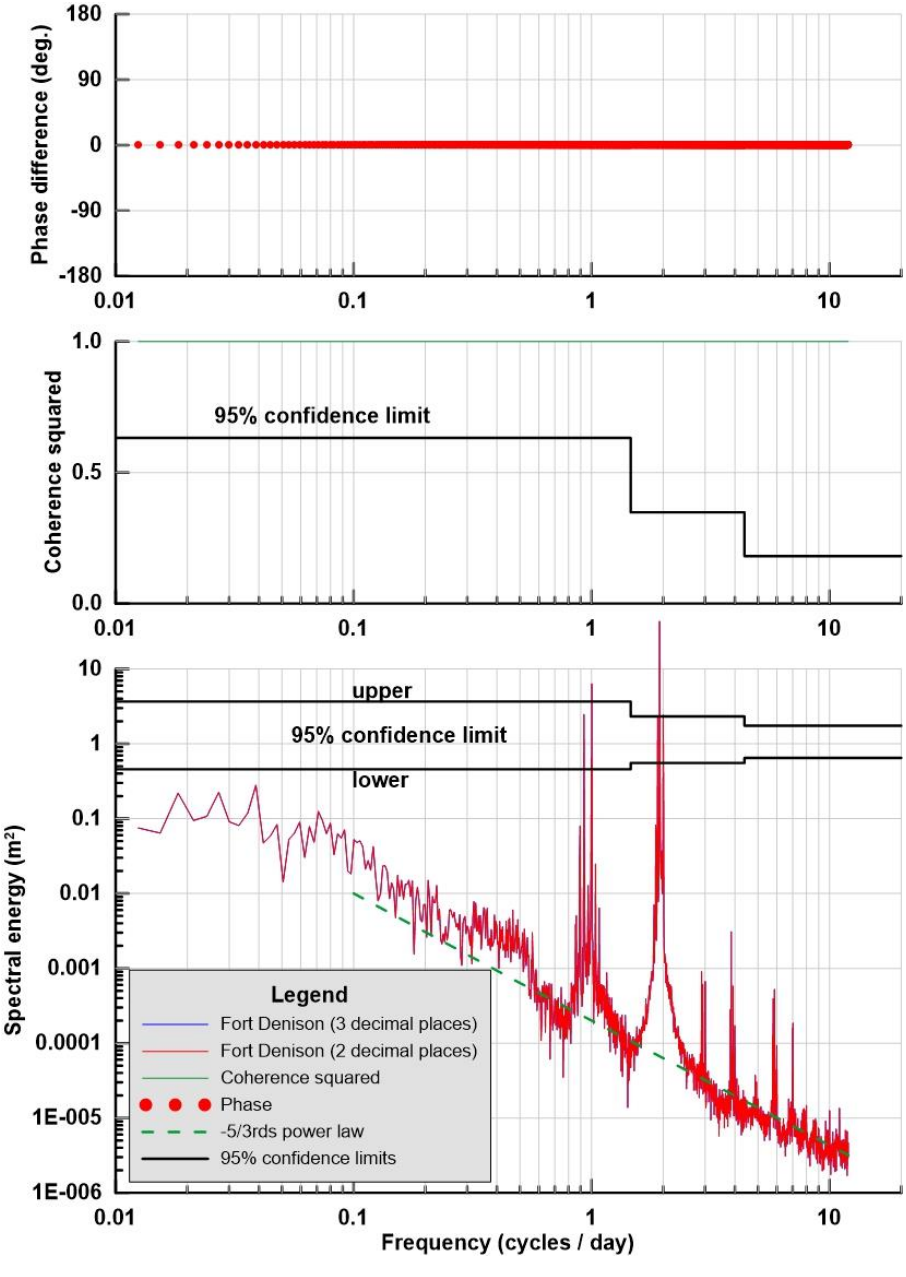


Figure 5–23. Periodogram, coherence squared and phase difference spectra for two identical water level data sets – one quoted to 3 decimal placed and the other quoted to two decimal places.

Finally, what happens if the two data sets are formed from random numbers? In this example, two sets of random numbers were generated between zero and unity.

We would expect to see the following:

(a) Both periodograms should be relatively flat. As shown in Section 2.5, a data set comprising random numbers should result in a flat periodogram.
(b) Correlation between the two data sets will be random. This should translate to a coherence squared spectrum that is highly variable, but whose values generally lie below the 95% confidence limit.
(c) Similarly, the phase lag of both data sets will be random and the phase difference spectrum will be highly variable.

Again, this is what is observed in Figure 5–24.

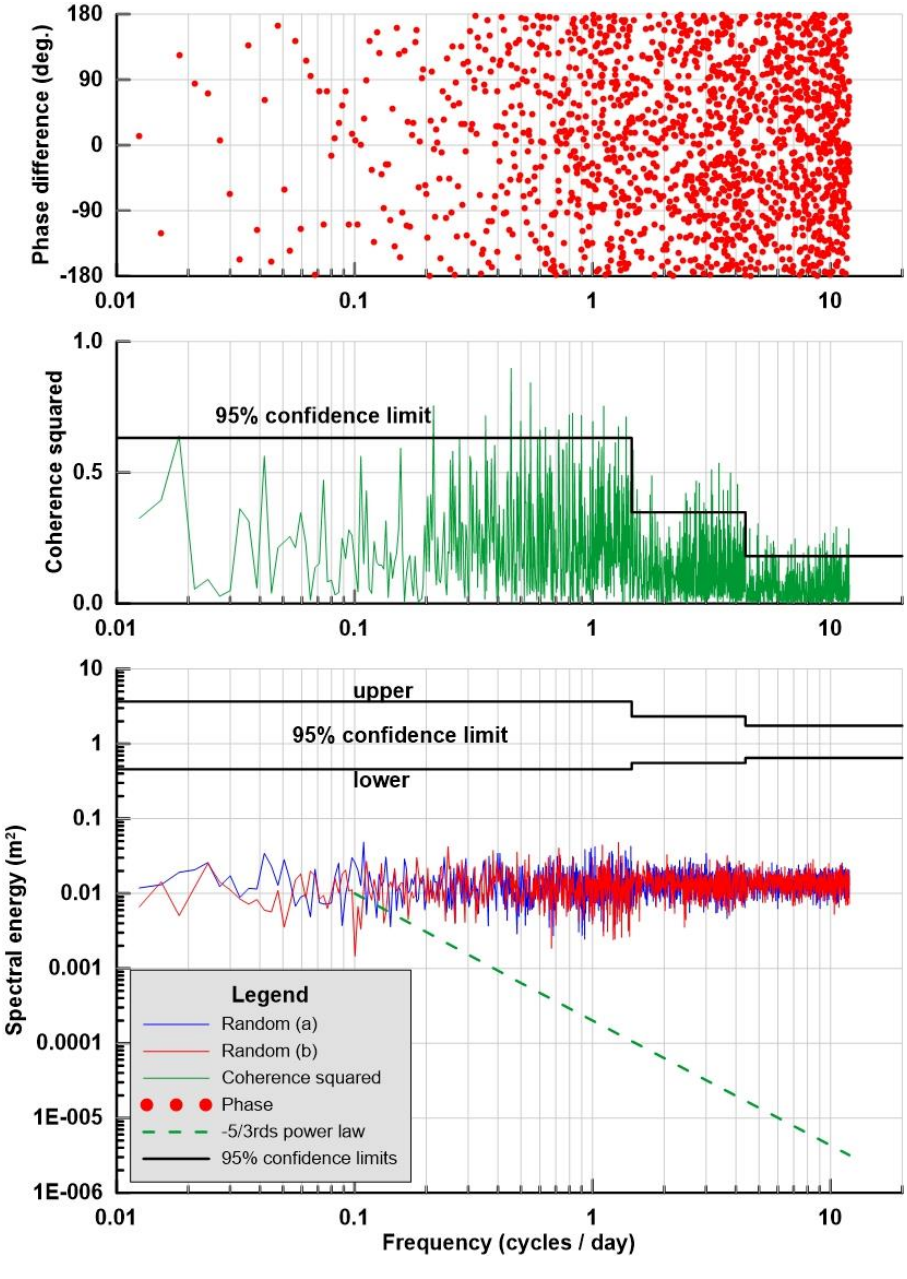


Figure 5–24. Periodogram, coherence squared and phase difference spectra for two random data sets.

5.3. Summary

Cross-spectrum analysis allows us to spectrally compare two different data sets. The data sets may be two scalars, a scalar and a vector component or the two components of a vector. The units for each variable do not need to be the same and the data do not need to be normalised.

- The two data sets do not necessarily have to be from the same time period. However, they do need to be of the same length and have the same time step.
- With a single data set, we generate an amplitude spectrum and phase lag spectrum. Using cross-spectral analysis, we can uniquely specify the spectral characteristics using three spectra: e.g. the cross-amplitude spectrum, the coherence squared spectrum and the phase difference spectrum. There are several other ways in which we can specify this combination. The cross-amplitude spectrum can be replaced with the 'gain' spectrum or we can use the energy spectra of each variable (which is our preference).
- In undertaking a cross-spectrum analysis using vector components (e.g. winds or currents) the results will depend on the orientation of the axes. The optimal orientation can be obtained using principal axis analysis. 'Optimal' is in the sense that the variance of the data along one axis is maximised, while the variance along its orthogonal axis is minimised. The orientation of the principal axes should be close to that of the local bathymetry or the slope generated by regression line between the two components.

The axis direction ambiguity can be resolved using rotary spectra. This is the subject of the next chapter.

Chapter 6. Rotary Spectra

Suppose we have a vector (e.g. current or wind components) and we want to undertake a cross-spectral analysis of its components. In the previous chapter, we demonstrated that the result we get depends on the orientation of the axes we choose. We used principal axis analysis to find the orthogonal axes about which the variance was maximised and minimised. Invariably, this orientation aligns well with the direction of the local bathymetry.

This works well when the bathymetry has a dominant direction. But what if the bathymetry is convoluted – it varies considerably over short distances? Selecting a suitable axis or orientation (even using principal axis analysis) may not provide the best result. We can solve this problem by using rotary spectral analysis. Instead of resolving the vector along cartesian co-ordinates, rotary spectral analysis uses a pair of counter rotating vectors of different magnitude to replace the cartesian axes.

The two most common way of representing a vector are in radial form (using magnitude and direction) or component form (e.g. north-south and east-west components). A less common way to represent a vector is as the sum of two counter rotating vectors. The counter rotating vectors will have the same frequency but not necessarily the same amplitudes.

Now this is starting to sound a bit complex. But it is really just an expansion of the work done on cross-spectral analysis in the previous chapter. The mechanics are a bit easier to undertake than the topic is to understand – they are based on simple combinations of the information we have already generated. So, what's the advantage in rotary spectra? We see three advantages in using rotary spectral analysis.

(a) Many of the spectral results (including the spectral energy) do not depend on the orientation of the axes we choose. Hence, we don't have to think about optimising the orientation of our axes.
(b) It allows us to simultaneously analyse a pair of vectors, i.e. four components, two components from each vector.
(c) It provides another way to look at our data.

A flowchart of the rotary spectra output is shown in Figure 6–1. The auto spectra for each of the four time series is embedded in the flowchart, as is the cross spectrum for pairs of the time series. From that point, there is a departure and the complexity substantially increases. While two cross spectra can be constructed from any two time series contained in the four data sets, we take advantage of the fact that the four data sets are from a pair of vectors (each vector containing two components or two time series) from which further information can be obtained.

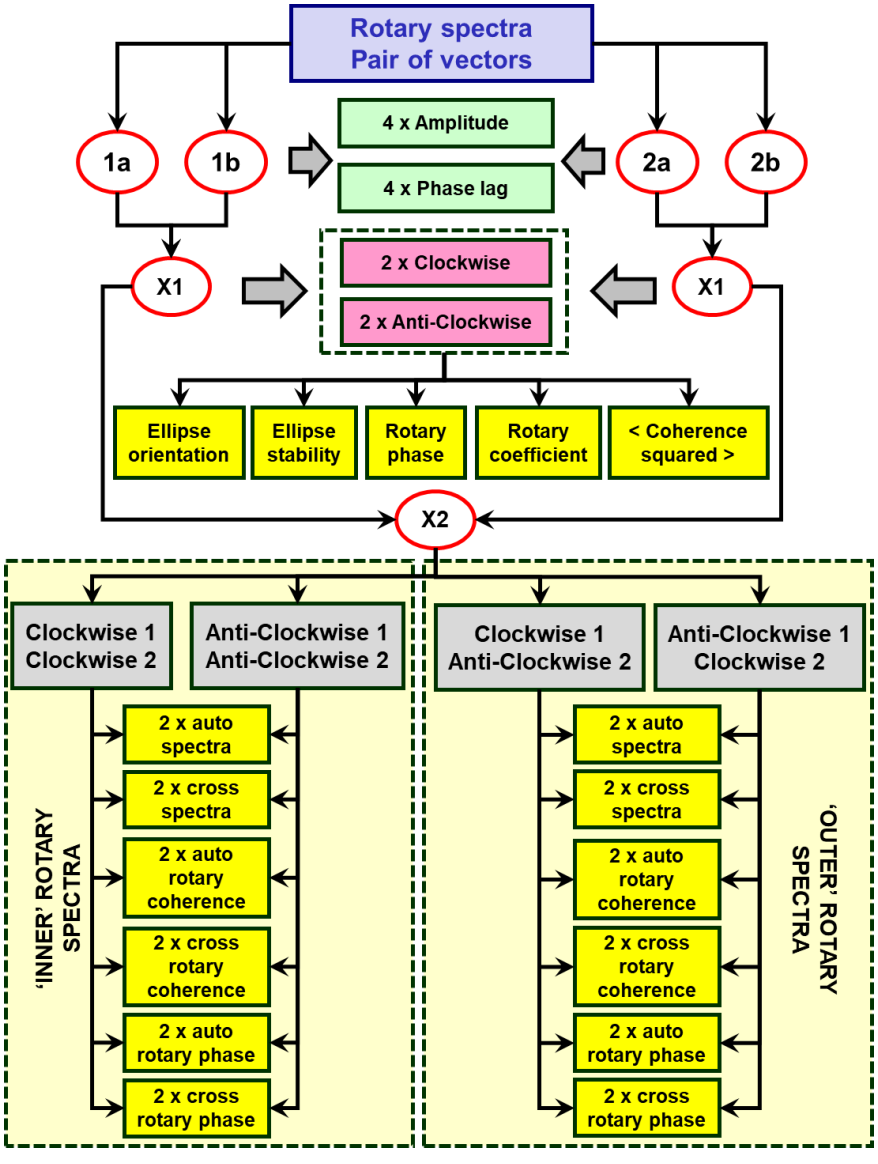


Figure 6–1. Flow chart of the rotary spectra. The two components of the first vector time series are '1a' and '1b', while those of the second vector are '2a' and '2b'. 'X1' indicates analyses involving the 'cross' of the components of the first vector (i.e. 1a and 1b) and the analyses involving the 'cross' of the components of the second vector (i.e. 2a and 2b). 'X2' indicates analyses involving the 'cross' of all four data sets (i.e. both vectors).

6.1. Theory behind rotary spectra

Firstly, we'll examine rotary spectra using a single vector. It will then be expanded to incorporate a pair of vectors.

6.1.1. A single vector

The following is primarily based on Gonella (1972) and Hayashi (1979). Recall from Section 2.3 our definition of the finite Fourier transform for a single variable is given by:

$$x_t = x(t) = \sum_{p=0}^{N/2} \left[a_p \cos\left(\frac{2\pi pt}{N}\right) + b_p \sin\left(\frac{2\pi pt}{N}\right) \right] \quad for \quad t = 1, 2, 3, \ldots N$$

The Fourier transform, $X(p)$, gives:

$$a_p = \frac{2}{N} \sum_{t=1}^{N} \left[x(t) \cos\left(\frac{2\pi pt}{N}\right) \right] \quad and \quad b_p = \frac{2}{N} \sum_{t=1}^{N} \left[x(t) \sin\left(\frac{2\pi pt}{N}\right) \right]$$

If x is a vector, we can represent it as a complex number:

$$x(t) = u(t) + i\, v(t)$$

And its Fourier transform is given by:

$$X(p) = U(p) + i\, V(p)$$

$$= \sum_{p=0}^{\frac{N}{2}} \left[a_p \cos\left(\frac{2\pi pt}{N}\right) + b_p \sin\left(\frac{2\pi pt}{N}\right) \right]$$

$$+ \; i \sum_{p=0}^{\frac{N}{2}} \left[c_p \cos\left(\frac{2\pi pt}{N}\right) + d_p \sin\left(\frac{2\pi pt}{N}\right) \right] \quad for \quad t = 1, 2, 3, \ldots N$$

$$= \sum_{p=0}^{\frac{N}{2}} [a_p + i\, c_p]\ cos\left(\frac{2\pi pt}{N}\right) + \sum_{p=0}^{\frac{N}{2}} [b_p + i\, d_p]\ sin\left(\frac{2\pi pt}{N}\right)$$

Let A, B, C and D be specific values of a_p, b_p and c_p, d_p for a given value of p and let $\theta = \frac{2\,\pi\,p\,t}{N}$. Then,

$$X(p) = [A + iC]\, cos(\theta) + [B + iD]\, sin(\theta)$$

Where p is an integer.

We can also write the Fourier transform of our vector as a pair of counter rotating vectors, (R^+, R^-), viz:

$$R(p) = R^+ e^{i\theta} + R^- e^{-i\theta}$$

$$= R^+ [cos(\theta) + i\, sin(\theta)] + R^- [cos(\theta) - i\, sin(\theta)]$$

$$= [R^+ + R^-]\, cos(\theta) + i\, [R^+ - R^-]\, sin(\theta)$$

The superscript $'+'$ indicates an anti-clockwise rotating vector and the superscript $'-'$ indicates a clockwise rotating vector.

We can also represent $R(p)$ as follows (Mooers, 1973):

$$R(p) = \begin{cases} |R^+|\, e^{i\theta} & for \quad p \geq 0 \\ |R^-|\, e^{-i\theta} & for \quad p \leq 0 \end{cases}$$

Where we've split the spectrum into two halves with the $'+'$ superscript representing the anti-clockwise rotating component ($p \geq 0$) and the $'-'$ superscript representing a clockwise rotating component ($p \leq 0$).

Notes:

$|R^+| = R^+ R^{+*}$ and $|R^-| = R^- R^{-*}$ where the asterisk $(*)$ indicates the complex conjugate.

The frequencies associated with $p < 0$ are not really negative. The frequency is determined using $|p|$. They are written as $p < 0$ simply to indicate the clockwise rotating component of the vector, while $p > 0$ indicates the anti-clockwise rotating component of the vector.

Equating the cosine and sine components for the expressions for $X(p)$ and $R(p)$ we get:

$$A + iC = [R^+ + R^-]$$

$$B + iD = i[R^+ - R^-]$$

Multiply the first equation by i and adding the two equations, we find the anti-clockwise rotating vector is given by:

$$R^+ = \frac{1}{2}(A + D) + \frac{i}{2}(C - B)$$

Multiply the second equation by i and adding the two equations, we find the clockwise rotating vector is given by:

$$R^- = \frac{1}{2}(A - D) + \frac{i}{2}(C + B)$$

The magnitude of the anti-clockwise rotating vector is defined as:

$$Anti\ Clockwise = \frac{1}{2} R^+ R^{+*}$$

Where the asterisk $(*)$ represents the complex conjugate. Substituting for R^+ and rearranging, we get:

$$Anti\ Clockwise = \frac{1}{8}[A^2 + B^2 + C^2 + D^2 + 2(AD - BC)]$$

We note that $(A^2 + B^2)$ is simply the square of the spectral components of the u component of the vector and $(C^2 + D^2)$ is the square of the spectral components of the v component of the vector. Further, we note that $(AD - BC)$ is the quadrature spectrum (see Section 5.1).

Similarly, the clockwise rotating vector is defined as:

$$Clockwise = \frac{1}{2} R^{-} R^{-*}$$

Substituting for R^{-} and rearranging, we get:

$$Clockwise = \frac{1}{8}[A^2 + B^2 + C^2 + D^2 - 2(AD - BC)]$$

These expressions for the clockwise and anti-clockwise rotating vector components are the same as those in Gonella (1972).



These vectors are rotating at the same frequency but in opposite directions and with different amplitudes. They map out an ellipse, with amplitudes R^{+} and R^{-} representing the major and minor axes (respectively) of the ellipse.

$$R^{+} = \frac{1}{2}(A + D) + \frac{i}{2}(C - B)$$

$$|R^{+}| = \sqrt{R^{+} R^{+*}} = \frac{1}{2}\sqrt{[(A + D)^2 + (C - B)^2]}$$

And

$$R^{-} = \frac{1}{2}(A - D) + \frac{i}{2}(C + B)$$

$$|R^{-}| = \sqrt{R^{-} R^{-*}} = \frac{1}{2}\sqrt{[(A - D)^2 + (C + B)^2]}$$

Further, since R^+ and R^- are complex numbers, 'angles' can be obtained. The first angle is obtained from the anti-clockwise rotating vector and is given by:

$$\psi^+ = tan^{-1}\left(\frac{C-B}{A+D}\right)$$

While the second angle is obtained from the clockwise rotating vector. It is given by:

$$\psi^- = tan^{-1}\left(\frac{C+B}{A-D}\right)$$

The orientation of the semi-major axis of the ellipse is given by:

$$\theta = \frac{1}{2}(\psi^+ + \psi^-) + k\pi = \frac{1}{2}\left[tan^{-1}\left(\frac{C-B}{A+D}\right) + tan^{-1}\left(\frac{C+B}{A-D}\right)\right] + k\pi$$

The orientation is defined as positive being anti-clockwise from the positive 'x' axis (or 'east' or 'real' axis). Note that there is a $k\pi$ ambiguity in this direction, resulting from the 'negative direction' of the semi-major axis. Ellipse orientation is not invariant under axis rotation. The units for the ellipse axes are the same as those of the original time series (usually m/sec or cm/sec).

The phase (or 'starting point' of the vector on the ellipse) is given by:

$$\phi = \frac{1}{2}(\psi^+ - \psi^-) = \frac{1}{2}\left[tan^{-1}\left(\frac{C-B}{A+D}\right) - tan^{-1}\left(\frac{C+B}{A-D}\right)\right]$$

The phase is relative to the major axis of the ellipse. Again, it is not invariant under rotation of axes.

The 'rotary stability' (γ_R) is defined as:

$$\gamma_R = \frac{(R^+R^{+*})\ (R^-R^{-*})}{(R^+R^{+*})^2 + (R^-R^{-*})^2}$$

$$= \frac{[A^2+B^2+C^2+D^2]^2 - 4[(A^2+B^2)(C^2+D^2) - (AC+BD)^2]}{[A^2+B^2+C^2+D^2]^2 - 4[(AD-CB)^2]}$$

It gives us an indication of the reliability of (or confidence in) the ellipse orientation. A stability value close to unity means that we have high statistical confidence in the orientation of the ellipse. A stability value close to zero means that we have little statistical confidence in the orientation of the ellipse.

The 'rotary coefficient' (C_R) defines the rotational motion (at a particular frequency) i.e. the tendency for the vector to rotate in a clockwise or anti-clockwise direction around the ellipse. It is given by:

$$C_R = \frac{R^- R^{-*} - R^+ R^{+*}}{R^- R^{-*} + R^+ R^{+*}} = \frac{-2\,(AD - BC)}{A^2 + B^2 + C^2 + D^2}$$

The rotary coefficient has a value between -1 and 1. Negative values indicate clockwise rotation while positive values indicate anti-clockwise rotation. If $|C_R| \longrightarrow 1$ then the motion tends to circular. If $|C_R| \longrightarrow 0$ then the motion tends to rectilinear.

The rotary coherence squared spectrum is non-trivial only when there is band averaging (averaging in the frequency domain).

We can define the following (band averaged) quantities:

$P_{uu} = \langle A^2 + B^2 \rangle$ auto-spectrum (u) squared

$P_{vv} = \langle C^2 + D^2 \rangle$ auto-spectrum (v) squared

$P_{uv} = \langle AC + \mathrm{BD} \rangle$ coincident spectrum

$Q_{uv} = \langle AD - \mathrm{BC} \rangle$ quadrature-spectrum

Then the following band averaged quantities apply:

$$Anti\ Clockwise = \frac{1}{8}[P_{uu} + P_{vv} + 2Q_{uv}]$$

$$Clockwise = \frac{1}{8}[P_{uu} + P_{vv} - 2Q_{uv}]$$

$$Rotary\ coefficient = \frac{-2Q_{uv}}{P_{uu} + P_{vv}}$$

$$Rotary\ stability = \frac{(P_{uu} + P_{vv})^2 - 4\,(P_{uu}\,P_{vv} - \mathrm{P}_{uv}^2)}{(P_{uu} + P_{vv})^2 - 4\,\mathrm{Q}_{uv}^2}$$

$$Ellipse\ orientation = \frac{1}{2}\tan^{-1}\left(\frac{2\,P_{uv}}{P_{uu} - P_{vv}}\right)$$

$$Rotary\ phase = \tan^{-1}\left(\frac{Q_{uv}}{P_{uv}}\right)$$

$$Rotary\ coherence\ squared = \frac{\mathrm{P}_{uv}^2 + \mathrm{Q}_{uv}^2}{P_{uu}\,P_{vv}}$$

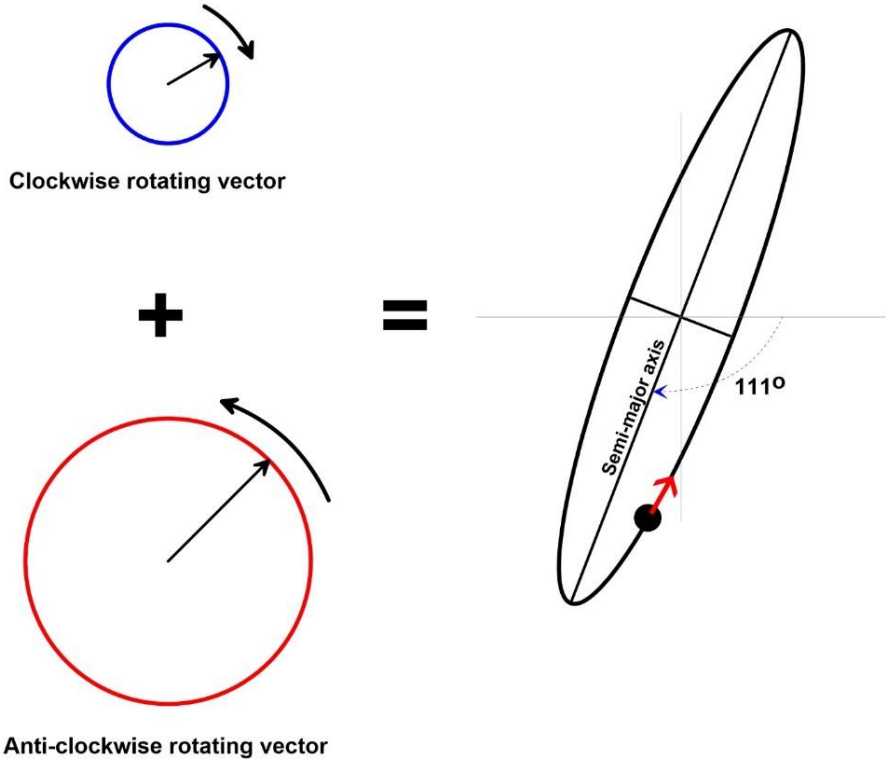


Figure 6–2. Graphical representation of the rotary spectral results at a frequency of 1 cycle per day. The two counter rotating vectors combine to form the ellipse.

Example: In one of the following examples the rotary spectral analysis resulted in the following: at a frequency of 1 cycle per day, clockwise spectral energy of 54.5 $(cm/s)^2$, anticlockwise spectral energy of 106.2 $(cm/s)^2$, rotary coefficient of -0.32 ellipse stability of 0.91, ellipse orientation of 111° and a rotary phase of 28°. The length of the semi-major axis was 3,748 and the length of the semi-minor axis was 683. These results are presented in Figure 6–2. A figure such as this can be generated for every frequency in the averaged rotary spectrum.

The anti-clockwise rotating vector is considerably larger than the clockwise rotating vector. This will result in the ellipse being somewhat elongated, as indicated by the magnitude of the rotary coefficient (0.32). The semi-major axis of the ellipse is rotated 111° from the x-axis. The negative sign of the rotary coefficient indicates that the direction of rotation is anti-clockwise, indicated by the red arrow on the ellipse. The black, filled circle on the ellipse is the rotary phase, which indicates the 'starting position' of the vector, relative to the start of the time series.

In summary, all that we need to specify the rotary components are the components of the two original energy spectra. The clockwise, anti-clockwise spectral components, together with the rotary coefficient and rotary stability are invariant under rotation. The orientation of the ellipse formed from the rotary spectral components, the rotary phase and the rotary coherence squared are all relative to the orientation of the original axes used and hence they are not invariant under rotation.

6.1.2. A pair of vectors

We can also undertake rotary spectra using a pair of vectors (Mooers, 1973; Middleton, 1982). This enables us to examine, for example, the relationship between ocean currents at two different locations (or depths) or the relationship between the wind and the currents.

This increases the complexity of the spectral analysis. When we had a scalar (or a single component of a vector) we generated two quantities: the amplitude (or energy) spectrum and the phase lag spectrum. When we examined the cross spectrum, each component generated two quantities and we also generated the cross-amplitude spectrum and the phase difference spectrum. (Note that the coherence squared spectrum is a product of the cross-amplitude spectrum).

When we have a single vector, we generate four rotary spectra: the clockwise and anti-clockwise rotary (energy) spectra, the rotary phase and the ellipse orientation. The other information we used in the previous section (e.g. rotary coefficient, rotary coherence squared etc) are products of these four quantities.

We now have a pair of vectors. Each vector will have these four quantities (which are each described in Section 6.1.1). However, we also have a further eight quantities, arising from the interactions of the two vectors. In total, we'll have 16 quantities (Mooers, 1973) and here we'll focus only on the 8 quantities that arise from the interactions (as the 4 rotary spectral components for each vector have already been described in Section 6.1.1). There will be:

- An amplitude and phase associated with the two anti-clockwise rotating vectors and an amplitude and phase associated with the two clockwise rotating vectors. These four quantities are referred to as the 'inner' spectra.
- Similarly, we'll have an amplitude a phase associated with the anti-clockwise component of data set one with the clockwise component of data set two and an amplitude a phase associated with the clockwise component of data set one with the anti-clockwise component of data set two. These four quantities are referred to as the 'outer' spectra.

Mooers (1973) used these terms, stating that their operations resembled the 'inner' (or scalar) product and the 'outer' (or vector) product.

Below, subscripts refer to the times series ($1\ or\ 2$) while the superscripts refer to the anti-clockwise $'+'$ and clockwise $'-'$ rotating vectors.

Note. Below, the p value is defined as positive and negative. This is simply a convenient way to represent the anti-clockwise and clockwise rotating components in a single figure. While we will talk about negative frequencies, they are simply the usual frequencies multiplied by -1. Positive frequencies or $p > 0$ are associated with the anti-clockwise rotating component while negative frequencies or $p < 0$ are associated with the clockwise rotating component.

Following on from Section 6.1.1, each of the two vectors $R_1(p)$ and $R_2(p)$ can be written as:

$$R_1(p) = \begin{cases} |R_1^+|\, e^{i\theta_1^+} & for \quad p > 0 \\ |R_1^-|\, e^{-i\theta_1^-} & for \quad p < 0 \end{cases}$$

Where, for each value of p (recall that $A, B, C, and\ D$ are specific components of R for a given value of p):

$$R_1^+ = \frac{1}{2}(A_1 + D_1) + \frac{i}{2}(C_1 - B_1)$$

$$R_1^- = \frac{1}{2}(A_1 - D_1) + \frac{i}{2}(C_1 + B_1)$$

$$tan(\theta_1^+) = \frac{C_1 - B_1}{A_1 + D_1}$$

$$tan(\theta_1^-) = \frac{C_1 + B_1}{A_1 - D_1}$$

And

$$R_2(p) = \begin{cases} |R_2^+|\, e^{i\theta_2^+} & for \quad p > 0 \\ |R_2^-|\, e^{-i\theta_2^-} & for \quad p < 0 \end{cases}$$

Where, for each value of p:

$$R_2^+ = \frac{1}{2}(A_2 + D_2) + \frac{i}{2}(C_2 - B_2) \qquad and \qquad R_2^- = \frac{1}{2}(A_2 - D_2) + \frac{i}{2}(C_2 + B_2)$$

$$tan(\theta_2^+) = \frac{C_2 - B_2}{A_2 + D_2} \qquad and \qquad tan(\theta_2^-) = \frac{C_2 + B_2}{A_2 - D_2}$$

The $'+'$ superscript represents the anti-clockwise rotating component, while the $'-'$ superscript represents the clockwise rotating component. Each of these vectors can be treated separately to obtain the equivalent results to those presented in Section 6.1.1. That is: the rotary spectra, rotary coefficient, ellipse orientation, rotary phase, ellipse stability and coherence squared.

For time series '1' we get:

Anti-clockwise and clockwise spectra, respectively:

$$|R_1^+| = \sqrt{R_1^+ \; R_1^{+*}} = \frac{1}{2}\sqrt{(A_1 + D_1)^2 + (C_1 - B_1)^2}$$

$$|R_1^-| = \sqrt{R_1^- \; R_1^{-*}} = \frac{1}{2}\sqrt{(A_1 - D_1)^2 + (C_1 + B_1)^2}$$

Major and minor ellipse axes are, respectively:

$$|R_1^+| + |R_1^-| = \frac{1}{2}\sqrt{(A_1 + D_1)^2 + (C_1 - B_1)^2} + \frac{1}{2}\sqrt{(A_1 - D_1)^2 + (C_1 + B_1)^2}$$

$$|R_1^+| - |R_1^-| = \frac{1}{2}\sqrt{(A_1 + D_1)^2 + (C_1 - B_1)^2} - \frac{1}{2}\sqrt{(A_1 - D_1)^2 + (C_1 + B_1)^2}$$

We define the following band averaged quantities:

$P_{u_1u_1} = \langle A_1^2 + B_1^2 \rangle$ auto-spectrum (u_1) squared

$P_{v_1v_1} = \langle C_1^2 + D_1^2 \rangle$ auto-spectrum (v_1) squared

$P_{u_1v_1} = \langle A_1C_1 + B_1D_1 \rangle$ coincident spectrum

$Q_{u_1v_1} = \langle A_1D_1 - C_1B_1 \rangle$ quadrature-spectrum

Then we can define the following in terms of these new quantities.

Rotary coherence squared (this is analogous to the coherence squared spectrum for the cross spectrum, Section 5.1):

$$\gamma_{R_1} = \frac{{P_{u_1v_1}}^2 + {Q_{u_1v_1}}^2}{P_{u_1u_1}\,P_{v_1v_1}}$$

Rotary phase (again, this is analogous to the phase difference spectrum for the cross spectrum, Section 5.1):

$$\theta_{R_1} = tan^{-1}\left(\frac{-Q_{u_1v_1}}{P_{u_1v_1}}\right)$$

Rotary coefficient:

$$C_{R_1} = \frac{-2\ Q_{u_1v_1}}{P_{u_1u_1} + P_{v_1v_1}}$$

Ellipse orientation:

$$\phi_{R_1} = \frac{1}{2} tan^{-1}\left(\frac{2\ P_{u_1v_1}}{P_{u_1u_1} - P_{v_1v_1}}\right)$$

For time series '2' we get:

Anti-clockwise and clockwise spectra, respectively:

$$|R_2^+| = \sqrt{R_2^+\ R_2^{+*}} = \frac{1}{2}\sqrt{(A_2 + D_2)^2 + (C_2 - B_2)^2}$$

$$|R_2^-| = \sqrt{R_2^-\ R_2^{-*}} = \frac{1}{2}\sqrt{(A_2 - D_2)^2 + (C_2 + B_2)^2}$$

Major and minor ellipse axes are, respectively:

$$|R_2^+| + |R_2^-| = \frac{1}{2}\sqrt{(A_2 + D_2)^2 + (C_2 - B_2)^2} + \frac{1}{2}\sqrt{(A_2 - D_2)^2 + (C_2 + B_2)^2}$$

$$|R_2^+| - |R_2^-| = \frac{1}{2}\sqrt{(A_2 + D_2)^2 + (C_2 - B_2)^2} - \frac{1}{2}\sqrt{(A_2 - D_2)^2 + (C_2 + B_2)^2}$$

Again, we define the following band averaged quantities:

$P_{u_2u_2} = \langle A_2^2 + B_2^2 \rangle$ auto-spectrum (u_2) squared

$P_{v_2v_2} = \langle C_2^2 + D_2^2 \rangle$ auto-spectrum (v_2) squared

$P_{u_2v_2} = \langle A_2C_2 + B_2D_2 \rangle$ coincident spectrum

$Q_{u_2v_2} = \langle A_2D_2 - C_2B_2 \rangle$ quadrature-spectrum

And we define the following in terms of these new quantities.

Rotary coherence squared (this is analogous to the coherence squared spectrum for the cross spectrum, Section 5.1):

$$\gamma_{R_2} = \frac{P_{u_2v_2}{}^2 + Q_{u_2v_2}{}^2}{P_{u_2u_2}\, P_{v_2v_2}}$$

Rotary phase (again, this is analogous to the phase difference spectrum for the cross spectrum, Section 5.1):

$$\theta_{R_2} = tan^{-1}\left(\frac{-Q_{u_2v_2}}{P_{u_2v_2}}\right)$$

Rotary coefficient:

$$C_{R_2} = \frac{-2\ Q_{u_2v_2}}{P_{u_2u_2} + P_{v_2v_2}}$$

Ellipse orientation:

$$\phi_{R_2} = \frac{1}{2} tan^{-1}\left(\frac{2\ P_{u_2v_2}}{P_{u_2u_2} - P_{v_2v_2}}\right)$$

With a pair of vectors, there are also the interactions between the different components of the two vectors. These interactions give us what is known and the 'inner' and 'outer' rotary components. The inner rotary components are obtained when we compare components of the two vectors that are rotating in the same direction, while the outer rotary

components result from comparing components of the two vectors that are rotating in opposite directions.

We define the Fourier transform of a pair of vectors, $R_j(p)$ and $R_k(p)$ below. (The 'j' vector might represent the winds and the 'k' vector might represent the currents). Both 'j' and 'k' have values of 1 or 2, representing the orthogonal (u and v) components of each vector. Thence:

$$R_j(p) = \begin{cases} |R_j^+|\, e^{i\theta_j^+} = |R_j^+|\, exp(i\theta_j^+) & for \quad p > 0 \\ |R_j^-|\, e^{-i\theta_j^-} = |R_j^-|\, exp(-i\theta_j^-) & for \quad p < 0 \end{cases}$$

$$R_k(p) = \begin{cases} |R_k^+| e^{i\theta_k^+} = |R_k^+|\, exp(i\theta_k^+) & for \quad p > 0 \\ |R_k^-|\, e^{-i\theta_k^-} = |R_k^-|\, exp(-i\theta_k^-) & for \quad p < 0 \end{cases}$$

Inner rotary components

Firstly, we'll look at the 'inner' rotary components; those components that are rotating in the same direction i.e. both $'+'$ and both $'-'$. They are:

$$R_j\, R_k^* = Inner_{jk} = \begin{cases} R_j^+ R_k^+\, exp\left[i\left(\theta_j^+ - \theta_k^+\right)\right] & for \quad p > 0 \\ R_j^- R_k^-\, exp\left[-i\left(\theta_j^- - \theta_k^-\right)\right] & for \quad p < 0 \end{cases}$$

The asterisk represents the complex conjugate of the vector.

If $j = k\ (= 1\ or\ 2)$ the result is the 'inner rotary auto energy spectra'. The imaginary parts of the vectors disappear because $\theta_j^+ = \theta_k^+$ and $\theta_j^- = \theta_k^-$ when $j = k$. Since j, k represent different vectors (e.g. j might represent a wind vector and k might represent a current vector), there are two inner rotary auto energy spectra.

$$Inner_{11} = \begin{cases} R_1^+ R_1^+ = (R_1^+)^2 & for \quad p > 0 \\ R_1^- R_1^- = (R_1^-)^2 & for \quad p < 0 \end{cases}$$

$$Inner_{22} = \begin{cases} R_2^+ R_2^+ = (R_2^+)^2 & for \quad p > 0 \\ R_2^- R_2^- = (R_2^-)^2 & for \quad p < 0 \end{cases}$$

If $j = 1, k = 2$ we get the 'inner rotary cross energy spectra'

$$Inner_{12} = \begin{cases} R_1^+ R_2^+ \, exp[i(\theta_1^+ - \theta_2^+)] & for \quad p > 0 \\ R_1^- R_2^- \, exp[-i(\theta_1^- - \theta_2^-)] & for \quad p < 0 \end{cases}$$

$$= \begin{cases} R_1^+ R_2^+ \, cos[(\theta_1^+ - \theta_2^+)] + i \, R_1^+ R_2^+ \, sin[(\theta_1^+ - \theta_2^+)] & for \; p > 0 \\ R_1^- R_2^- \, cos[(\theta_1^- - \theta_2^-)] - i \, R_1^- R_2^- \, sin[(\theta_1^- - \theta_2^-)] & for \; p < 0 \end{cases}$$

If $j = 2, and \; k = 1$ the result is simply the complex conjugate of the above (i.e. $Inner_{12} = Inner_{21}^*$) and no further information is gained by including both spectra.

$$Inner_{21} = \begin{cases} R_2^+ R_1^+ \, exp[i(\theta_2^+ - \theta_1^+)] & for \quad p > 0 \\ R_2^- R_1^- \, exp[-i(\theta_2^- - \theta_1^-)] & for \quad p < 0 \end{cases}$$

$$= \begin{cases} R_2^+ R_1^+ \, cos[(\theta_2^+ - \theta_1^+)] + i \, R_2^+ R_1^+ \, sin[(\theta_2^+ - \theta_1^+)] & for \; p > 0 \\ R_2^- R_1^- \, cos[(\theta_2^- - \theta_1^-)] - i \, R_2^- R_1^- \, sin[(\theta_2^- - \theta_1^-)] & for \; p < 0 \end{cases}$$

$$= \begin{cases} R_2^+ R_1^+ \, cos[(\theta_1^+ - \theta_2^+)] - i \, R_2^+ R_1^+ \, sin[(\theta_1^+ - \theta_2^+)] & for \; p > 0 \\ R_2^- R_1^- \, cos[(\theta_1^- - \theta_2^-)] + i \, R_2^- R_1^- \, sin[(\theta_1^- - \theta_2^-)] & for \; p < 0 \end{cases}$$

$$= complex \; conjugate \; of \; Inner_{12}$$

We can also determine four inner rotary phases, comprising two inner rotary auto phase spectra, $\phi_{11} \; and \; \phi_{22}$, and two inner rotary cross phase spectra, $\phi_{12} \; and \; \phi_{21}$. They are given by the following:

$$tan(\phi_{11}) = \begin{cases} \dfrac{\langle R_1^+ R_1^+ \, sin(\theta_1^+ - \theta_1^+) \rangle}{\langle R_1^+ R_1^+ \, cos(\theta_1^+ - \theta_1^+) \rangle} = 0 & for \quad p > 0 \\ \dfrac{\langle -R_1^- R_2^- \, sin(\theta_1^- - \theta_1^-) \rangle}{\langle R_1^- R_1^- \, cos(\theta_1^- - \theta_1^-) \rangle} = 0 & for \quad p < 0 \end{cases}$$

$$tan(\phi_{22}) = \begin{cases} \dfrac{\langle R_2^+ R_2^+ \, sin(\theta_2^+ - \theta_2^+) \rangle}{\langle R_2^+ R_2^+ \, cos(\theta_2^+ - \theta_2^+) \rangle} = 0 & for \quad p > 0 \\ \dfrac{\langle -R_2^- R_2^- \, sin(\theta_2^- - \theta_2^-) \rangle}{\langle R_2^- R_2^- \, cos(\theta_2^- - \theta_2^-) \rangle} = 0 & for \quad p < 0 \end{cases}$$

$$tan(\phi_{12}) = \begin{cases} \dfrac{\langle R_1^+ R_2^+ \, sin(\theta_1^+ - \theta_2^+) \rangle}{\langle R_1^+ R_2^+ \, cos(\theta_1^+ - \theta_2^+) \rangle} & for \quad p > 0 \\ \dfrac{\langle -R_1^- R_2^- \, sin(\theta_1^- - \theta_2^-) \rangle}{\langle R_1^- R_2^- \, cos(\theta_1^- - \theta_2^-) \rangle} & for \quad p < 0 \end{cases}$$

$$tan(\phi_{21}) = \begin{cases} \dfrac{\langle R_2^+ R_1^+ \, sin(\theta_2^+ - \theta_1^+)\rangle}{\langle R_2^+ R_1^+ \, cos(\theta_2^+ - \theta_1^+)\rangle} & for \quad p > 0 \\ \dfrac{\langle -R_2^- R_1^- \, sin(\theta_2^- - \theta_1^-)\rangle}{\langle R_2^- R_1^- \, cos(\theta_2^- - \theta_1^-)\rangle} & for \quad p < 0 \end{cases}$$

Where $\langle \quad \rangle$ represents a band average (i.e. an average in frequency space). The two inner rotary auto phase spectra [i.e. $tan(\phi_{11})$ and $tan(\phi_{22})$] are identically zero, while the two inner rotary cross phase spectra are complex conjugates. Hence, only one inner rotary phase spectrum is non-zero and unique. Generally, we use the inner rotary cross phase defined by $tan(\phi_{12})$.

Emery and Thompson (1998) point out that, if $R_1^+ R_2^+$ and $R_1^- R_2^-$ 'do not change much' over the band average, then

$$\phi_{12} \approx \begin{cases} \theta_1^+ - \theta_2^+ & for \quad p > 0 \\ \theta_1^- - \theta_2^- & for \quad p < 0 \end{cases}$$

Similarly, we can also determine four inner rotary coherence squared spectra, comprising two inner rotary auto coherence squared spectra, γ_{11} *and* γ_{22}, and two inner rotary cross coherence squared spectra, γ_{12} *and* γ_{21}. They are given by the following:

$$\gamma_{11}^2 = \begin{cases} \dfrac{\langle R_1^+ R_1^+ \, cos(\theta_1^+ - \theta_1^+)\rangle^2 + \langle R_1^+ R_1^+ \, sin(\theta_1^+ - \theta_1^+)\rangle^2}{\langle (R_1^+)^2\rangle \, \langle (R_1^+)^2\rangle} = 1 & for \quad p > 0 \\ \dfrac{\langle R_1^- R_1^- \, cos(\theta_1^- - \theta_1^-)\rangle^2 + \langle R_1^- R_1^- \, sin(\theta_1^- - \theta_1^-)\rangle^2}{\langle (R_1^-)^2\rangle \, \langle (R_1^-)^2\rangle} = 1 & for \quad p < 0 \end{cases}$$

$$\gamma_{22}^2 = \begin{cases} \dfrac{\langle R_2^+ R_2^+ \, cos(\theta_2^+ - \theta_2^+)\rangle^2 + \langle R_2^+ R_2^+ \, sin(\theta_2^+ - \theta_2^+)\rangle^2}{\langle (R_2^+)^2\rangle \, \langle (R_2^+)^2\rangle} = 1 & for \quad p > 0 \\ \dfrac{\langle R_2^- R_2^- \, cos(\theta_2^- - \theta_2^-)\rangle^2 + \langle R_2^- R_2^- \, sin(\theta_2^- - \theta_2^-)\rangle^2}{\langle (R_2^-)^2\rangle \, \langle (R_2^-)^2\rangle} = 1 & for \quad p < 0 \end{cases}$$

$$\gamma_{12}^2 = \begin{cases} \dfrac{\langle R_1^+ R_2^+ \cos(\theta_1^+ - \theta_2^+)\rangle^2 + \langle R_1^+ R_2^+ \sin(\theta_1^+ - \theta_2^+)\rangle^2}{\langle (R_1^+)^2\rangle \ \langle (R_2^+)^2\rangle} & for \quad p > 0 \\ \dfrac{\langle R_1^- R_2^- \cos(\theta_1^- - \theta_2^-)\rangle^2 + \langle R_1^- R_2^- \sin(\theta_1^- - \theta_2^-)\rangle^2}{\langle (R_1^-)^2\rangle \ \langle (R_2^-)^2\rangle} & for \quad p < 0 \end{cases}$$

$$\gamma_{21}^2 = \begin{cases} \dfrac{\langle R_2^+ R_1^+ \cos(\theta_2^+ - \theta_1^+)\rangle^2 + \langle R_2^+ R_1^+ \sin(\theta_2^+ - \theta_1^+)\rangle^2}{\langle (R_2^+)^2\rangle \ \langle (R_1^+)^2\rangle} & for \quad p > 0 \\ \dfrac{\langle R_2^- R_1^- \cos(\theta_2^- - \theta_1^-)\rangle^2 + \langle R_2^- R_1^- \sin(\theta_2^- - \theta_1^-)\rangle^2}{\langle (R_2^-)^2\rangle \ \langle (R_1^-)^2\rangle} & for \quad p < 0 \end{cases}$$

The two inner rotary auto coherence squared spectra (i.e. γ_{11}^2 and γ_{22}^2) are identically unity, while the two inner rotary cross coherence squared spectra are complex conjugates. Hence, only one inner rotary coherence squared spectrum is non-trivial and unique. Generally, we use the inner rotary cross coherence squared defined by γ_{12}^2.

Outer rotary components

We now look at those components that are rotating in opposite directions i.e. the $'+'$ component of one vector and the $'-'$ component of the second vector. These result in the 'outer' rotary components.

$$R_j \, R_k^* = Outer_{jk} = \begin{cases} R_j^- R_k^+ \, exp\left[i\left(\theta_k^+ - \theta_j^-\right)\right] & for \quad p > 0 \\ R_j^+ R_k^- \, exp\left[i\left(\theta_j^+ - \theta_k^-\right)\right] & for \quad p < 0 \end{cases}$$

If $j = k \ (= 1 \ or \ 2)$ the result is the 'outer rotary auto energy spectra'. Note that j, k are for each vector (e.g. j representing the wind vector and k representing the current vector).

$$Outer_{11} = \begin{cases} R_1^- R_1^+ \, exp[i(\theta_1^+ - \theta_1^-)] & for \quad p > 0 \\ R_1^+ R_1^- \, exp[i(\theta_1^+ - \theta_1^-)] & for \quad p < 0 \end{cases}$$

$$= \begin{cases} R_1^- R_1^+ \cos[(\theta_1^+ - \theta_1^-)] + i \, R_1^- R_1^+ \sin[(\theta_1^+ - \theta_1^-)] & for \quad p > 0 \\ R_1^+ R_1^- \cos[(\theta_1^+ - \theta_1^-)] + i \, R_1^+ R_1^- \sin[(\theta_1^+ - \theta_1^-)] & for \quad p < 0 \end{cases}$$

$$Outer_{22} = \begin{cases} R_2^- R_2^+ \, exp[i(\theta_2^+ - \theta_2^-)] & for \quad p > 0 \\ R_2^+ R_2^- \, exp[i(\theta_2^+ - \theta_2^-)] & for \quad p < 0 \end{cases}$$

$$= \begin{cases} R_2^- R_2^+ \, cos[(\theta_2^+ - \theta_2^-)] + i \, R_2^- R_2^+ \, sin[(\theta_2^+ - \theta_2^-)] & for \quad p > 0 \\ R_2^+ R_2^- \, cos[(\theta_2^+ - \theta_2^-)] + i \, R_2^+ R_2^- \, sin[(\theta_2^+ - \theta_2^-)] & for \quad p < 0 \end{cases}$$

The component for the negative frequencies $(p < 0)$ is exactly the same as that for the positive frequencies $(p > 0)$. The two outer auto energy spectra are symmetric about $p = 0$.

If $j = 1, k = 2$ we get the 'outer rotary cross energy spectrum'

$$Outer_{12} = \begin{cases} R_1^- R_2^+ \, exp[i(\theta_2^+ - \theta_1^-)] & for \quad p > 0 \\ R_1^+ R_2^- \, exp[i(\theta_1^+ - \theta_2^-)] & for \quad p < 0 \end{cases}$$

$$= \begin{cases} R_1^- R_2^+ \, cos[(\theta_2^+ - \theta_1^-)] + i \, R_1^- R_2^+ \, sin[(\theta_2^+ - \theta_1^-)] & for \quad p > 0 \\ R_1^+ R_2^- \, cos[(\theta_1^+ - \theta_2^-)] + i \, R_1^+ R_2^- \, sin[(\theta_1^+ - \theta_2^-)] & for \quad p < 0 \end{cases}$$

If $j = 2, k = 1$ the result is the 'frequency inverse' of above (i.e. the two outer cross energy spectra are reflections about $p = 0$. The outer cross energy spectrum for positive frequencies is the same as that for negative frequencies and vice versa.

$$Outer_{21} = \begin{cases} R_2^- R_1^+ \, exp[i(\theta_1^+ - \theta_2^-)] & for \quad p > 0 \\ R_2^+ R_1^- \, exp[i(\theta_2^+ - \theta_1^-)] & for \quad p < 0 \end{cases}$$

$$= \begin{cases} R_2^- R_1^+ \, cos[(\theta_1^+ - \theta_2^-)] + i \, R_2^- R_1^+ \, sin[(\theta_1^+ - \theta_2^-)] & for \quad p > 0 \\ R_2^+ R_1^- \, cos[(\theta_2^+ - \theta_1^-)] + i \, R_2^+ R_1^- \, sin[(\theta_2^+ - \theta_1^-)] & for \quad p < 0 \end{cases}$$

Again, no further information is gained by including both outer cross energy spectra. Generally, we use the $Outer_{12}$ to define the outer rotary cross energy spectrum.

Four outer rotary phases can be defined: two outer rotary auto phase spectra, $\psi_{11} \, and \, \psi_{22}$, and two outer rotary cross phase spectra, $\psi_{12} \, and \, \psi_{21}$. They are given by the following:

$$tan(\psi_{11}) = \begin{cases} \dfrac{\langle R_1^- R_1^+ \, sin(\theta_1^- - \theta_1^+) \rangle}{\langle R_1^- R_1^+ \, cos(\theta_1^- - \theta_1^+) \rangle} & for \quad p > 0 \\ \dfrac{\langle R_1^+ R_1^- \, sin(\theta_1^- - \theta_1^+) \rangle}{\langle R_1^+ R_1^- \, cos(\theta_1^- - \theta_1^+) \rangle} & for \quad p < 0 \end{cases}$$

$$tan(\psi_{22}) = \begin{cases} \dfrac{\langle R_2^- R_2^+ \, sin(\theta_2^- - \theta_2^+)\rangle}{\langle R_2^- R_2^+ \, cos(\theta_2^- - \theta_2^+)\rangle} & for \quad p > 0 \\ \dfrac{\langle R_2^+ R_2^- \, sin(\theta_2^- - \theta_2^+)\rangle}{\langle R_2^+ R_2^- \, cos(\theta_2^- - \theta_2^+)\rangle} & for \quad p < 0 \end{cases}$$

$$tan(\psi_{12}) = \begin{cases} \dfrac{\langle R_1^- R_2^+ \, sin(\theta_1^- - \theta_2^+)\rangle}{\langle R_1^- R_2^+ \, cos(\theta_1^- - \theta_2^+)\rangle} & for \quad p > 0 \\ \dfrac{\langle R_1^+ R_2^- \, sin(\theta_2^- - \theta_1^+)\rangle}{\langle R_1^+ R_2^- \, cos(\theta_2^- - \theta_1^+)\rangle} & for \quad p < 0 \end{cases}$$

$$tan(\psi_{21}) = \begin{cases} \dfrac{\langle R_2^- R_1^+ \, sin(\theta_2^- - \theta_1^+)\rangle}{\langle R_2^- R_1^+ \, cos(\theta_2^- - \theta_1^+)\rangle} & for \quad p > 0 \\ \dfrac{\langle R_2^+ R_1^- \, sin(\theta_1^- - \theta_2^+)\rangle}{\langle R_2^+ R_1^- \, cos(\theta_1^- - \theta_2^+)\rangle} & for \quad p < 0 \end{cases}$$

Where $\langle \quad \rangle$ represents an average in frequency space (i.e. a band average). For the two outer rotary auto phase spectra, ψ_{11} *and* ψ_{22}, we note that the spectral estimates for negative frequencies ($p < 0$) are identical to those for positive frequencies ($p > 0$). Hence, it is only necessary to define the outer rotary auto phase spectra for positive frequencies. Further, the negative frequencies associated with $tan(\psi_{12})$ equal the positive frequencies associated with $tan(\psi_{21})$ and vice versa. Therefore, only one of $tan(\psi_{12})$ and $tan(\psi_{21})$ are required. Generally, the former is used.

Again, Emery and Thompson (1998) point out that if $R_1^- R_2^+$ and $R_1^+ R_2^-$ 'do not change much' over the band average, then

$$\psi_{12} \approx \begin{cases} \theta_1^- - \theta_2^+ & for \quad p > 0 \\ \theta_2^- - \theta_1^+ & for \quad p < 0 \end{cases}$$

Four outer rotary coherence squared spectra can be defined: two outer rotary auto coherence squared spectra, λ_{11}^2 *and* λ_{22}^2, and two outer rotary cross coherence squared spectra, λ_{12}^2 *and* λ_{21}^2. They are given by the following:

$$\lambda_{11}^2 = \begin{cases} \dfrac{\langle R_1^- R_1^+ \rangle^2 [\langle cos(\theta_1^+ - \theta_1^-))^2 + \langle sin(\theta_1^+ - \theta_1^-))^2]}{\langle (R_1^+)^2 \rangle \ \langle (R_1^-)^2 \rangle} & for \ \ p > 0 \\ \dfrac{\langle R_1^+ R_1^- \rangle^2 [\langle cos(\theta_1^+ - \theta_1^-))^2 + \langle sin(\theta_1^+ - \theta_1^-))^2]}{\langle (R_1^+)^2 \rangle \ \langle (R_1^-)^2 \rangle} & for \ \ p < 0 \end{cases}$$

$$\lambda_{22}^2 = \begin{cases} \dfrac{\langle R_2^- R_2^+ \rangle^2 [\langle cos(\theta_2^+ - \theta_2^-))^2 + \langle sin(\theta_2^+ - \theta_2^-))^2]}{\langle (R_2^+)^2 \rangle \ \langle (R_2^-)^2 \rangle} & for \ \ p > 0 \\ \dfrac{\langle R_2^+ R_2^- \rangle^2 [\langle cos(\theta_2^+ - \theta_2^-))^2 + \langle sin(\theta_2^+ - \theta_2^-))^2]}{\langle (R_2^+)^2 \rangle \ \langle (R_2^-)^2 \rangle} & for \ \ p < 0 \end{cases}$$

$$\lambda_{12}^2 = \begin{cases} \dfrac{\langle R_1^- R_2^+ \rangle^2 [\langle cos(\theta_2^+ - \theta_1^-))^2 + \langle sin(\theta_2^+ - \theta_1^-))^2]}{\langle (R_2^+)^2 \rangle \ \langle (R_1^-)^2 \rangle} & for \ \ p > 0 \\ \dfrac{\langle R_1^+ R_2^- \rangle^2 [\langle cos(\theta_1^+ - \theta_2^-))^2 + \langle sin(\theta_1^+ - \theta_2^-))^2]}{\langle (R_1^+)^2 \rangle \ \langle (R_2^-)^2 \rangle} & for \ \ p < 0 \end{cases}$$

$$\lambda_{21}^2 = \begin{cases} \dfrac{\langle R_2^- R_1^+ \rangle^2 [\langle cos(\theta_1^+ - \theta_2^-))^2 + \langle sin(\theta_1^+ - \theta_2^-))^2]}{\langle (R_1^+)^2 \rangle \ \langle (R_2^-)^2 \rangle} & for \ \ p > 0 \\ \dfrac{\langle R_2^+ R_1^- \rangle^2 [\langle cos(\theta_2^+ - \theta_1^-))^2 + \langle sin(\theta_2^+ - \theta_1^-))^2]}{\langle (R_2^+)^2 \rangle \ \langle (R_1^-)^2 \rangle} & for \ \ p < 0 \end{cases}$$

For the two outer rotary auto coherence squared spectra, $\lambda_{11}^2 \ and \ \lambda_{22}^2$, we find that the spectral estimates for negative frequencies ($p < 0$) are identical to those for positive frequencies ($p > 0$). Hence, it is only necessary to define the outer rotary auto coherence squared spectra for positive frequencies. Further, the negative frequencies associated with λ_{12}^2 equal the positive frequencies associated with λ_{21}^2 and vice versa. Therefore, only one of λ_{12}^2 and λ_{21}^2 are required. Usually, the former is used.

Summary

We now have a set of inner and outer auto, cross, phase and coherency squared spectra that describe the relative motion of two vectors. Although the mechanics appears a little complex, all of these quantities comprise simple combinations of the real and imaginary components of the Fourier transforms of each time series.

The unique and non-trivial spectra comprise:

(a) Inner rotary spectra

The inner rotary auto energy spectra

$$Inner_{11} = \begin{cases} R_1^+ R_1^+ = (R_1^+)^2 & for \quad p > 0 \\ R_1^- R_1^- = (R_1^-)^2 & for \quad p < 0 \end{cases}$$

$$Inner_{22} = \begin{cases} R_2^+ R_2^+ = (R_2^+)^2 & for \quad p > 0 \\ R_2^- R_2^- = (R_2^-)^2 & for \quad p < 0 \end{cases}$$

The inner rotary cross energy spectrum

$$Inner_{12} = \begin{cases} R_1^+ R_2^+ \, exp[i(\theta_1^+ - \theta_2^+)] & for \quad p > 0 \\ R_1^- R_2^- \, exp[-i(\theta_1^- - \theta_2^-)] & for \quad p < 0 \end{cases}$$

The inner rotary cross phase spectrum

$$tan(\phi_{12}) = \begin{cases} \dfrac{\langle R_1^+ R_2^+ \, sin(\theta_1^+ - \theta_2^+) \rangle}{\langle R_1^+ R_2^+ \, cos(\theta_1^+ - \theta_2^+) \rangle} & for \quad p > 0 \\ \dfrac{\langle -R_1^- R_2^- \, sin(\theta_1^- - \theta_2^-) \rangle}{\langle R_1^- R_2^- \, cos(\theta_1^- - \theta_2^-) \rangle} & for \quad p < 0 \end{cases}$$

The inner rotary cross coherence squared spectrum

$$\gamma_{12}^2 = \begin{cases} \dfrac{\langle R_1^+ R_2^+ \, cos(\theta_1^+ - \theta_2^+) \rangle^2 + \langle R_1^+ R_2^+ \, sin(\theta_1^+ - \theta_2^+) \rangle^2}{\langle (R_1^+)^2 \rangle \, \langle (R_2^+)^2 \rangle} & for \;\; p > 0 \\ \dfrac{\langle R_1^- R_2^- \, cos(\theta_1^- - \theta_2^-) \rangle^2 + \langle R_1^- R_2^- \, sin(\theta_1^- - \theta_2^-) \rangle^2}{\langle (R_1^-)^2 \rangle \, \langle (R_2^-)^2 \rangle} & for \;\; p < 0 \end{cases}$$

(b) Outer rotary spectra

The outer rotary auto energy spectra (noting the reflection about $p = 0$ and only $p > 0$ needs to be specified).

$$Outer_{11} = R_1^- R_1^+ \, exp[i(\theta_1^+ - \theta_1^-)] \qquad for \;\; p > 0 \; and \; p < 0$$

$$Outer_{22} = R_2^- R_2^+ \, exp[i(\theta_2^+ - \theta_2^-)] \qquad for \;\; p > 0 \; and \; p < 0$$

The outer rotary cross energy spectrum

$$Outer_{12} = \begin{cases} R_1^- R_2^+ \, exp[i(\theta_2^+ - \theta_1^-)] & for \quad p > 0 \\ R_1^+ R_2^- \, exp[i(\theta_1^+ - \theta_2^-)] & for \quad p < 0 \end{cases}$$

The outer rotary auto phase spectra (again, noting that only values for $p > 0$ need to be specified)

$$tan(\psi_{11}) = \frac{\langle R_1^- R_1^+ \, sin(\theta_1^- - \theta_1^+)\rangle}{\langle R_1^- R_1^+ \, cos(\theta_1^- - \theta_1^+)\rangle} \qquad for \;\; p > 0 \; and \; p < 0$$

$$tan(\psi_{22}) = \frac{\langle R_2^- R_2^+ \, sin(\theta_2^- - \theta_2^+)\rangle}{\langle R_2^- R_2^+ \, cos(\theta_2^- - \theta_2^+)\rangle} \qquad for \;\; p > 0 \; and \; p < 0$$

The outer rotary cross phase spectrum

$$tan(\psi_{12}) = \begin{cases} \dfrac{\langle R_1^- R_2^+ \, sin(\theta_1^- - \theta_2^+)\rangle}{\langle R_1^- R_2^+ \, cos(\theta_1^- - \theta_2^+)\rangle} & for \quad p > 0 \\ \dfrac{\langle R_1^+ R_2^- \, sin(\theta_2^- - \theta_1^+)\rangle}{\langle R_1^+ R_2^- \, cos(\theta_2^- - \theta_1^+)\rangle} & for \quad p < 0 \end{cases}$$

The outer rotary auto coherence squared spectra (noting the reflection about $p = 0$ and only $p > 0$ needs to be specified).

$$\lambda_{11}^2 = \frac{\langle R_1^- R_1^+ \rangle^2 [\langle cos(\theta_1^+ - \theta_1^-)\rangle^2 + \langle sin(\theta_1^+ - \theta_1^-)\rangle^2]}{\langle (R_1^+)^2 \rangle \; \langle (R_1^-)^2 \rangle} \qquad for \;\; p > 0 \; and \; p < 0$$

$$\lambda_{22}^2 = \frac{\langle R_2^- R_2^+ \rangle^2 [\langle cos(\theta_2^+ - \theta_2^-)\rangle^2 + \langle sin(\theta_2^+ - \theta_2^-)\rangle^2]}{\langle (R_2^+)^2 \rangle \; \langle (R_2^-)^2 \rangle} \qquad for \;\; p > 0 \; and \; p < 0$$

The outer rotary cross coherence squared spectrum

$$\lambda_{12}^2 = \begin{cases} \dfrac{\langle R_1^- R_2^+ \rangle^2 [\langle cos(\theta_2^+ - \theta_1^-)\rangle^2 + \langle sin(\theta_2^+ - \theta_1^-)\rangle^2]}{\langle (R_2^+)^2 \rangle \ \langle (R_1^-)^2 \rangle} & for \quad p > 0 \\ \dfrac{\langle R_1^+ R_2^- \rangle^2 [\langle cos(\theta_1^+ - \theta_2^-)\rangle^2 + \langle sin(\theta_1^+ - \theta_2^-)\rangle^2]}{\langle (R_1^+)^2 \rangle \ \langle (R_2^-)^2 \rangle} & for \quad p < 0 \end{cases}$$

6.2. Examples of the rotary spectrum – a single vactor

In the examples involving vectors, data processing is carried out using the components of the vector. This is more convenient than using speed and direction.

Further, as in the previous examples, we'd ideally want about 5-10 cycles to effectively interpret features at the low frequency end of the spectrum. For example, if we have a year of data, the lowest frequency features that can be effectively interpreted is about 1-2 months. This differs from the maximum allowable resolution, which is one year.

At the high frequency end of the spectrum, we ideally want about 5 times the Nyquist frequency to effectively interpret features. For example, if we sample every 5 minutes (the Nyquist period is then 10 minutes), we generally restrict ourselves to interpreting features that are about one hour in duration. Again, this differs from the resolution, which is governed by the Nyquist period.

6.2.1. Example 6-1: A single current vector – hourly data

This example uses the same vector as used in Example 5-8 (Section 5.2.8) and Example 5-9 (Section 5.2.9). We'll repeat the example using the east-west and north-south components and also using the along-shore and across-shore components to illustrate those components that are invariant under rotation.

Questions: What do the rotary energy spectra look like? At what frequencies is the correlation strongest between the clockwise and anti-clockwise rotating current components? Do we get the same results when we rotate the axes?

Data used:

- The original data were sampled at 5 minute intervals.
- A 7-point Gaussian average was used to smooth the data, which were then subsampled to hourly intervals.
- The total length of the data set is 4,096 hours, or about 6 months.
- The Nyquist period is 2 hours (Nyquist frequency is ½ cycles per hour).
- The data set is about 170 days long, indicating the longest period features we can realistically interpret is about 20-40 days.
- This restricts any answer to our question to the frequency band bounded by 12 and (about) 0.05 cycles per day.

Pre-processing:

- The linear trend was removed from the data.
- The time series plots of the data are shown in Figure 6–3 in stick vector form. (For plotting purposes, the data were averaged into 3-hourly bins using box-car averaging). The length of each stick is proportional to the current speed, while the orientation of the sticks indicates the direction of the currents (using the oceanographic convention).
- Visually, the dominant currents are oriented slightly east of north, although there is considerable variation throughout the data record.

Testing:

- Stationarity: Results from the Dickey-Fuller test give Tau calculated values of -21.15 (for the east-west current anomalies) and -6.29 (for the north-south current component anomalies); the Tau critical value is - 2.86. As the calculated Tau is less than Tau critical for both data sets, we can reject the null hypothesis that the data sets are not stationary.
- Results from the statistical tests for stationarity indicate that the data were not obtained from the same, respective, populations. Noting this, we proceeded with the analysis.

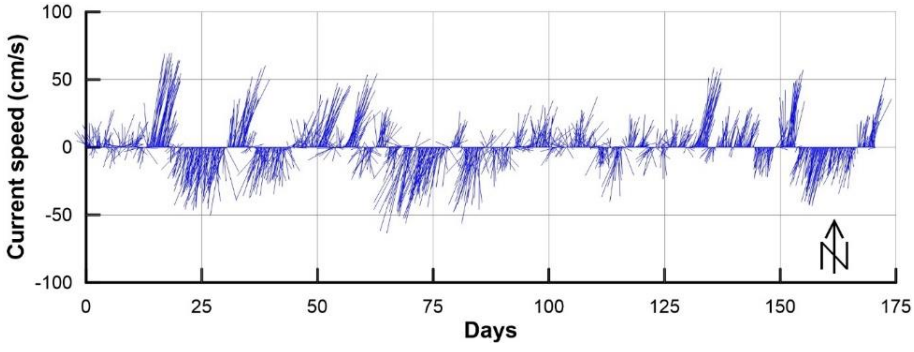


Figure 6–3. Time series of the currents in stick vector form.

Smoothing of the periodogram:

- Both periodograms were averaged as follows: 4 spectral bands were averaged over the first 200 spectral estimates, 8 spectral bands over the next 800 spectral estimates and 16 spectral bands were averaged for the remainder of both spectra. The frequency separating the regions where 4 and 8 spectral bands were averaged was 1.4 cycles per day and the frequency separating the regions where 8 and 16 spectral bands were averaged occurred at 3.2 cycles per day.
- Spectral averaging enabled confidence limits to be generated.
- The 95% confidence limits are used for the energy spectra, the stability spectrum and the coherence spectrum.
- No confidence limits for the other rotary spectra are presented.

Description: The resulting clockwise and anti-clockwise rotary spectral energy and the rotary coefficient are presented in Figure 6–4 (the bottom and top plots, respectively). The 95% confidence limits for the energy spectra are superimposed on the respective plots, as is the '-5/3^{rds} power law'. The stability, rotary phase and ellipse orientation spectra are presented in Figure 6–5 (bottom, middle and top plots, respectively.

Results from the rotary spectral energy for the two data sets (Figure 6–4, bottom plot), the rotary coefficient (Figure 6–4, middle plot) and the rotary coherence squared (Figure 6–4, top plot) indicate the following.

- At frequencies less than about one cycle per day, there is little difference in the energy of the clockwise and anti-clockwise rotary components. The rotary coefficient is close to zero, indicating that the flow is rectilinear (almost bi-directional). Referring back to the stick vector plot (Figure 6–3) we can see there is a dominant axis of flow, aligned along an axis slightly east of north.
- Between frequencies of 1 and 5 cycles per day, there is considerably more energy in the clockwise rotating spectrum than in the anti-clockwise rotating spectrum. The rotary coefficient lies between (approximately) 0.25 and 0.75, suggesting the flow (at these frequencies) moves in a clockwise direction, mapping out an ellipse (recall that a rotary coefficient close to zero indicates rectilinear flow, while a rotary coefficient of unity indicates circular flow).
- Large spectral peaks are observed at the diurnal and semi-diurnal periods (near 1 and 2 cycles per day), although the spectral energy at these frequencies is substantially lower than that at low frequencies.
- Rotary coherence squared lies above the 95% confidence limits at frequencies less than about 0.2 cycles per day, and at many specific higher frequencies, some of which correspond with the tidal signals.
- Both spectra generally follow the '-5/3^{rds} power law' between frequencies of 0.05 and 5 cycles per day. At frequencies greater than about 5 cycles per day, the spectral energy of the anti-clockwise rotating component flattens slightly.
- Between about 1 and 5 cycles per day, the clockwise rotating component contains considerably more energy than the anti-clockwise rotating component.

Results from the stability (Figure 6–5, bottom plot), rotary phase (Figure 6–5, middle plot) and ellipse orientation (Figure 6–5, top plot) indicate the following.

- Generally, the stability lies above the 95% confidence limits at frequencies less than about 1 cycle per day. At higher frequencies the stability often lies above the 95% confidence limits. The coherence between the two rotary components is greatest at low frequencies (less than about 1 cycle per day). There are specific frequencies (viz. 1, 2 and 3 cycles per day) at which there is large coherence between the two rotary components.
- The rotary phase is close to zero at low frequencies (less than about 0.3 cycles per day). That is: the starting position on the ellipse of the rotating current vector is at the point that it furthest from the ellipse centre. At higher frequencies the rotary phase is highly variable.
- The ellipse orientation lies close to 90° at low frequencies (less than about 2 cycles per day), and close to zero at higher frequencies.

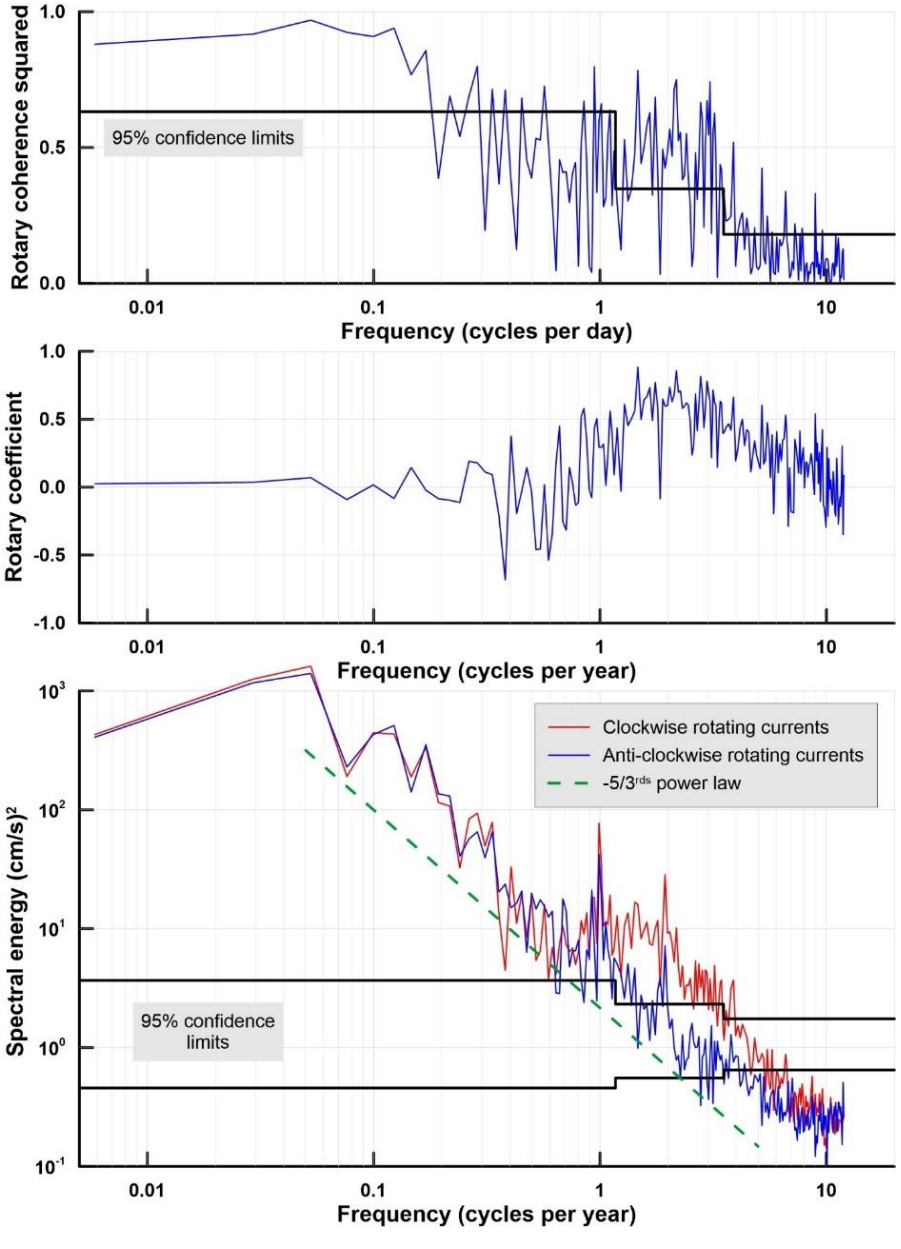


Figure 6–4. Periodogram of the clockwise and anti-clockwise rotary components, the rotary coefficient and the rotary coherence squared generated using the north-south and east-west current components.

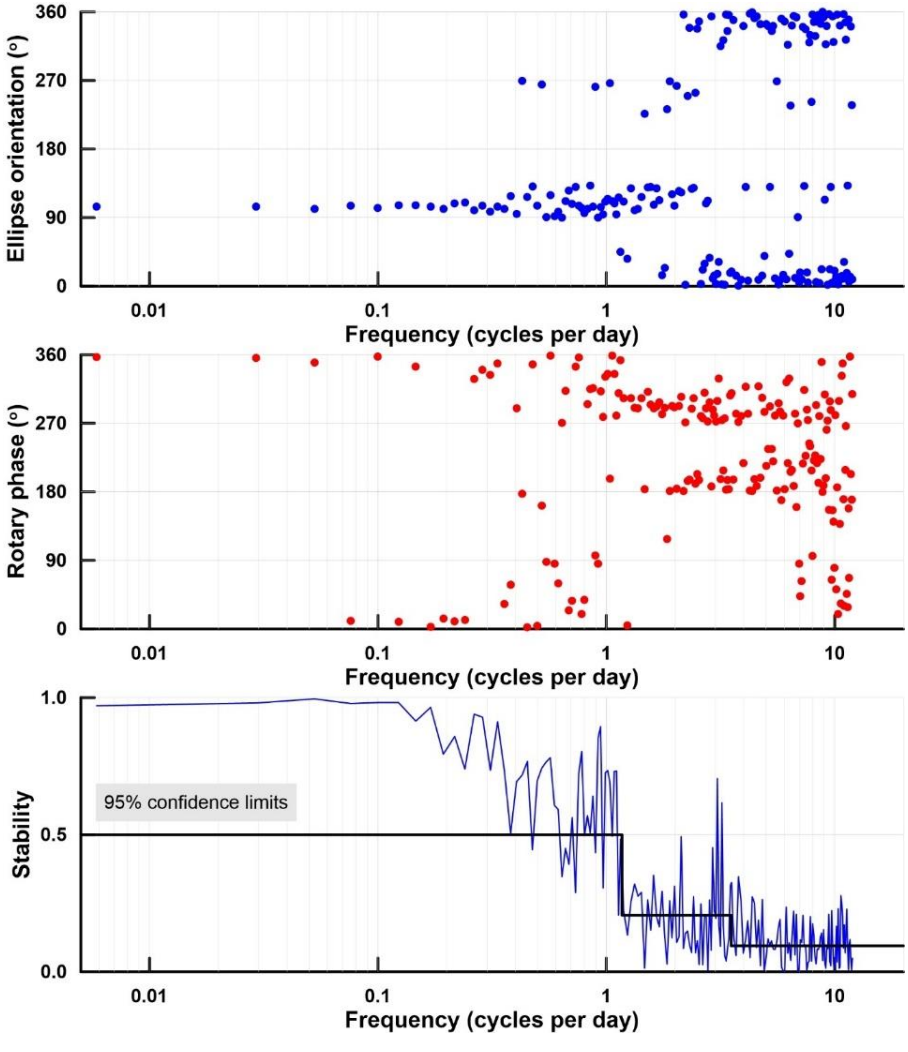


Figure 6–5. Stability, rotary phase and ellipse orientation spectra of the rotary coefficients generated from the north-south and east-west current components.

Sometimes it is more convenient to plot the rotary spectral energy as a 'two-sided' spectrum (Figure 6–6). Here, the frequency axis is linear (not the usual log axis). The anti-clockwise component has 'positive' frequencies and the clockwise component is associated with 'negative'

frequencies. The frequencies themselves are not negative, they are just written this way just to separate the clockwise and anti-clockwise components. The actual frequencies are the absolute values of the numbers on the frequency axis.

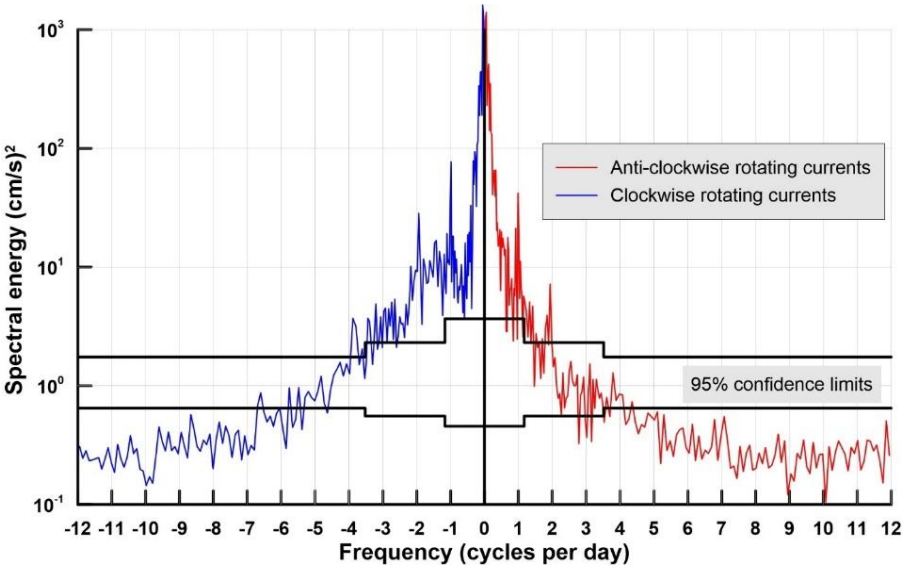


Figure 6–6. Periodogram of the clockwise (negative frequencies) and anti-clockwise (positive frequencies) rotary components and the rotary coefficient.

Returning to our original questions.

(a) For a single vector, several periodograms result: the clockwise and anti-clockwise rotating components of the energy spectra, the rotary coefficient, the rotary coherence squared, ellipse stability, rotary phase and ellipse orientation. This provides considerably more information (at least the information is presented in a different manner) than does the cross spectral analysis.

(b) The rotary coherence squared spectrum lies above the 95% confidence limits at frequencies less than about 0.2 cycles per day, between 1 and 4 cycles per day and at several higher frequencies.

This is essentially the same result as obtained for the cross spectral analysis (Example 5-8, Section 5.2.8).

(c) From the figures below, we note the following. The clockwise, anti-clockwise, rotary coefficient and stability and all invariant under axis rotation. However, the rotary phase and ellipse orientation do depend on the orientation of the axes chosen. So, we find that some rotary spectra are invariant under axis rotation, while others are not.

One of the advantages in using rotary spectra is that many of the spectra are invariant under rotation. Recall from Sections 5.2.8 and 5.2.9, that the results we obtained using cartesian co-ordinates depended on the selection of axes – different results were obtained when we resolved the axes into along-shore and across-shore components compared with east-west and north-south axes.

We'll repeat the analysis using the data from above, but with axes rotated by -45^{o}, to demonstrate those rotary spectral components that are invariant under axis rotation.

The set up of the data is identical to that for the unrotated data set above. Results from the Dickey-Fuller test give Tau calculated values of -11.57 (for one rotated current component anomalies) and -7.82 (for the second rotated current component anomalies); the Tau critical value is - 2.86. As the calculated Tau is less than Tau critical for both data sets, we can reject the null hypothesis that the data sets are not stationary.

Comparing Figure 6–4 with Figure 6–7 shows that the clockwise and anti-clockwise spectra and the rotary coefficient are identical. Similarly, comparing the stability spectra of Figure 6–5 and Figure 6–8 shows that they are also identical.

These quantities are all invariant under axis rotation.

The ellipse orientation of Figure 6–5 showed values lying close to 100° at low frequency and 0° at high frequencies. With the axes rotated by -45°, these values lie close to 250° and 45° / 135°, respectively. Allowing for an ambiguity of 180° in the orientation of the ellipse, these values differ by about 45°, the same angle by which the axes were rotated. The rotary phase in Figure 6–5 was close to zero at low frequencies (less than about 0.5 cycles per day) and near 90° / 270° in the rotated data set. At higher frequencies the phases become noisy and are difficult to interpret.

The ellipse orientation and rotary phase are not invariant under axis rotation. This is understandable as the ellipse orientation needs to be relative to something, which happens to be the axis about which the data are rotated. The rotary phase is relative to the orientation of the ellipse and will differ from the original values accordingly.

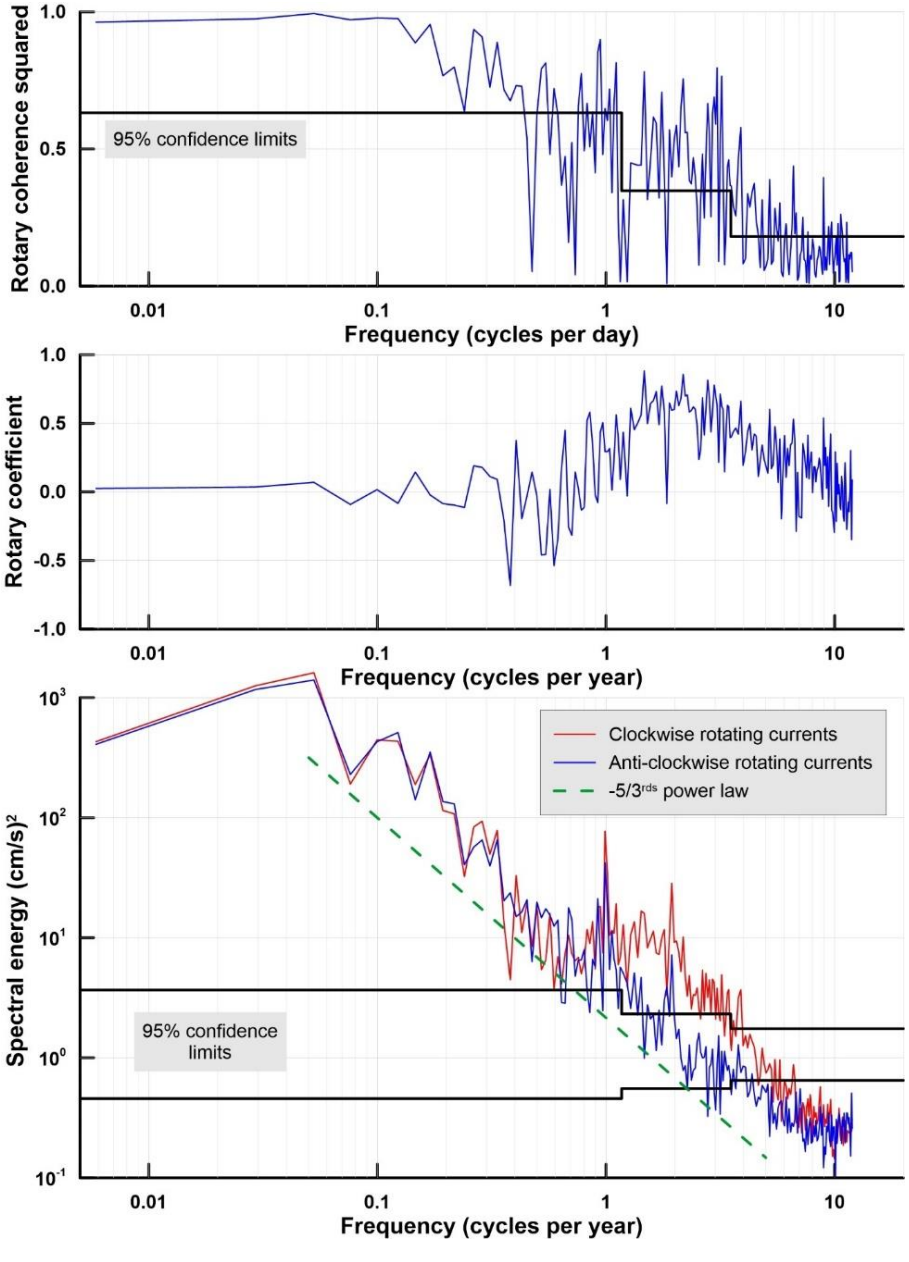


Figure 6–7. Periodogram of the clockwise and anti-clockwise rotary components, the rotary coefficient and the rotary coherence squared using axes rotated by - 45°.

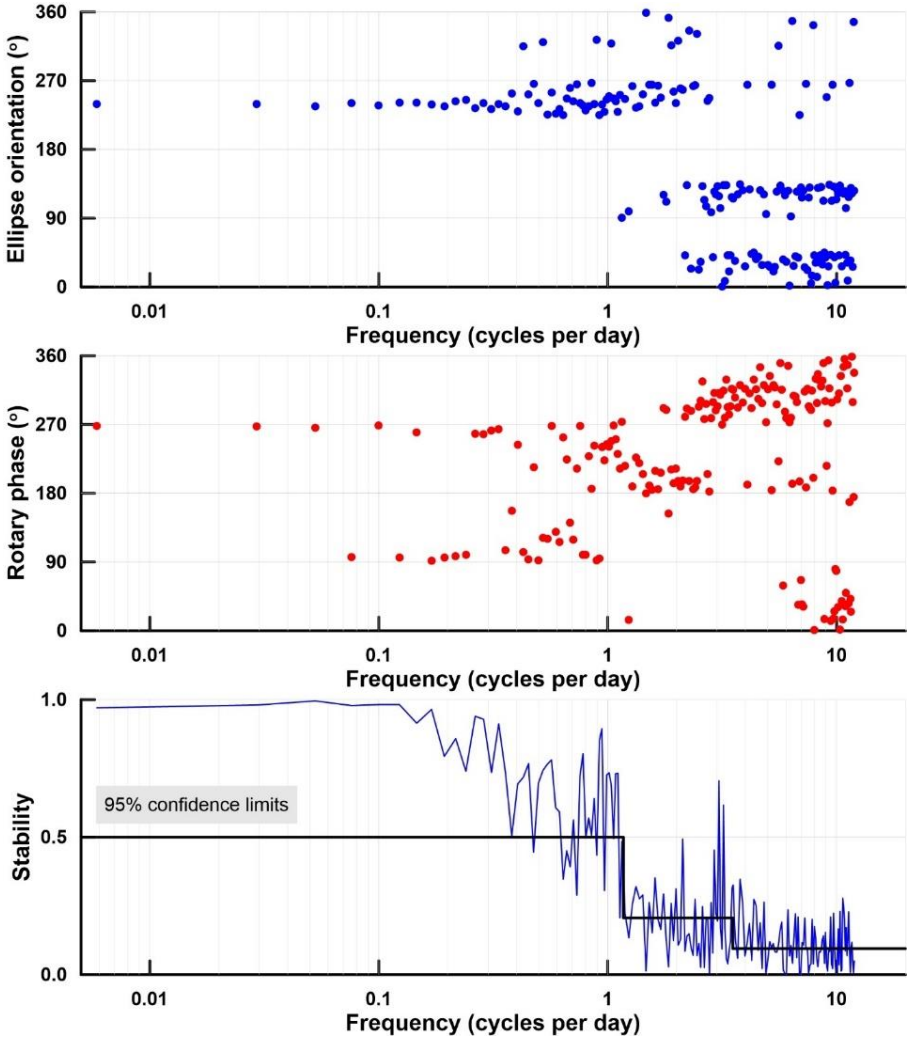


Figure 6–8. Stability, rotary phase and ellipse orientation spectra of the rotary coefficients using axes rotated by - 45º.

6.2.2. Example 6-2: A single current vector – 5 min data

This example examines the rotary spectra using high frequency ocean current speed and direction data.

Questions: At what frequencies is the correlation strongest between the clockwise and anti-clockwise rotating current components?

Data used:

- The current component data were sampled at 5 minute intervals.
- The Nyquist period is 10 minutes (Nyquist frequency is 6 cycles per hour).
- The rotary spectra are undertaken using 4,096 data points or about 14 days long. The longest period features we can realistically interpret is about 1-2 days.
- This restricts any answer to our question to the frequency band bounded by about 0.5 and 6 cycles per day.

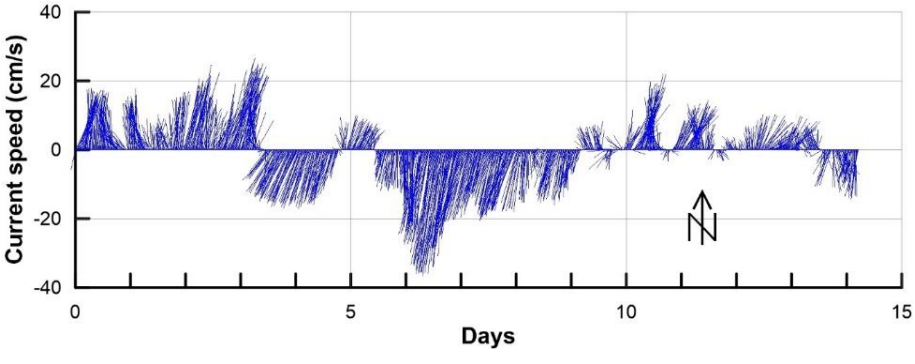


Figure 6–9. Time series of the currents in stick vector form.

Pre-processing:

- The linear trend was removed from the data.
- The time series plots of the data are shown in Figure 6–9 in stick vector form. The length of each stick is proportional to the

current speed, while the orientation of the sticks indicates the direction of the currents (using the oceanographic convention).

- Visually, the dominant currents are rotated slightly east of north, although there is considerable variation throughout the data record.
- Currents in the middle of the record flow towards the south, while those at either end of the record flow generally towards the north.

Testing:

- Stationarity: Results from the Dickey-Fuller test give Tau calculated values of -18.20 (for the east-west current anomalies) and -5.25 (for the north-south current component anomalies); the Tau critical value is - 2.86. As the calculated Tau is less than Tau critical for both data sets, we can reject the null hypothesis that the data sets are not stationary.
- Results from the statistical tests for stationarity indicate that the data were not obtained from the same, respective, populations. Noting this, we proceeded with the analysis.

Smoothing of the periodogram:

- Both periodograms were averaged: 4 spectral bands were averaged over the first 200 spectral estimates, 8 spectral bands over the next 800 spectral estimates and 16 spectral bands were averaged for the remainder of both spectra. The frequency separating the regions where 4 and 8 spectral bands were averaged was 1.4 cycles per day and the frequency separating the regions where 8 and 16 spectral bands were averaged occurred at 3.2 cycles per day.
- Spectral averaging enabled confidence limits to be generated.
- The 95% confidence limits are presented.

Description: The resulting clockwise and anti-clockwise rotary spectral energy and the rotary coefficient are presented in Figure 6–10 (the bottom and top plots, respectively). The 95% confidence limits for the energy spectra is superimposed on the respective plots, as is the '-5/3^{rds} power

law'. The stability, rotary phase and ellipse orientation spectra are presented in Figure 6–11 (bottom, middle and top plots, respectively.

Results from the rotary spectral energy for the two data sets (Figure 6–10, bottom plot), the rotary coefficient (Figure 6–10, middle plot) and the rotary coherence squared (Figure 6–10, top plot) indicate the following.

- Despite having removed the linear trend, the region of greatest spectral energy lies at the low frequency end of both spectra.
- There is little difference in the energy of the clockwise and anti-clockwise rotary components across the entire spectrum. The rotary coefficient is close to zero at low frequency, indicating that the flow is rectilinear (almost bi-directional). Referring back to the stick vector plot (Figure 6–9) we can see there is a dominant axis of flow, aligned along an axis slightly east of north.
- The rotary coefficient generally lies in the range +/-0.5 indicating that the flow across all frequencies looks more rectilinear / elliptical, rather than circular.
- The rotary coherence squared generally lies above the 95% confidence limit at frequencies less than about 1 cycle per day and at about 11 cycles per day (2 hour period). The latter may be in response to high frequency internal waves.
- The largest spectral peak occurs near 2 cycles per day, and there is a broad peak near 1 cycle per day. These are likely to be due to the semi-diurnal and diurnal tides, respectively.
- At frequencies less than about 10 cycles per day, both spectra generally follow the '$-5/3^{rds}$ power law'. At higher frequencies the spectral energy of both components flattens, suggesting an input of energy at higher frequencies. This would warrant further investigation.

Results from the stability (Figure 6–11, bottom plot), rotary phase (Figure 6–11, middle plot) and ellipse orientation (Figure 6–11, top plot) indicate the following.

- Generally, the stability lies above the 95% confidence limits at frequencies less than about 1 cycle per day. At higher frequencies

the stability lies above the 95% confidence limits only at selected frequencies, although many such frequencies are apparent.

- The rotary phase is close to zero at low frequencies (less than about 3 cycles per day). At higher frequencies, the rotary phase spectrum is noisy and difficult to interpret. However, the rotary phase is close to zero at those frequencies where the stability lies about the 95% confidence limit.
- The ellipse orientation lies near 100° at frequencies less than about 3 cycles per day. There is a certain symmetry about 180°, with four regions each of 45° in which no ellipse orientation information occurs. This results from the 180° 'ambiguity' in the ellipse. The orientation definition refers to the angle formed by the 'positive' semi-major ellipse axis with the horizontal (or 'x' or 'real') axis. The 'missing' information would be filled if the angle of the 'negative' semi-major ellipse axis and the negative horizontal axis was also determined.

Returning to the question we asked.

(a) The frequencies at which the rotary coherence squared spectrum are greatest are: frequencies less than about 1.2 cycles per day, near 12 cycles per day and at several isolated higher frequencies. These frequencies don't immediately correlate with expected processes and would likely require further investigation.

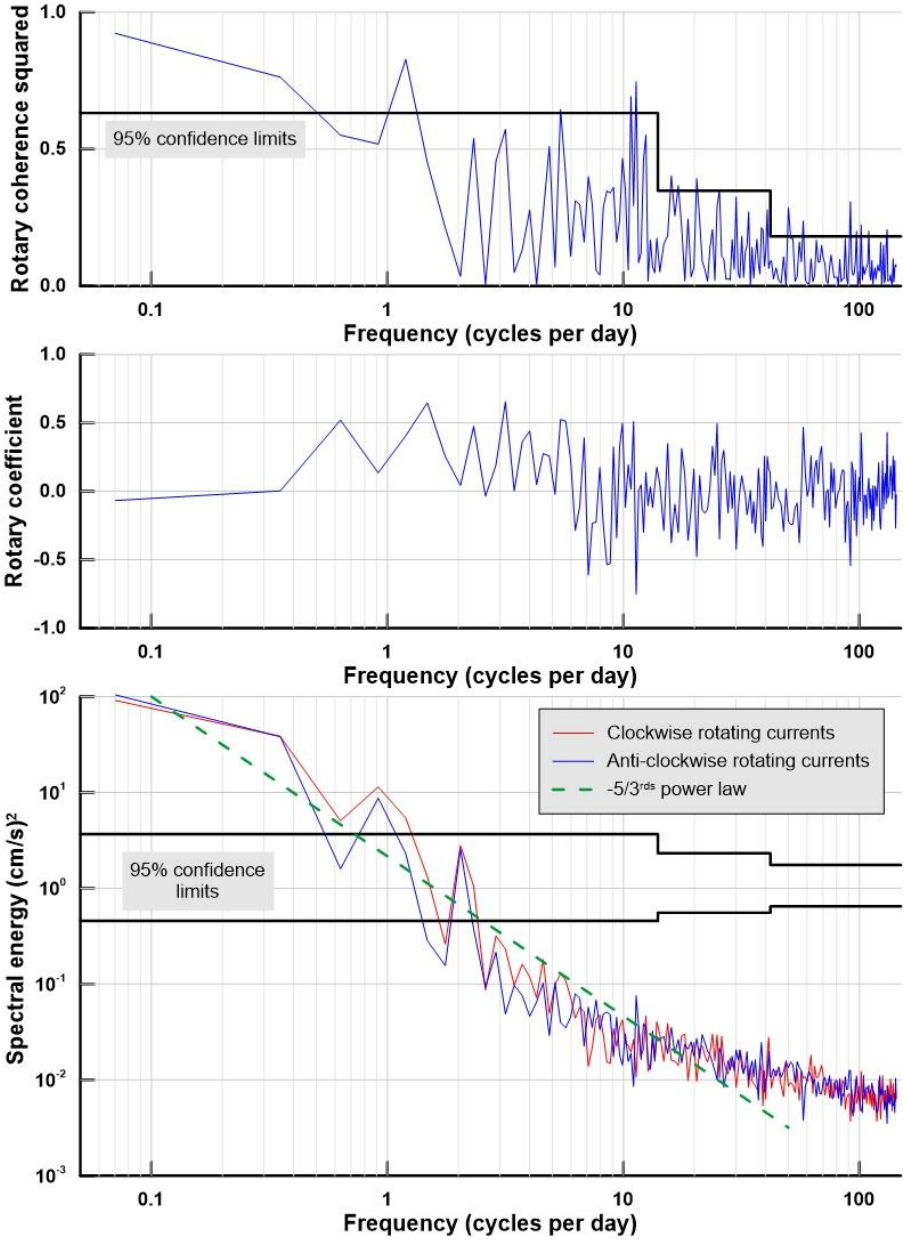


Figure 6–10. Periodogram of the clockwise and anti-clockwise rotary components, the rotary coefficient and rotary coherence squared generated using the north-south and east-west current components.

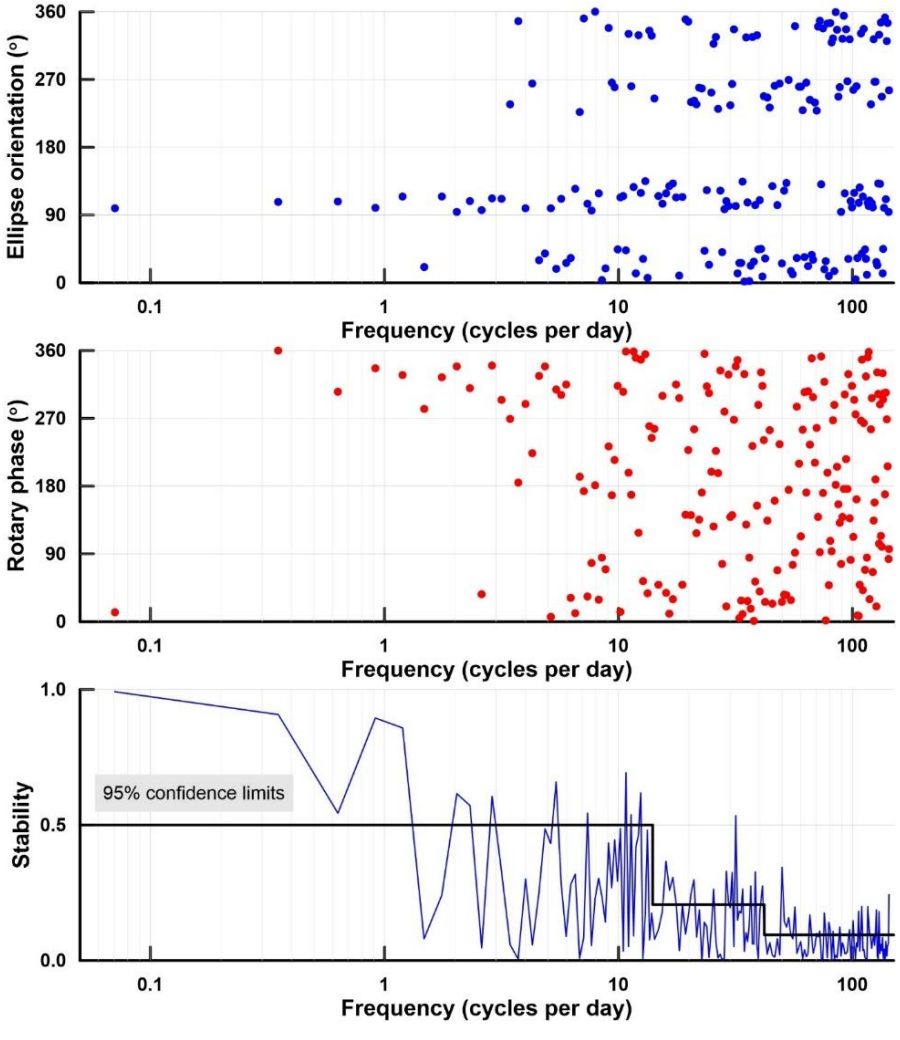


Figure 6–11. Stability, rotary phase and ellipse orientation spectra of the rotary coefficients generated from the north-south and east-west current components.

We can compare the rotary spectral energy (Figure 6–10, bottom plot) with the cartesian spectral energy (Figure 6–12), generated from the east-west and north-south current components. At frequencies less than about 3 cycles per day there is a discernible difference between the east-west and north-south spectra, that is not apparent in the rotary spectra. The high frequency end of both the cartesian and rotary spectra show similar behaviour.

Whether one chooses rotary or cartesian spectral analysis is a matter of personal preference. Provided we can judiciously choose an axis rotation (i.e. if we can clearly identify a dominant axis or direction of water motion), the cartesian spectra can produce 'maximum' and 'minimum' energy spectra. In general, we find the results obtained using the cartesian spectra a little easier to interpret. The rotary approach does not depend on the choice of axes and does provide additional information to assist in data interpretation.

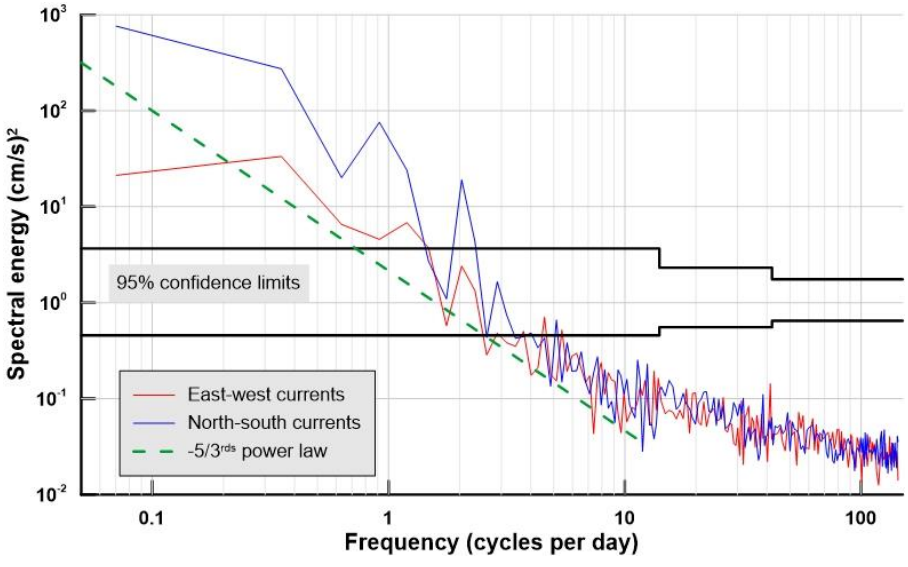


Figure 6–12. Periodogram of east-west and north-south components of the current.

6.2.3. Example 6-3: A single current vector – 10 day data

This example examines the rotary spectra using very low frequency (tides and inertial periods removed) ocean current speed and direction data.

Questions: At what frequencies is the correlation strongest between the clockwise and anti-clockwise rotating current components? Are there substantial differences when data are padded?

Data used:

- The original data were sampled at 5 minute intervals.
- A 7-point Gaussian filter was used to smooth the data, which were then subsampled to hourly intervals.
- Hourly data were then subject to the 120i913 filter (Thompson, 1983) to remove features with tidal and inertial periods. This filter is 271 hours long hence the data were subsampled at 10 day intervals. (Theoretically, we should subsample at 11.3 days or longer, but '10 days is convenient').
- The result is a data set 462 records long (i.e. 4,620 days).
- To use the FFT algorithm we will need to pad the data by adding 25 data points to each end of the data set (or we could truncate the data set to give two sets each with 256 records with some overlap, or use the DFT). Later, we'll use all three approaches and compare the results.
- The Nyquist period is 20 days (Nyquist frequency is 0.05 cycles per day or about 18 cycles per year).
- We have about 12.5 years of data so the longest period features we can realistically interpret is about 1-2 years.
- Answers to our questions will be restricted to the frequency band bounded by about 0.5 and 2 cycles per year.

Pre-processing:

- The time series plots of the data are shown in Figure 6–13 in stick vector form. The length of each stick is proportional to the current speed, while the orientation of the sticks indicates the direction of the currents (using the oceanographic convention).

- Visually, the dominant currents are rotated slightly east of north, although there is considerable variation throughout the data record.
- Currents flow mainly towards the south although northerly flows are often experienced for short time periods.
- Twenty-five random values (generated using the mean and variance of the available data) were added to each end of the data set to bring the total number of data records to 512, and enable the FFT to be used. We could have added zeroes instead of random values, but we did this to show that there are different ways in which the padding can be done.
- The linear trend was removed from the 10 day data to produce anomalies.
- Two further analyses were undertaken. One in which the data were also divided into two sets each of length 256 records (with an overlap of 50 records between the two data sets). The second analysis uses the DFT with 462 records.

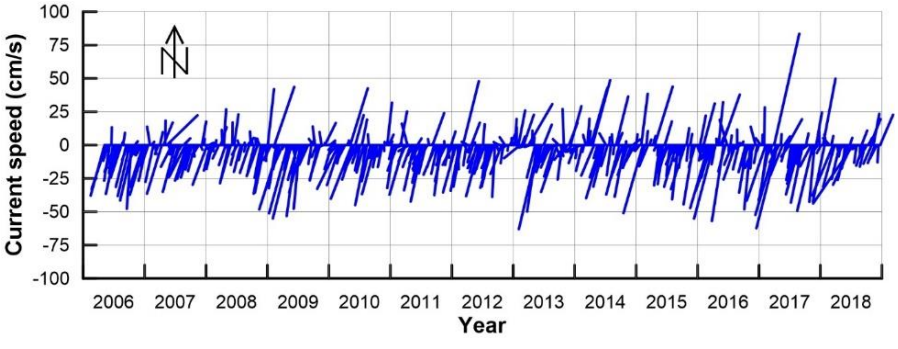


Figure 6–13. Time series of the currents in stick vector form.

Testing:

- Stationarity: Results from the Dickey-Fuller test give Tau calculated values of -19.73 (for the east-west current anomalies) and -18.40 (for the north-south current component anomalies); the Tau critical value is - 2.86. As the calculated Tau is less than

Tau critical for both data sets, we can reject the null hypothesis that the data sets are not stationary.

- To undertake the tests for stationarity, the data were split into 16 groups each comprising 32 data points. Bartlett's test for homogeneity of variance resulted in calculated values of 18.6 and 29.9 for the east-west and north-south anomalies, respectively. The critical value at the 1% level is 30.6 and, at the 5% level is 25.0. Therefore, we conclude that the 16 data 'east-west' sets are from the same population at the 1% and 5% levels. The 'north-south' data sets are from the same populations at the 1% level but not at the 5% level. The calculated ANOVA values were 0.59 and 1.00 for the 'east-west' and 'north-south' data sets, respectively. The critical values at the 1% level and 5% level of significance are 0.35 and 0.48 respectively. Therefore, we conclude that the means for each data set are not from the same populations. Strictly, these statistical tests indicate that the data are not stationary. However, recall from Section 3.8 that even a data set comprising randomly generated values does not always meet the stationarity criteria. The analysis in this example does not worse than that for a randomly generated data set and we would be comfortable in saying that this data set is close to stationary. Again, we note these results and proceeded with the analysis.

Smoothing of the periodogram:

- Both periodograms were averaged as follows: 4 spectral bands were averaged over the entire spectrum.
- Spectral averaging enabled confidence limits to be generated.
- The 95% confidence limits are shown.

Description: The resulting clockwise and anti-clockwise rotary spectral energy, the rotary coefficient and rotary coherence squared are presented in Figure 6–14 (the bottom, middle and top plots, respectively). The 95% confidence limits for the energy spectra is superimposed on the respective plots, as is the '-5/3^{rds} power law'. The stability, rotary phase and ellipse orientation spectra are presented in Figure 6–15 (bottom, middle and top plots, respectively.

Results indicate the following.

- The rotary energy spectra are both relatively flat (Figure 6–14, bottom plot) indicating that these data are random-like. The tests for stationarity (above) indicated that the data were 'almost' stationary (in so far as the testing gave similar results to those obtained for a random spectrum, Section 3.8). Neither rotary energy spectra follow the '$-5/3^{rds}$ power law'. This suggests that the linear trend, the tides and the inertial period features (all of which were removed in the data processing) are the major driving forces in these data sets.
- There is little difference in the energy of the clockwise and anti-clockwise rotary components across the entire spectrum.
- The rotary coefficient (Figure 6–14, middle plot) is close to zero across all frequencies, indicating that the flow is rectilinear (almost bi-directional). Referring back to the stick vector plot (Figure 6–13) we can see there is a dominant axis of flow, aligned along an axis running slightly east of north to slightly west of south.
- The rotary coherence squared spectrum (Figure 6–14, top plot) generally lies about the 95% confidence limit across the entire spectrum. The correlation between the clockwise and anti-clockwise rotating components is strong. This is somewhat self-evident from the spectral energy plots which show both components virtually the same.

Results from the stability (Figure 6–15, bottom plot), rotary phase (Figure 6–15, middle plot) and ellipse orientation (Figure 6–15, top plot) indicate:

- The stability always lies above the 95% confidence limits. This suggests considerable confidence in the ellipse orientation.
- The rotary phase is close to zero across all frequencies.
- The ellipse orientation lies near 100^{o} across all frequencies. The ellipse orientation is defined as 'positive being anti-clockwise'.

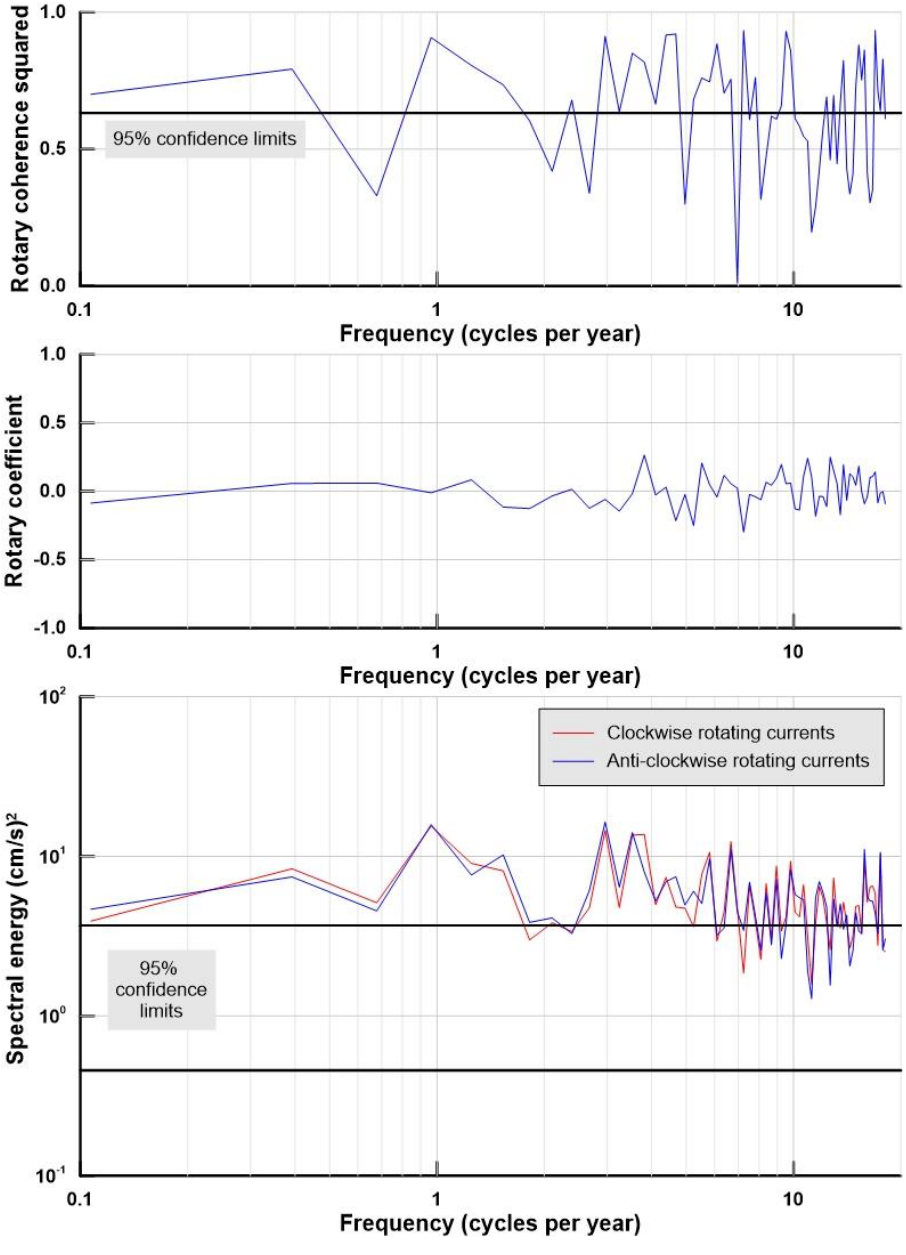


Figure 6–14. Periodogram of the clockwise and anti-clockwise rotary components, the rotary coefficient and the rotary coherence squared generated using the north-south and east-west current components.

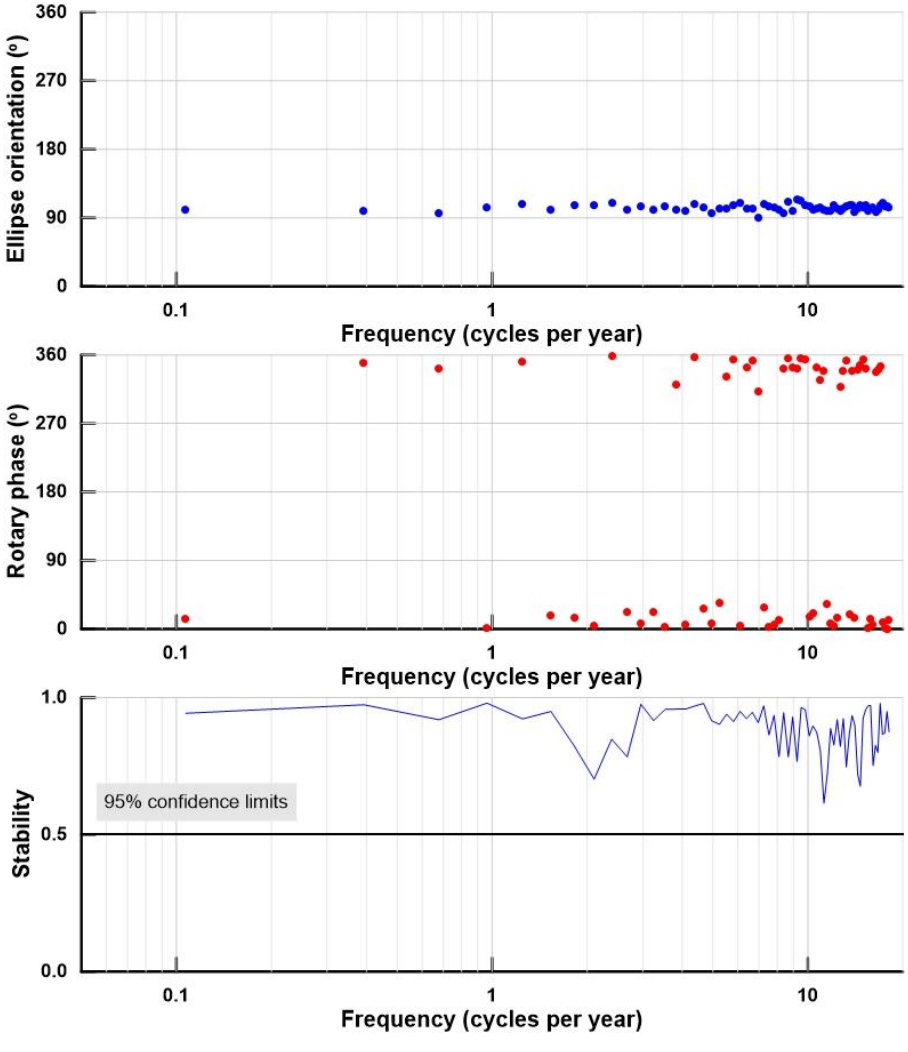


Figure 6–15. Stability, rotary phase and ellipse orientation spectra of the rotary coefficients generated from the north-south and east-west current components.

Above, we stated that there were three options we could take: (a) pad the data to bring the length the 512 records and used the FFT, (b) split the data into two sets of 256 data points, overlapping by 50 data points, then used the FFT, and (c) use the DFT with all 462 data points. Spectral

estimates for (a) and (c) were averaged over 4 consecutive spectral bands. Spectral estimates for (b) were each averaged over 2 consecutive spectral bands and the results from the two data sets averaged (effectively the same as averaging over 4 consecutive spectral bands).

The results are presented in Figure 6–16 and Figure 6–17. While some differences among the three options can be observed, the results are essentially equivalent. The only difference occurs using the 2 x 256 FFT data sets, which loses information at the low end of the spectrum, as expected. (Recall from Section 3.7, if we're interested in the low frequency end of the spectrum then we need to sample over a long period of time. 512 data points is twice as long as 256 data points).

For this particular example there is no meaningful difference when we padded the data by adding 5% of random numbers (or zeroes) to each end of the data set and used a single 512 FFT, whether we used 2 x 256 FFTs with an overlap of 50 records, or whether we used the DFT and all 462 data points.

Returning to the question we asked.

(a) Both rotary spectra are relatively flat, indicating that there are no processes in the data set, other than those that have already been removed during the pre-processing. Other rotary spectra are relatively uniform across the whole frequency domain.
(b) Whether the data are padded, the DFT is used or two smaller FFTs are used, the results are basically the same. Noting however, that the smaller FFT will limit the extent of low frequency resolution in the spectra.

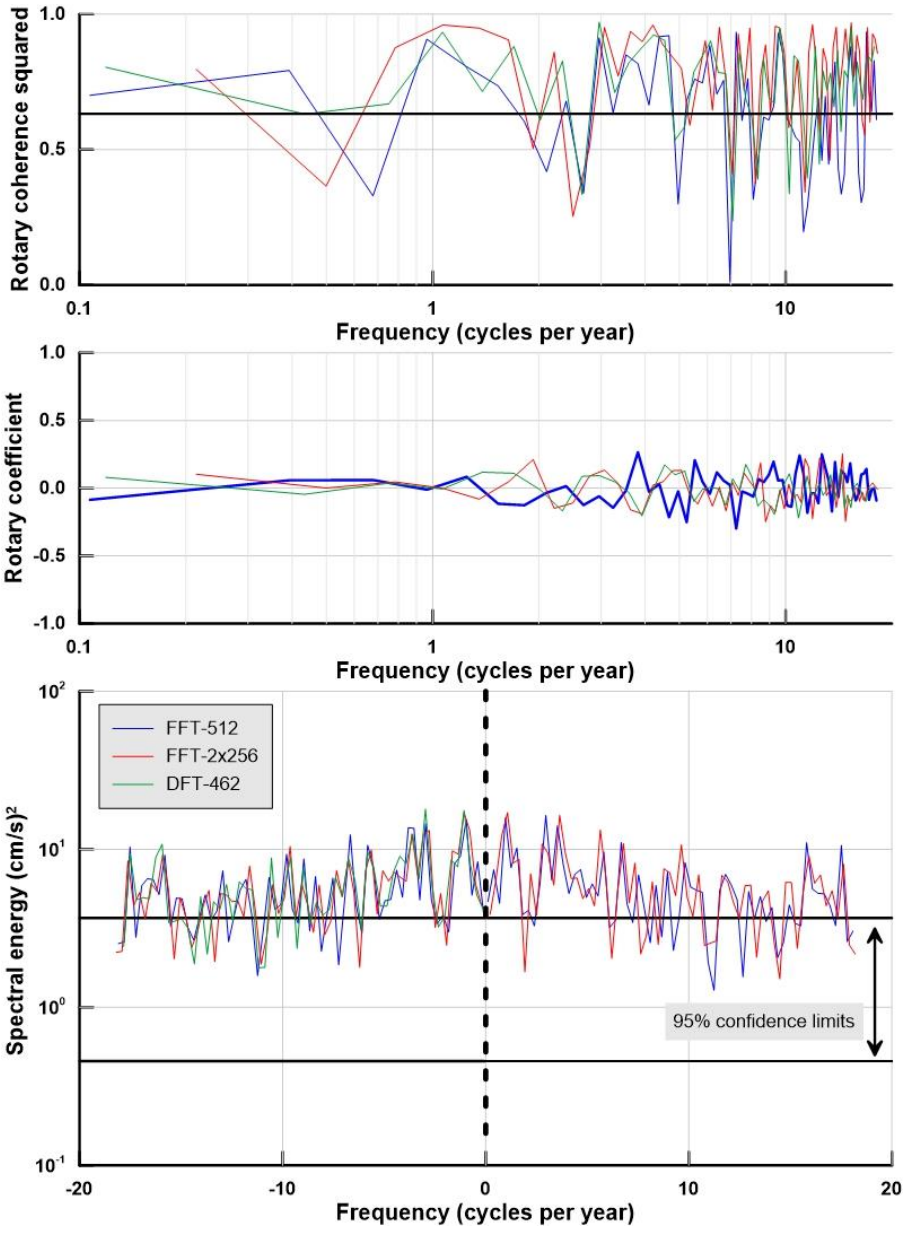


Figure 6–16. Periodogram of the clockwise and anti-clockwise rotary components, the rotary coefficient and the rotary coherence squared using the three options.

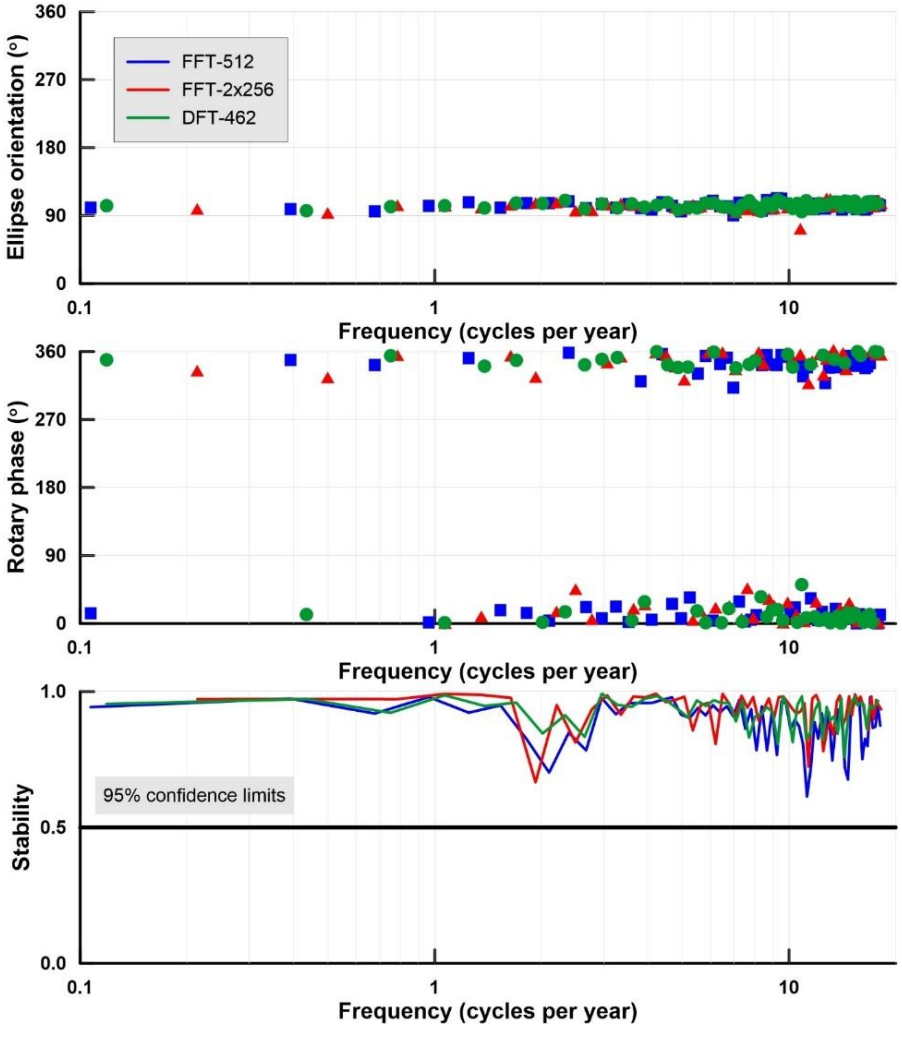


Figure 6–17. Stability, rotary phase and ellipse orientation spectra of the rotary coefficients for each of the three options.

6.3. Examples of the rotary spectrum – a pair of vectors

6.3.1. Example 6-4: A pair of vectors – hourly current data

This example examines the rotary spectra using a pair of currents, separated vertically by 2 m. As the separation of the two current meters is very small, we expect there to be very high coherence between the two data sets.

Questions: How similar are the rotary spectral products generated from two current meters separated vertically by 2 m?

Data used:

- Currents in component form (east-west and north-south).
- The two sets of current data were separated vertically by 2 m.
- The length of the data was 4,096 records, at a sampling interval of 1 hour. Hence the total length of the data was about 6 months.
- The Nyquist period is 2 hours (Nyquist frequency is ½ cycles per hour).
- We have about 170 days so the longest period features we can realistically interpret is about 20 days.

Pre-processing:

- The original data were sampled at 5 minute intervals.
- A 7-point Gaussian filter was used to smooth the data, which were then subsampled to hourly intervals.
- The time series plots of the data are shown in Figure 6–18 in stick vector form. The length of each stick is proportional to the current speed, while the orientation of the sticks indicates the direction of the currents (using the oceanographic convention).
- Visually, there is little difference between the two plots. This is expected as the two data sets are separated, vertically by only 2 m.
- The dominant currents are rotated slightly east of north, although there is considerable variation throughout the data record.

- The linear trend was removed from each of the four time series components.

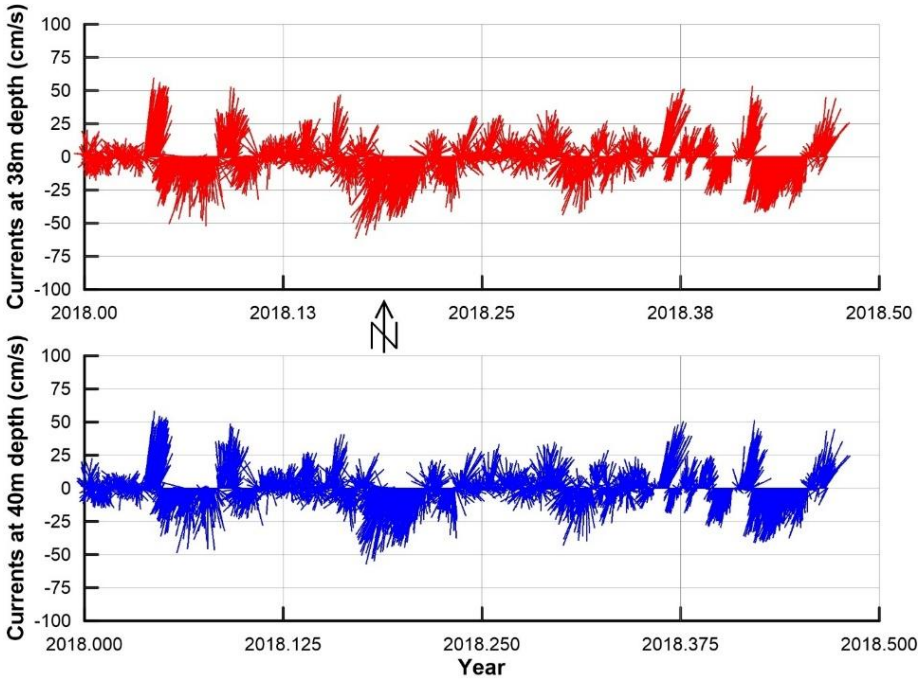


Figure 6–18. Time series of the currents in stick vector form.

Testing:

- Stationarity: Results from the Dickey-Fuller test give Tau calculated values of -21.58 and -9.18 (for the east-west current anomalies of the first data set) and -21.15 and -8.91 (for the north-south current anomalies of the second data set); the Tau critical value is - 2.86. As the calculated Tau is less than Tau critical for both data sets, we can reject the null hypothesis that the data sets are not stationary. The similarity between the tau values for the two east-west components and again for the two north-south components is expected as the data are only separated by 2 m.

- To undertake the tests for stationarity, the data were split into 16 groups each comprising 256 data points. Tests for normality (chi-squared), homogeneity of variance (Bartlett) and equality of means (ANOVA) all failed. The hypothesis that the means and variances are from the same populations is rejected. We note these results and proceeded with the analysis.

Smoothing of the periodogram:

- Four consecutive spectral estimates were averaged over the entire spectrum.
- As undertaken previously, the number of spectral estimates that are averaged may differ – usually with a larger number of spectral estimates being averaged at higher frequencies and a smaller number of spectral estimates averaged at low frequencies.
- Again, spectral averaging enables confidence limits to be generated.
- The 95% confidence limits are presented.

Description: The inner rotary spectral energy (Figure 6–19), the inner rotary phase (Figure 6–20) and the and inner rotary coherence squared (Figure 6–21) are presented. The outer rotary spectral energy (Figure 6–22), the outer rotary phase (Figure 6–23) and the outer rotary coherence squared (Figure 6–24) are all presented. Each set comprises four spectra ('11', '22', '12' and '21', where '1' represents vector number '1' and '2' represents vector number '2'), although not all spectra are unique.

In keeping with many presentations of rotary spectra, these plots use 'negative' frequencies for the clockwise rotating component and 'positive' frequencies for the anti-clockwise rotating component.

Inner rotary energy spectra. 'Inner 11' (Figure 6–19a), 'Inner 22' (Figure 6–19b) and 'Inner 12' (Figure 6–19c) spectra are unique. However, spectra 'Inner 21' (Figure 6–19d) is identical to 'Inner 12' (Figure 6–19c) and only one of these cross rotary spectra need be used.

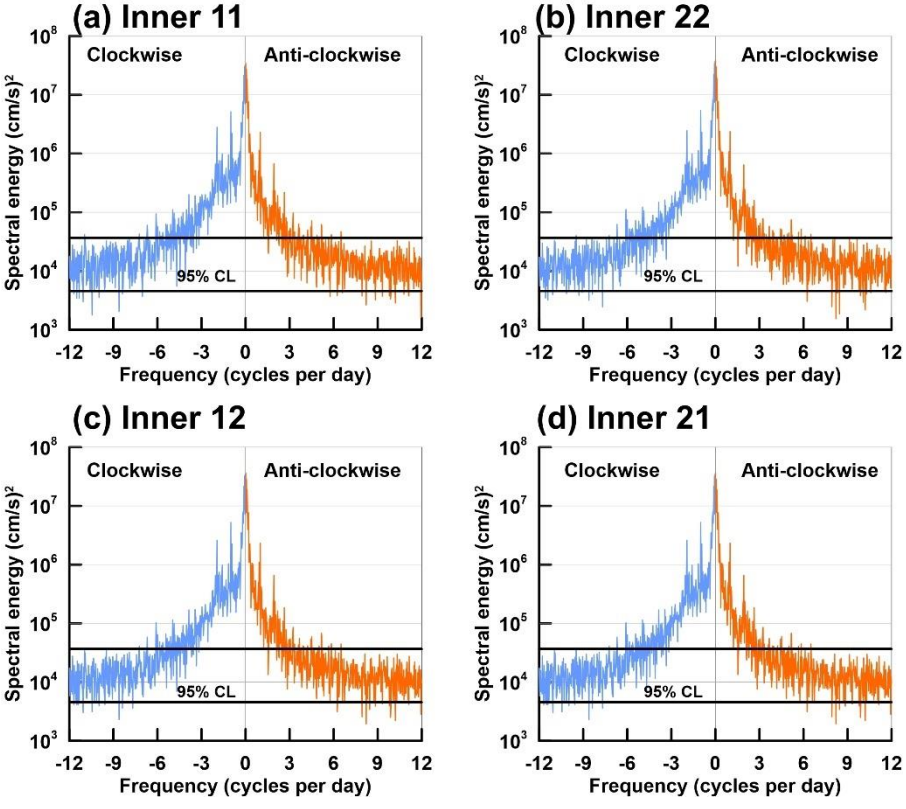


Figure 6–19. Periodogram of the inner rotary spectra (two auto spectra, top plots, and two cross spectra, bottom plots). The horizontal black lines represent the upper and lower 95% confidence limits (CL).

Inner rotary phase spectra. The two inner auto rotary phase spectra ('Inner 11', Figure 6–20a and 'Inner 22', Figure 6–20b) are both zero across all frequencies. This is understandable as each of these spectra are effectively comparing the phase of the vector components with itself. 'Inner 12' (Figure 6–20c), the inner rotary cross phase spectrum is unique. The inner rotary cross phase spectrum 'Inner 21' (Figure 6–20d) is a reflection (about a phase of zero) of 'Inner 12' (Figure 6–20c). Hence, two of the inner rotary phase spectra produce trivial results and one phase spectra is a reflection of the other. Therefore, only one of these inner rotary phase spectra need be used (generally, this is the inner rotary cross phase 'Inner 12').

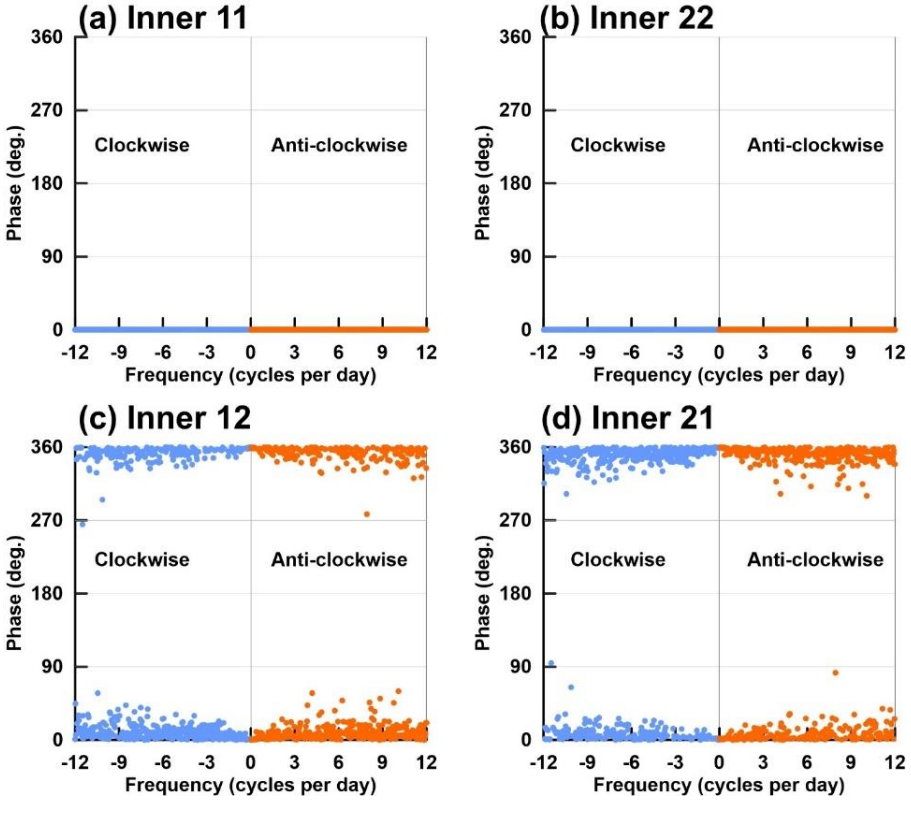


Figure 6–20. Inner rotary phase spectra.

Inner rotary coherence squared spectra. The two inner auto rotary coherence squared spectra ('Inner 11', Figure 6–21a and 'Inner 22', Figure 6–21b) are both unity across all frequencies. Again, this is understandable as each of these spectra is effectively comparing the vector with itself. The inner rotary cross coherence squared spectrum ('Inner 12', Figure 6–21c) is unique. The spectrum 'Inner 21', (Figure 6–21d) is identical to 'Inner 12' (Figure 6–21c). Hence, two of the inner rotary coherence squared spectra produce trivial results and one is identical to the other. Therefore, only one of these inner cross rotary coherence squared spectra need be used (generally, the inner rotary cross coherence squared spectra, 'Inner 12', is used).

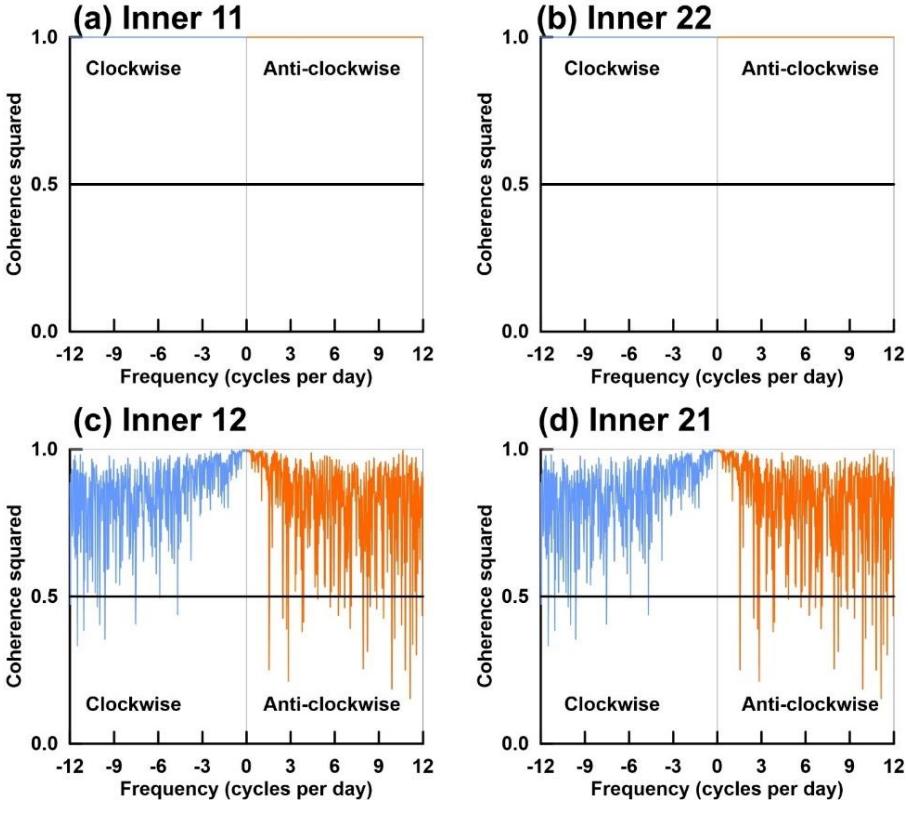


Figure 6–21. Inner rotary coherence squared spectra. The 95% confidence limit is indicated by the horizontal black lines.

Results for the inner rotary components indicate the following.

- Of the 12 inner rotary spectra that can be produced, only 5 are non-trivial or unique. They are: the two inner rotary auto energy spectra, the inner rotary cross energy spectrum, the inner rotary cross phase spectrum and the inner rotary cross coherence squared spectrum.
- The inner rotary energy spectra (both auto rotary and the cross rotary spectra) all show similar features (Figure 6–19). There is high energy at low frequency and spectral peaks at the diurnal and semi-diurnal periods. Several smaller peaks are also apparent at higher frequencies, probably a result of energy leakage into the harmonics.
- The clockwise and anti-clockwise components of the spectra show similar features, although the clockwise rotating component at the diurnal and semi-diurnal periods appear slightly larger than those for the anti-clockwise rotating components. This is, perhaps, unusual in this region.
- The inner rotary phase (Figure 6–20) is close to zero across most frequencies (both clockwise and anti-clockwise) although there is considerable variability in phase, particularly as the frequency increases.
- The inner rotary coherence squared (Figure 6–21) exceeds the 95% confidence level across most frequencies (both clockwise and anti-clockwise).

The high inner rotary coherence squared and inner rotary phase close to zero are expected. Recalling that the inner rotary components refer to functions rotating in the same direction (i.e. either both anti-clockwise rotation or both clockwise rotation). The two current meters are separated only by 2m vertically, so it is expected that vectors rotating in similar directions should be similar.

Outer rotary energy spectra. The clockwise and anti-clockwise rotating components of the 'Outer 11' (Figure 6–22a) and 'Outer 22' (Figure 6–22b) auto spectra are reflected about a frequency of 'zero'. The clockwise rotating cross spectrum of 'Outer 12' (Figure 6–22c) is a reflection of the anti-clockwise rotating cross spectrum 'Outer 21' (Figure 6–22d) and vice versa. Therefore, only the positive frequencies of the rotary auto spectra together with one of the cross rotary spectral components are required to uniquely specify the outer rotary energy spectra.

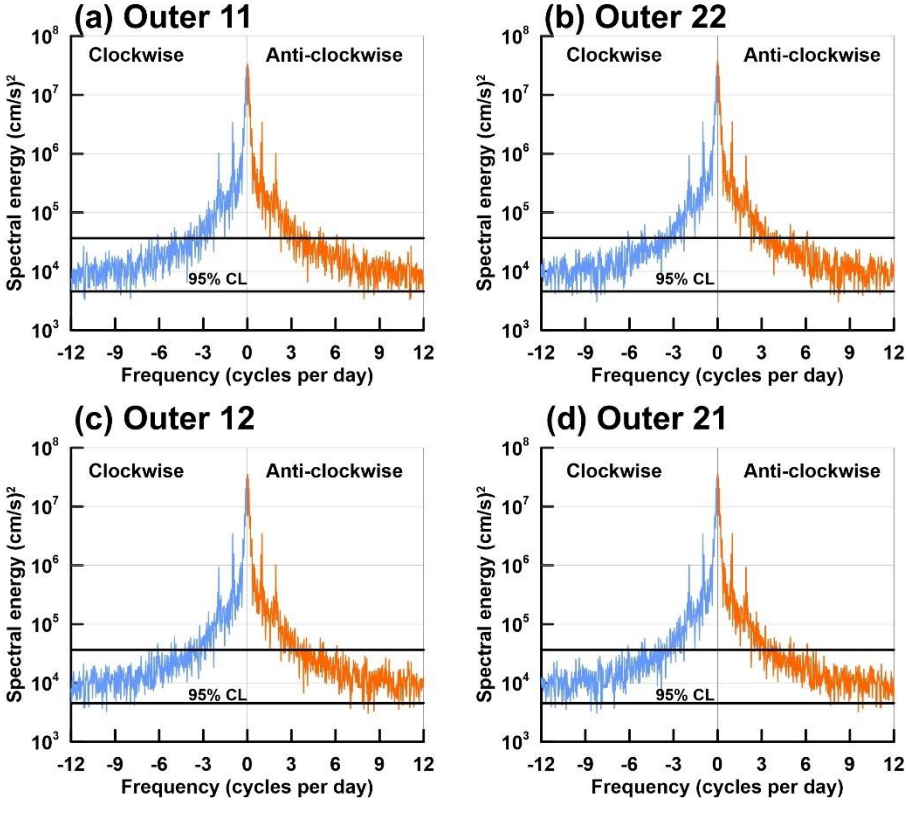


Figure 6–22. Periodogram of the outer rotary spectra (two auto spectra and two cross spectra). The horizontal black lines are the upper and lower 95% confidence limits (CL).

Outer rotary phase spectra. The outer rotary phase spectra have a similar form as the energy spectra. That is: the clockwise and anti-clockwise rotating components of the 'Outer 11' (Figure 6–23a) and 'Outer 22' (Figure 6–23b) rotary auto phase spectra are reflected about a frequency of 'zero'. The clockwise component of the rotating cross spectrum of 'Outer 12' (Figure 6–23c) is a reflection (about a frequency of 'zero') of the anti-clockwise component of the rotating cross spectrum 'Outer 21' (Figure 6–23d) and vice versa. Therefore, only the positive frequencies of the rotary phase spectra together with one of the cross rotary phase spectral components are required to uniquely specify the outer rotary phase spectra.

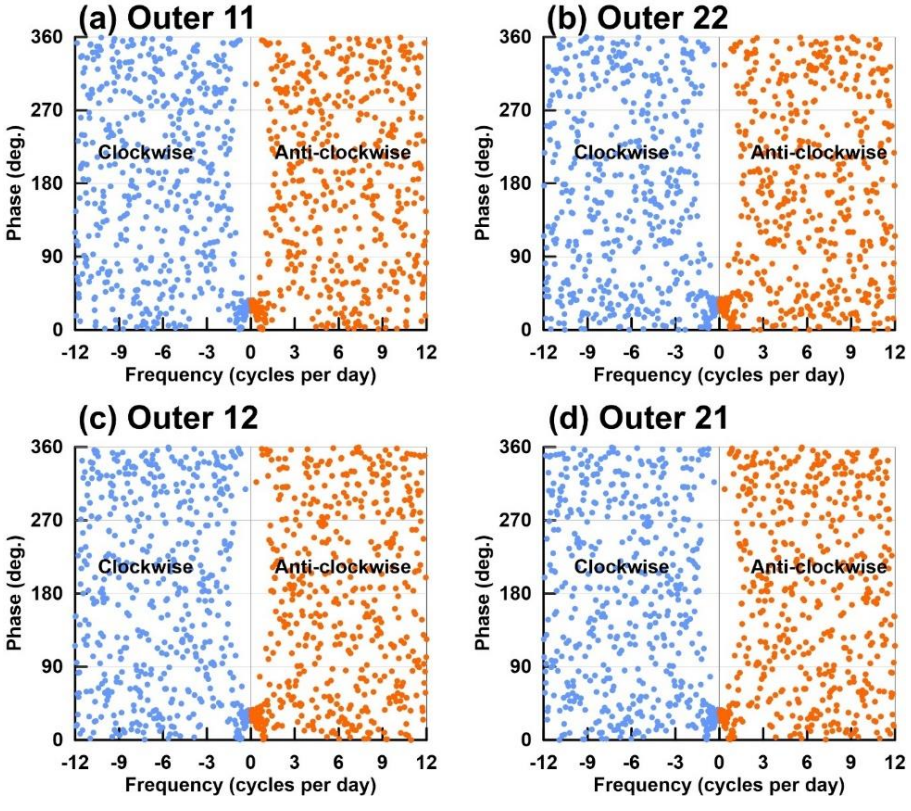


Figure 6–23. Outer rotary phase spectra.

Outer rotary coherence squared spectra. Again, the outer rotary coherence squared spectra have a similar form to that of the energy spectra and the phase spectra. The clockwise and anti-clockwise rotating components of the 'Outer 11' (Figure 6–24a) and 'Outer 22' (Figure 6–24b) rotary auto coherence squared spectra are reflected about a frequency of 'zero'. The clockwise rotating cross spectrum of 'Outer 12' (Figure 6–24c) is a reflection about a frequency of 'zero' of the anti-clockwise rotating cross spectrum 'Outer 21' (Figure 6–24d) and vice versa. Therefore, only the positive frequencies of the rotary coherence squared spectra together with one of the rotary cross coherence squared spectral components are required to uniquely specify the outer rotary coherence squared spectra.

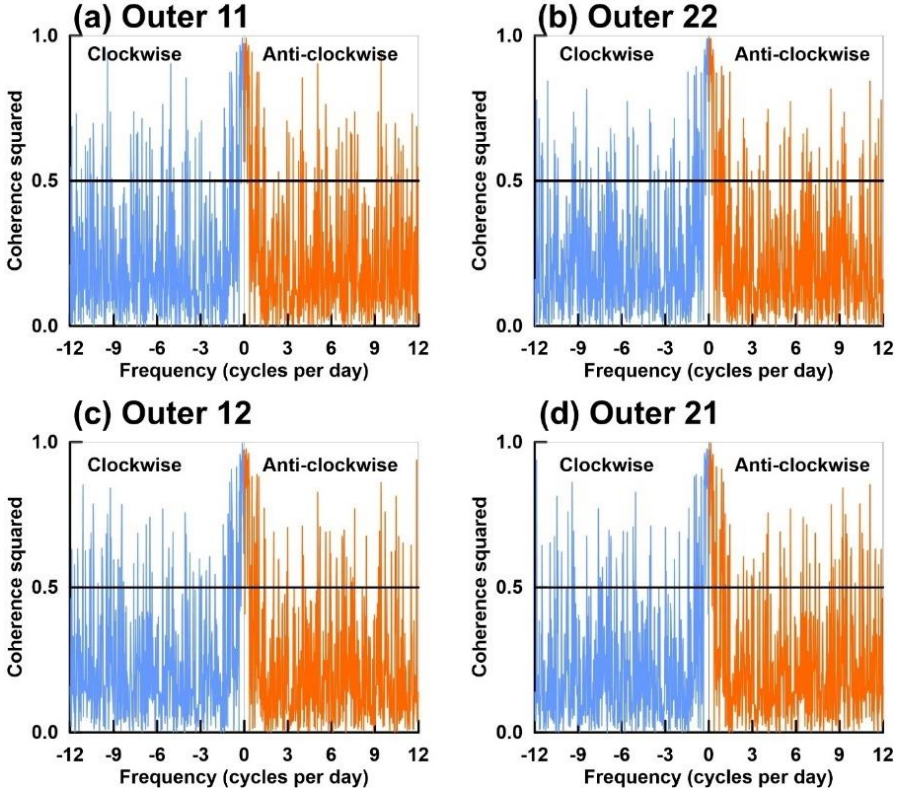


Figure 6–24. Outer rotary coherence squared spectra. The horizontal black lines represent the 95% confidence limits.

Results for the outer rotary components indicate the following.

- For the outer energy auto spectra (as well as the outer auto phase spectra and the outer auto coherence squared spectra) only positive frequencies are required. The values associated with negative frequencies are simply a reflection (about a frequency of 'zero') of those associated with positive frequencies. Only one of the outer cross spectra are required for the energy, phase and coherence squared rotary cross spectra.
- The outer rotary energy spectra (both auto rotary and the cross rotary spectra) all show similar features (Figure 6–22). There is high energy at low frequency and spectral peaks at the diurnal and semi-diurnal periods.
- The outer rotary phase (Figure 6–23) appears similar to the phase obtained from a random process across all frequencies (both clockwise and anti-clockwise). There is no distinct pattern in these spectra.
- The outer rotary coherence squared (Figure 6–24) exceed the 95% confidence level on occasions, but not as dominant as observed in the inner rotary coherence squared plots.

The generally low outer rotary coherence squared and the random like appearance of the outer rotary phase are both expected. There is no reason to expect a close association between the anti-clockwise rotating component of one vector with the clockwise rotating component of another vector.

Returning to the question we originally asked:

(a) As expected (because of the close proximity of the two current meters), the rotary spectra all indicate that the currents at these two levels in the water column are almost identical. The energy spectra are almost identical, the rotary phase is close to zero and the rotary coherence generally lies above the 95% confidence limits.

6.3.2. Example 6-5: A pair of vectors – 10-day currents and winds

Rotary spectra are used to examine currents at a depth of about 10m and winds from a shore-based station some 10 km distant. Both data sets are filtered by removing the tides and inertial periods and subsampled such that delta-t is 10 days.

For the spectral analysis, the current speeds were converted to m/s (from cm/s) so that both the winds and currents used the same units. While this won't substantially change the results, interpretation is made easier if all variables have the same units. Differences will appear in the magnitudes of the energy spectra, but not in the phase or coherence squared spectra.

Questions: What are the spectral relationships between the winds and the near surface currents?

Data used:

- Currents and winds in component form (east-west and north-south). The length of the data was 512 records, taken ten days apart. Hence the total length of the data was about 14 years.
- The Nyquist period is 10 days (Nyquist frequency is 1 cycle every 20 days).
- The data set is about 14 years long so the longest period features we can realistically interpret is about 2 years. This restricts any answer to our question to the frequency band bounded by about 0.05 and 0.5 cycles per year.

Pre-processing:

- The original data were sampled at 5 minute intervals.
- A 7-point Gaussian filter was used to smooth the data, which were then subsampled to hourly intervals.
- The hourly data were then subject to the 120i913 filter (the filter length is 240 hours) described in Thompson (1984) to remove tidal and inertial features.
- The linear trend was removed from each of the four time series.

- The final data set had a time step of 10 days.
- Wind directions were converted to the oceanographic direction convention (i.e. wind direction is now defined as the direction towards which the wind is blowing).
- The time series plots of the data are shown in Figure 6–25 in stick vector form. The length of each stick is proportional to the current speed, while the orientation of the sticks indicates the direction towards which the winds / currents are moving.
- The dominant currents are rotated slightly east of north, although there is considerable variation throughout the data record.

Testing:

- Stationarity: Results from the Dickey-Fuller test give Tau calculated values of -17.78 and -23.19 (for the east-west and north-south components of the wind anomalies) and -20.01 and -19.89 (for the east-west and north-south components of the current anomalies); the Tau critical value is - 2.86. As the calculated Tau is less than Tau critical for both data sets, we can reject the null hypothesis that the data sets are not stationary.
- All data anomalies (i.e. with the means removed) were normally distributed. Data were split into 16 groups each comprising 32 data points. Bartlett's test for homogeneity of variance indicated that the data sets comprised samples from the same population at the 5% level of significance. Similarly, ANOVA tests indicated no significant difference in the means of each data set at the 5% level of significance i.e. the different data sets are likely to have come from the same population. Therefore, the data are likely to be stationary and we proceed with the analysis.

Smoothing of the periodogram:

- Both periodograms were averaged using 4 spectral estimates in each band over the entire spectrum.
- Spectral averaging enabled confidence limits to be generated.
- The 95% confidence limits are presented.

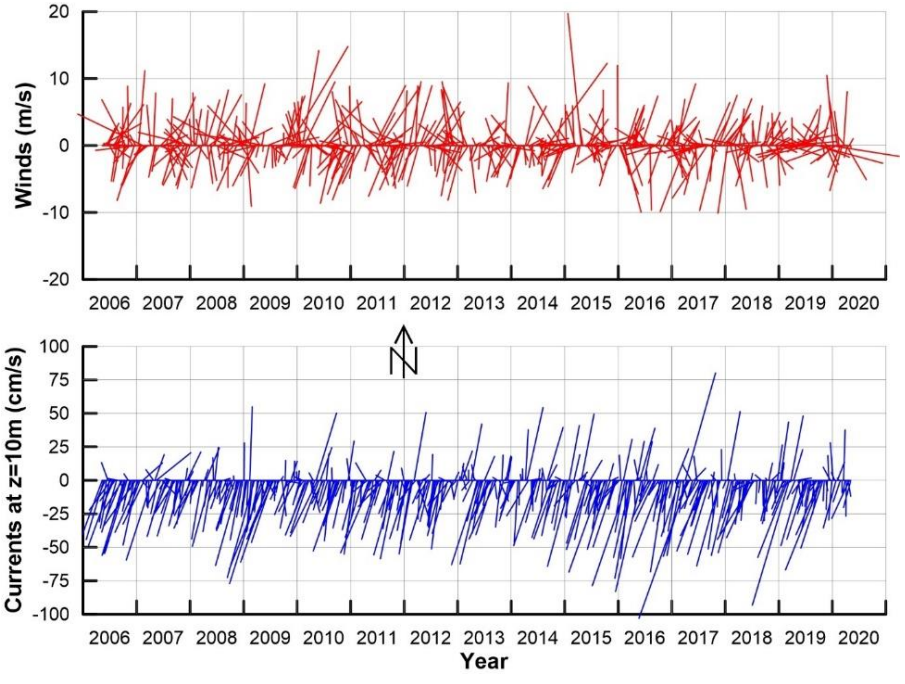


Figure 6–25. Time series of the winds (top) and currents (bottom) in stick vector form.

Description: As with the previous example, four inner and four outer energy spectra, four inner and four outer phase spectra, and four inner and four outer coherence squared spectra are produced. Again, not all of these spectra are unique. All are presented below, although the trivial and non-unique spectra are usually omitted when presenting the results.

The inner rotary spectral energy (Figure 6–26), the inner rotary phase (Figure 6–27) and the and inner rotary coherence squared (Figure 6–28) are presented. The outer rotary spectral energy (Figure 6–29), the outer rotary phase (Figure 6–30) and the outer rotary coherence squared (Figure 6–31) are all presented. Each set comprises four spectra ('11', '22', '12' and '21', where '1' represents vector number '1' and '2' represents vector number '2'), although not all spectra are unique.

Inner rotary energy spectra. 'Inner 11' (Figure 6–26a), 'Inner 22' (Figure 6–26b) and 'Inner 12' (Figure 6–26c) spectra are unique. However, the cross spectra 'Inner 21' (Figure 6–19d) is identical to 'Inner 12' (Figure 6–26c) and only one of these cross rotary spectra need be used.

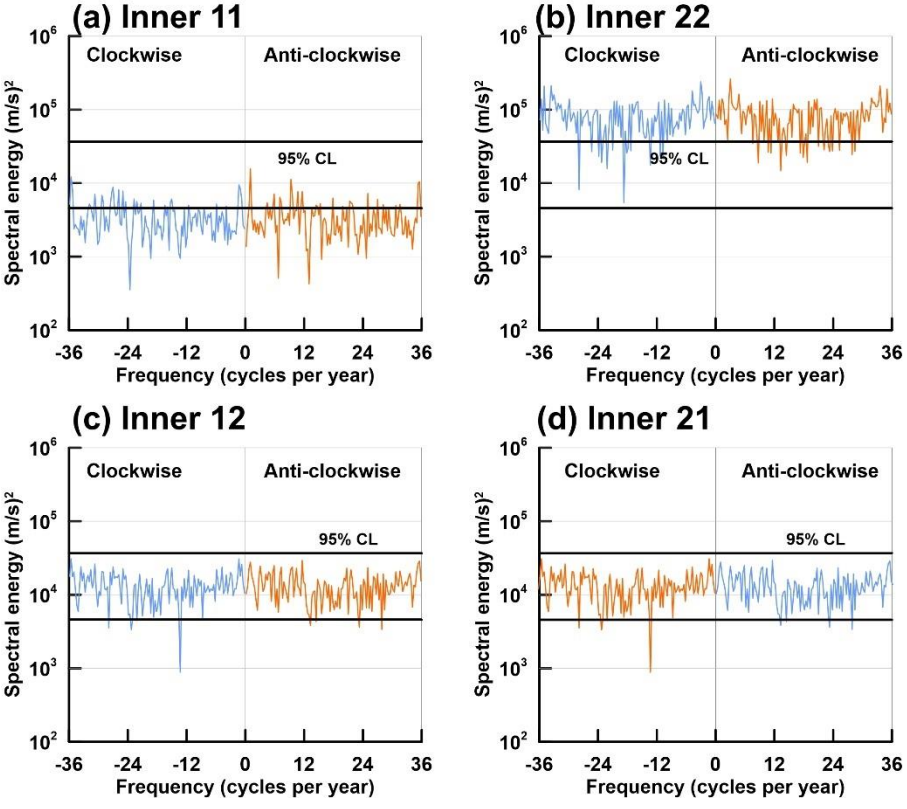


Figure 6–26. Periodogram of the inner rotary spectra (two auto spectra and two cross spectra). The horizontal black lines are the upper and lower 95% confidence limits (CL).

Inner rotary phase spectra. The two inner auto rotary phase spectra ('Inner 11', Figure 6–27a and 'Inner 22', Figure 6–27b) are both zero across all frequencies. This is understandable as each of these spectra are effectively comparing the phase of the vector components with itself. 'Inner 12' (Figure 6–27c), the inner rotary cross phase spectrum is unique. The inner rotary cross phase spectra 'Inner 21' (Figure 6–27d) is a reflection (about a phase of zero) of 'Inner 12' (Figure 6–27c). Hence, two of the inner rotary phase spectra produce trivial results and one inner rotary cross phase spectrum is a reflection of the other. Therefore, only one of these inner rotary phase spectra need be used (generally, this is the inner rotary cross phase 'Inner 12').

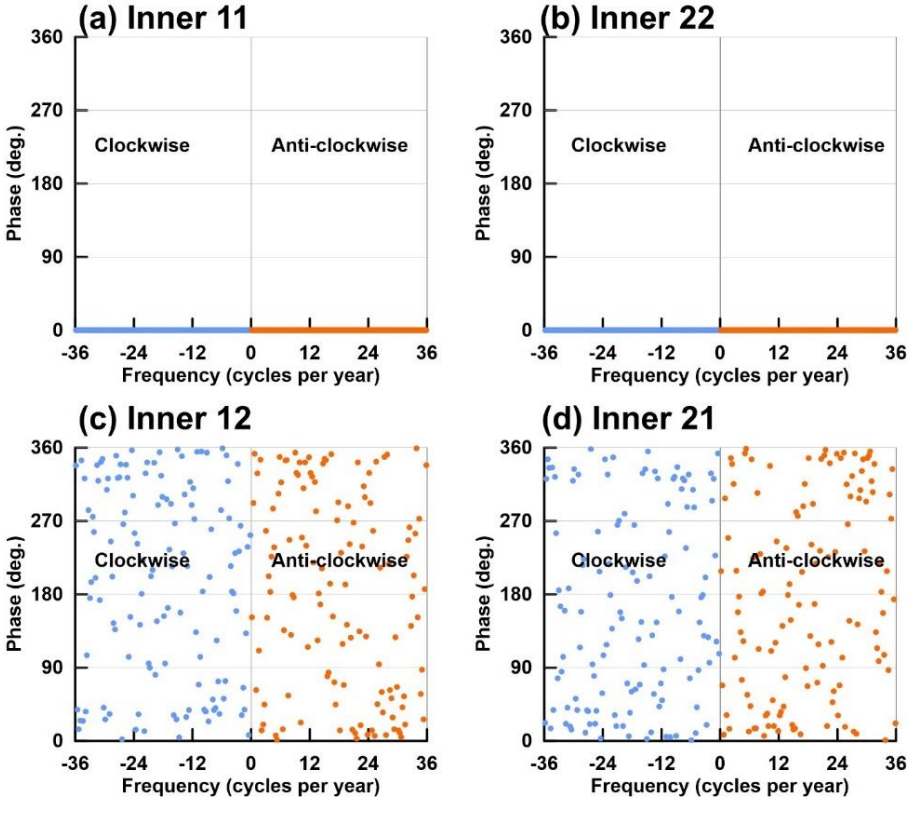


Figure 6–27. Inner rotary phase spectra.

Inner rotary coherence squared spectra. The two inner auto rotary coherence squared spectra ('Inner 11', Figure 6–28a and 'Inner 22', Figure 6–28b) are both unity across all frequencies. Again, this is understandable as each of these spectra is effectively comparing the vector with itself. The inner rotary cross coherence squared spectrum ('Inner 12', Figure 6–28c) is unique. The spectrum 'Inner 21' , (Figure 6–28d) is identical to 'Inner 12' (Figure 6–28c). Hence, two of the inner rotary coherence squared spectra produce trivial results and one is identical to the other. Therefore, only one of these inner cross rotary coherence squared spectra need be used (generally, the inner rotary cross coherence squared spectra, 'Inner 12', is used).

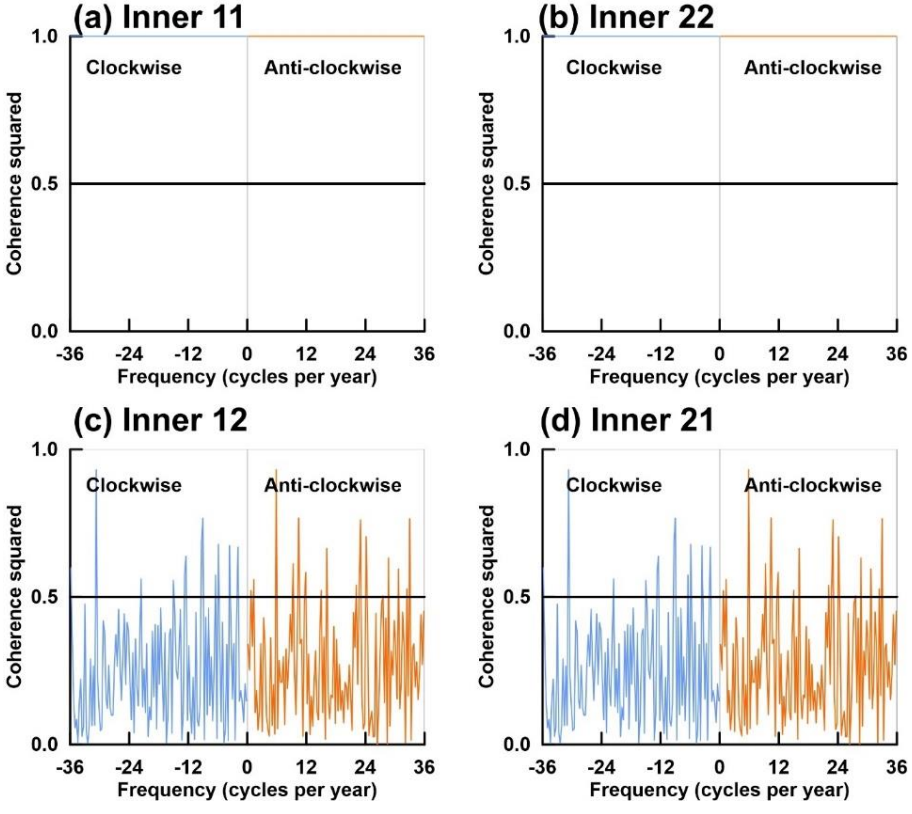


Figure 6–28. Inner rotary coherence squared spectra. The horizontal black lines represent the 95% confidence limits.

Results for the inner rotary components indicate the following.

- Of the 12 inner rotary spectra that can be produced, only 5 are non-trivial or unique. They are: the two inner rotary auto energy spectra, the inner rotary cross energy spectrum, the inner rotary cross phase spectrum and the inner rotary cross coherence squared spectrum.
- The inner rotary energy spectra (both auto rotary and the cross rotary spectra) are all relatively flat and featureless (Figure 6–26), although the energy does vary by about two orders of magnitude. A similar pattern was observed for a purely random process. This observed flat spectra suggests that the main features in the data set have been removed by the filtering process.
- Similarly, the inner rotary phase (Figure 6–27) indicates a random structure. There does not appear to be a clear pattern in the phase spectra.
- The inner rotary coherence squared (Figure 6–28) is generally less than the exceeds the 95% confidence level across most frequencies (both clockwise and anti-clockwise).

These results suggest that the filtering process undertaken as part of the data pre-processing, has removed all of the features in the data sets. Recall, that the trends have been removed and the filter removed features with tidal and inertial periods.

Outer rotary energy spectra. The clockwise and anti-clockwise rotating components of the 'Outer 11' (Figure 6–29a) and 'Outer 22' (Figure 6–29b) auto spectra are reflected about a frequency of 'zero'. The clockwise rotating cross spectrum of 'Outer 12' (Figure 6–29c) is a reflection of the anti-clockwise rotating cross spectrum 'Outer 21' (Figure 6–29d) and vice versa. Therefore, only the positive frequencies of the two rotary auto spectra together with one of the cross rotary spectral components are required to uniquely specify the outer rotary energy spectra.

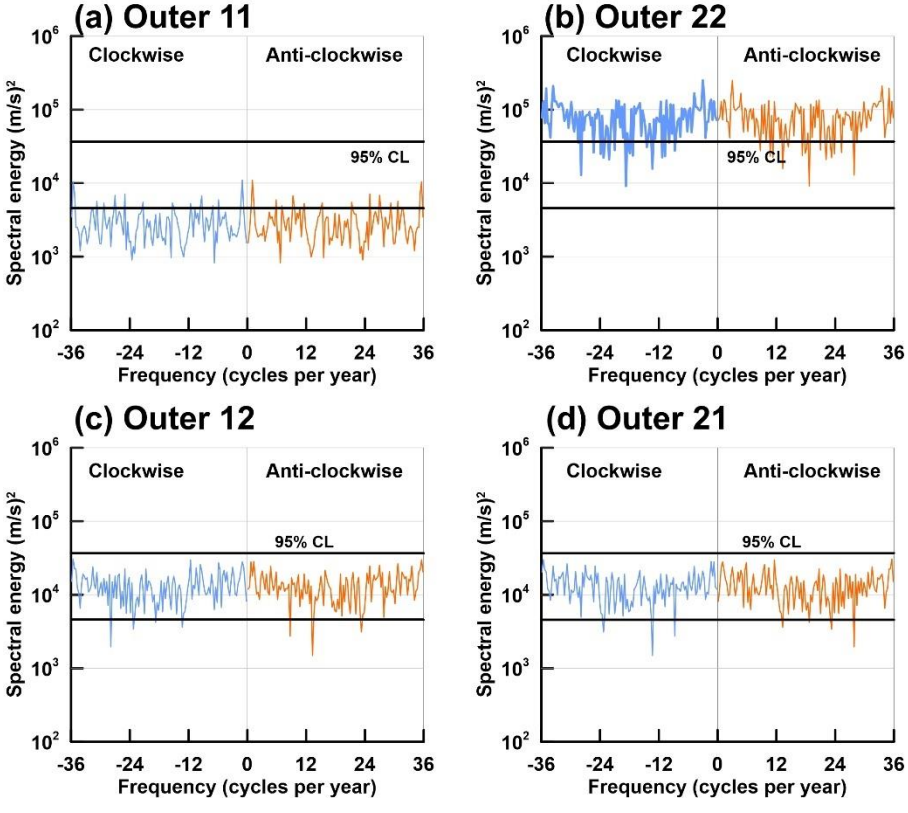


Figure 6–29. Periodogram of the outer rotary spectra (two auto spectra and two cross spectra). The horizontal black lines are the upper and lower 95% confidence limits (CL).

Outer rotary phase spectra. The outer rotary phase spectra have a similar form as the energy spectra. That is: the clockwise and anti-clockwise rotating components of the 'Outer 11' (Figure 6–30a) and 'Outer 22' (Figure 6–30b) rotary auto phase spectra are reflected about a frequency of 'zero'. The clockwise component of the rotating cross spectrum of 'Outer 12' (Figure 6–30c) is a reflection (about a frequency of 'zero') of the anti-clockwise component of the rotating cross spectrum 'Outer 21' (Figure 6–30d) and vice versa. Therefore, only the positive frequencies of the auto rotary phase spectra together with one of the cross rotary phase spectral components are required to uniquely specify the outer rotary phase spectra.

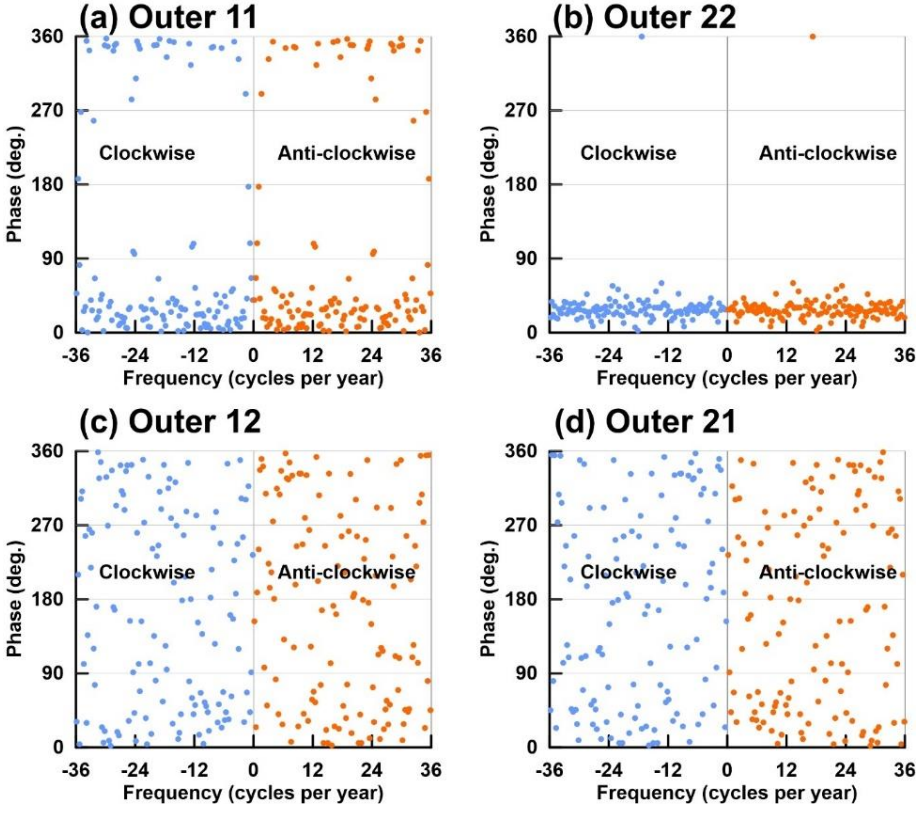


Figure 6–30. Outer rotary phase spectra.

Outer rotary coherence squared spectra. Again, the outer rotary coherence squared spectra have a similar form to that of the energy

spectra and the phase spectra. The clockwise and anti-clockwise rotating components of the 'Outer 11' (Figure 6–31a) and 'Outer 22' (Figure 6–31b) rotary auto coherence squared spectra are reflected about a frequency of 'zero'. The clockwise rotating cross spectrum of 'Outer 12' (Figure 6–31c) is a reflection (about a frequency of 'zero') of the anti-clockwise rotating cross spectrum 'Outer 21' (Figure 6–31d) and vice versa. Therefore, only the positive frequencies of the auto rotary coherence squared spectra together with one of the rotary cross coherence squared spectral components are required to uniquely specify the outer rotary coherence squared spectra.

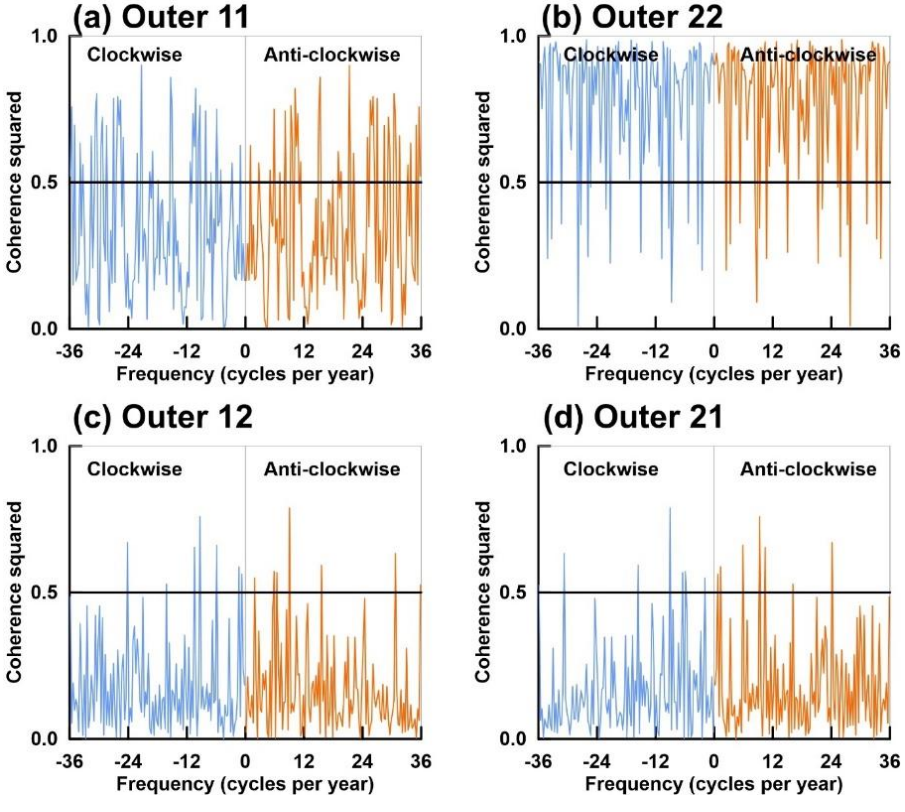


Figure 6–31. Outer rotary coherence squared spectra. The horizontal black lines represent the 95% confidence limits.

Results for the outer rotary components indicate the following.

- The outer rotary energy spectra show similar results to the inner rotary energy spectra. All four outer rotary energy spectra and relatively flat and featureless. Again, reminiscent of a random energy spectrum (Figure 6–29).
- The two outer cross rotary phase spectra (Figure 6–30) also appear to be result of random processes. However, the two outer auto rotary phase spectra indicate phases that have relatively small variations across the entire frequency band. Note that the outer auto rotary phase compares the clockwise and anti-clockwise rotating components of the winds (Outer 11) or the currents (Outer 22).
- The two outer cross rotary coherence squared spectra (Figure 6–31) rarely exceed the 95% confidence level. In contrast, the two outer auto rotary coherence spectra often exceed the 95% confidence level.

Returning to the question we originally asked:

(a) Some of the inner and outer auto rotary spectra are significant. However, only at selected frequencies do the (inner or outer) cross rotary spectra exceed the 95% confidence limits. Therefore, we conclude that there is some significant correlation between the local winds and the near surface currents in the 20-500 day period in the observed data set. We'd need to further investigate the data sets to determine whether these frequencies have any physical meaning.

6.4. Summary of rotary spectra

For a single vector, and at each frequency, p, we get the following rotary spectra:

$$Anti\ Clockwise = \frac{1}{8}[A^2 + B^2 + C^2 + D^2 + 2(AD - BC)]$$

$$Clockwise = \frac{1}{8}[A^2 + B^2 + C^2 + D^2 - 2(AD - BC)]$$

Where A and B are the real and imaginary elements of one component of the vector and C and D are the real and imaginary elements of the other component of the vector.

The orientation of the semi-major axis of the ellipse is given by:

$$\theta = \frac{1}{2}\left[tan^{-1}\left(\frac{C - B}{A + D}\right) + tan^{-1}\left(\frac{C + B}{A - D}\right)\right] + k\pi$$

The phase (or 'starting point' of the vector on the ellipse) is given by:

$$\phi = \frac{1}{2}\left[tan^{-1}\left(\frac{C - B}{A + D}\right) - tan^{-1}\left(\frac{C + B}{A - D}\right)\right]$$

The 'rotary stability' (γ_R) is given by:

$$\gamma_R = \frac{[A^2 + B^2 + C^2 + D^2]^2 - 4[(A^2 + B^2)(C^2 + D^2) - (AC + BD)^2]}{[A^2 + B^2 + C^2 + D^2]^2 - 4[(AD - CB)^2]}$$

The 'rotary coefficient' (C_R) is given by:

$$C_R = \frac{-2\,(AD - BC)}{[A^2 + B^2 + C^2 + D^2]^2}$$

The two counter rotating circles form an ellipse with the major and minor semi-axes given by (respectively):

$$R^+ = \frac{1}{2}(A + D) + \frac{i}{2}(C - B) \quad \text{and} \quad R^- = \frac{1}{2}(A - D) + \frac{i}{2}(C + B)$$

For a pair of vectors (labelled below with the subscripts '1' and '2'), we have:

$$R_1^+ = \frac{1}{2}(A_1 + D_1) + \frac{i}{2}(C_1 - B_1) \text{ and } R_1^- = \frac{1}{2}(A_1 - D_1) + \frac{i}{2}(C_1 + B_1)$$

$$R_2^+ = \frac{1}{2}(A_2 + D_2) + \frac{i}{2}(C_2 - B_2) \text{ and } R_2^- = \frac{1}{2}(A_2 - D_2) + \frac{i}{2}(C_2 + B_2)$$

There is a level of redundancy in these results – some components are trivial, while others are a reflection about the phase or the frequency. Only the unique results are given below.

The unique results are: in the inner rotary energy (two auto and one cross), the cross phase (the auto phase is zero) and cross coherence squared (the auto coherence squared is unity) spectra:

$$Inner_{11} = \begin{cases} R_1^+ R_1^+ = (R_1^+)^2 & for \quad p > 0 \\ R_1^- R_1^- = (R_1^-)^2 & for \quad p < 0 \end{cases}$$

$$Inner_{22} = \begin{cases} R_2^+ R_2^+ = (R_2^+)^2 & for \quad p > 0 \\ R_2^- R_2^- = (R_2^-)^2 & for \quad p < 0 \end{cases}$$

$$Inner_{12} = \begin{cases} R_1^+ R_2^+ \, exp[i(\theta_1^+ - \theta_2^+)] & for \quad p > 0 \\ R_1^- R_2^- \, exp[-i(\theta_1^- - \theta_2^-)] & for \quad p < 0 \end{cases}$$

$$tan(\phi_{12}) = \begin{cases} \dfrac{\langle R_1^+ R_2^+ \, sin(\theta_1^+ - \theta_2^+) \rangle}{\langle R_1^+ R_2^+ \, cos(\theta_1^+ - \theta_2^+) \rangle} & for \quad p > 0 \\ \dfrac{\langle -R_1^- R_2^- \, sin(\theta_1^- - \theta_2^-) \rangle}{\langle R_1^- R_2^- \, cos(\theta_1^- - \theta_2^-) \rangle} & for \quad p < 0 \end{cases}$$

$$\gamma_{12}^2 = \begin{cases} \dfrac{\langle R_1^+ R_2^+ \, cos(\theta_1^+ - \theta_2^+) \rangle^2 + \langle R_1^+ R_2^+ \, sin(\theta_1^+ - \theta_2^+) \rangle^2}{\langle (R_1^+)^2 \rangle \, \langle (R_2^+)^2 \rangle} & for \;\; p > 0 \\ \dfrac{\langle R_1^- R_2^- \, cos(\theta_1^- - \theta_2^-) \rangle^2 + \langle R_1^- R_2^- \, sin(\theta_1^- - \theta_2^-) \rangle^2}{\langle (R_1^-)^2 \rangle \, \langle (R_2^-)^2 \rangle} & for \;\; p < 0 \end{cases}$$

We also obtain the outer rotary energy (two auto and one cross), the outer phase (two auto for positive frequencies only and one cross) and outer coherence squared (two auto for positive frequencies only and one cross) spectra:

$$Outer_{11} = R_1^- R_1^+ \, exp[i(\theta_1^+ - \theta_1^-)] \qquad for \quad p > 0$$

$$Outer_{22} = R_2^- R_2^+ \, exp[i(\theta_2^+ - \theta_2^-)] \qquad for \quad p > 0$$

$$Outer_{12} = \begin{cases} R_1^- R_2^+ \, exp[i(\theta_2^+ - \theta_1^-)] & for \quad p > 0 \\ R_1^+ R_2^- \, exp[i(\theta_1^+ - \theta_2^-)] & for \quad p < 0 \end{cases}$$

$$tan(\psi_{11}) = \frac{\langle R_1^- R_1^+ \, sin(\theta_1^- - \theta_1^+)\rangle}{\langle R_1^- R_1^+ \, cos(\theta_1^- - \theta_1^+)\rangle} \qquad for \qquad p > 0$$

$$tan(\psi_{22}) = \frac{\langle R_2^- R_2^+ \, sin(\theta_2^- - \theta_2^+)\rangle}{\langle R_2^- R_2^+ \, cos(\theta_2^- - \theta_2^+)\rangle} \qquad for \qquad p > 0$$

$$tan(\psi_{12}) = \begin{cases} \dfrac{\langle R_1^- R_2^+ \, sin(\theta_1^- - \theta_2^+)\rangle}{\langle R_1^- R_2^+ \, cos(\theta_1^- - \theta_2^+)\rangle} & for \quad p > 0 \\ \dfrac{\langle R_1^+ R_2^- \, sin(\theta_2^- - \theta_1^+)\rangle}{\langle R_1^+ R_2^- \, cos(\theta_2^- - \theta_1^+)\rangle} & for \quad p < 0 \end{cases}$$

$$\lambda_{11}^2 = \frac{\langle R_1^- R_1^+ \rangle^2 [\langle cos(\theta_1^+ - \theta_1^-)\rangle^2 + \langle sin(\theta_1^+ - \theta_1^-)\rangle^2]}{\langle (R_1^+)^2\rangle \, \langle (R_1^-)^2\rangle} \qquad for \; p > 0$$

$$\lambda_{22}^2 = \frac{\langle R_2^- R_2^+ \rangle^2 [\langle cos(\theta_2^+ - \theta_2^-)\rangle^2 + \langle sin(\theta_2^+ - \theta_2^-)\rangle^2]}{\langle (R_2^+)^2\rangle \, \langle (R_2^-)^2\rangle} \qquad for \; p > 0$$

$$\lambda_{12}^2 = \begin{cases} \dfrac{\langle R_1^- R_2^+ \rangle^2 [\langle cos(\theta_2^+ - \theta_1^-)\rangle^2 + \langle sin(\theta_2^+ - \theta_1^-)\rangle^2]}{\langle (R_2^+)^2\rangle \, \langle (R_1^-)^2\rangle} & for \quad p > 0 \\ \dfrac{\langle R_1^+ R_2^- \rangle^2 [\langle cos(\theta_1^+ - \theta_2^-)\rangle^2 + \langle sin(\theta_1^+ - \theta_2^-)\rangle^2]}{\langle (R_1^+)^2\rangle \, \langle (R_2^-)^2\rangle} & for \quad p < 0 \end{cases}$$

6.5. Overview of spectral analysis

One final example examines all aspects of spectral analysis outlined in the previous three chapters. It involves an assessment of the structure of high frequency internal waves in the coastal ocean.

Internal waves

All water waves propagate along a density interface. The most commonly observed surface waves exist because of the density difference between the water and the air. Similarly, if there is a density difference between two different types of water (characterised by their density differences), then there is the potential for waves to propagate along the density interface formed by these two water masses. Such waves are known as internal waves. They propagate inside (i.e. internally within) the waters.

One way in which this density difference may be generated is by the repeated (solar) heating and (wind) mixing of the surface layers of the water. This creates warmer (lower density) water that sits above cooler (higher density) water. Internal waves may then propagate along the density interface that is formed by these waters of different density. Often, the water temperature is used as a surrogate for ocean density. In our region of interest, water temperature in summer is a good approximation for density (Tate, et al. 2019) as is used in this example.

Questions: Can the spectral signals in the water temperature and currents be used to characterise internal waves?

Data used:

- The data used were currents in component form (east-west and north-south) and water temperature throughout the water column. The total water depth was about 65m.
- Sampling was undertaken every 5 minutes. Hence, the Nyquist period is 10 minutes (Nyquist frequency is 1/10 cycles per minute or 6 cycles per hour).
- The length of the data was 32 records, about 2.7 hours.
- High frequency internal waves from this region generally occur in packets a few hours in duration. Each packet comprises 4-6 individual waves.

Pre-processing:

- Principal component analysis was used to determine the along shore and across shelf components of the currents.
- Linear interpolation was used to convert the temperature data from temperature at a fixed depth (as measured by the thermistors) to the depth of each isotherm.
- Linear trends were removed from all data resulting in a set of data anomalies.
- Analyses presented below were based on the depth of the 20°C isotherm (which was approximately at mid-water depth) and the

currents 15 m and 49 m above the sea floor (i.e. from depths above and below the depth of the 20°C isotherm).

Data are plotted in Figure 6–32. Several points can be noted from this figure.

- About 6 (or 7) internal waves can be identified in this packet. The displacement of the 20°C isotherm indicates a trough-to-crest internal waves height of almost 15 m (about one quarter of the total water depth).
- The along-shore component of the currents at both depths (15 m and 49 m above the sea floor) is substantially less than the corresponding across shore currents. This is unusual – the along shore currents are usually substantially greater than the across shore currents.
- Across shore components of the currents 15 m above the sea floor move in the opposite direction to the across shore components of the currents 49 m above the sea floor. The magnitude of the across shore currents at both depths is approximately the same.

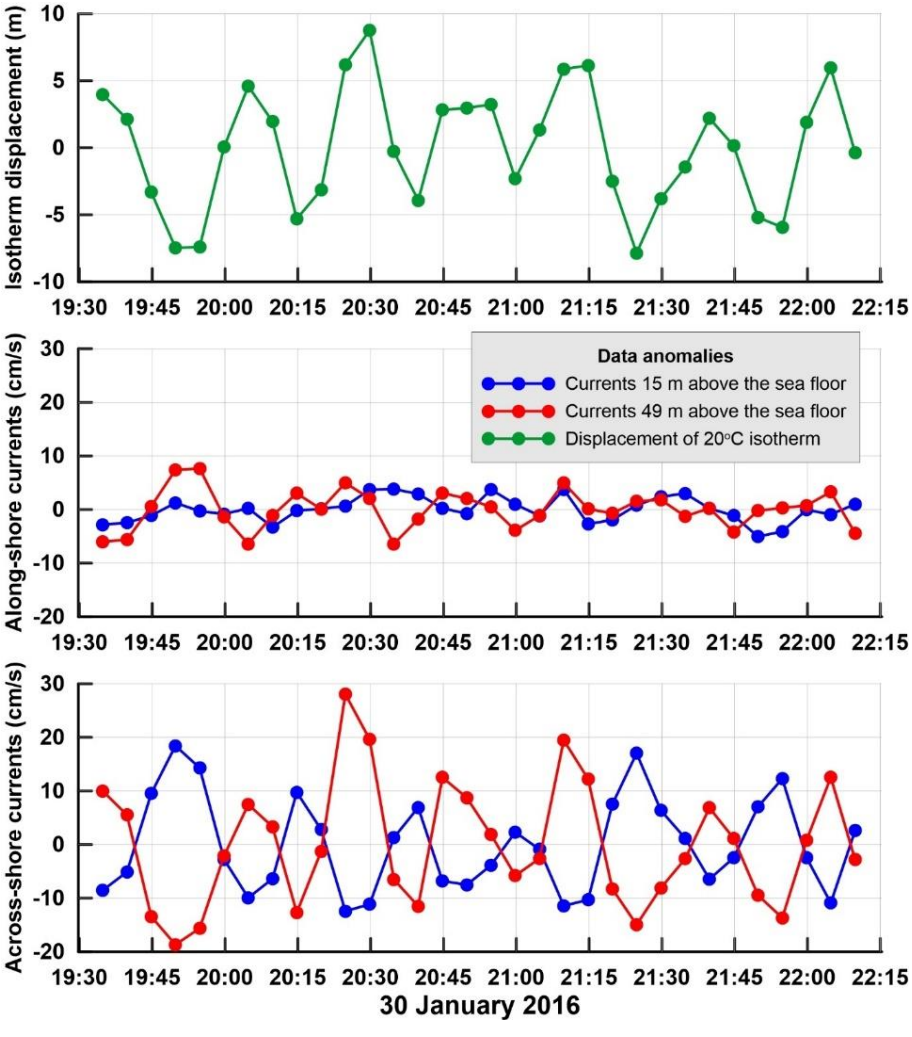


Figure 6–32. Time series plots of the data anomalies. The closed circles represent the actual values (the lines joining them are shown for convenience only).

For the auto and cross spectral analyses, focus will be placed on the across shore current components and the displacement of the 20°C isotherm. In this region, the direction of propagation of the internal waves is towards the (approximate) west, perpendicular to the bathymetry. Therefore, most

of the variations in currents due to the internal waves should be most noticeable in the across shore component of the currents.

The amplitude and phase spectra for the across shore components of the currents 15 m and 49 m above the sea floor and the displacement of the 20°C isotherm are shown in Figure 6–33. The 95% confidence limits were generated by band averaging over two adjacent frequency bands.

Two features are apparent from the spectra. Firstly, all three spectral energy plots (Figure 6–33, bottom plot) have an identical shape indicating the same processes affect all three variables in a similar manner. Spectral energy 49 m above the sea floor is slightly greater than the spectral energy 15 m above the sea floor. All three spectra show a broad peak between about 2-3 cycles per hour. It may be argued that these peaks do not exceed the 95% confidence limits. However, referring back to the original time series (Figure 6–32) it is clear that these peaks are real and dominate the data sets. The broadness of the peak suggests that the period of the waves varies slightly from wave to wave. The waves may not be stationary, which implies that spectral analysis may not be the most appropriate analysis tool. We'll examine this further Chapter 8 when we examine wavelet analysis.

Secondly, the phase lag (Figure 6–33, top plot) of the across shore current components differ by about 180° – they are out of phase. When the currents at 49 m above the sea floor flow on-shore, the currents 15 m above the sea floor flow off-shore, and vice versa. This can also be identified in the time series plots (Figure 6–32). Further, it can also be seen that the phase lag of the displacement of the 20°C isotherm is almost identical to that of the across shore component of the currents 49 m above the sea floor. This arises simply because of the direction convention used for the currents (i.e. currents moving in an offshore direction are defined as positive while currents moving in an onshore direction are defined as negative). If the opposite direction convention was adopted, then the phase lag spectrum of the displacement of the 20°C isotherm would be almost identical to that of the across shore current component 15 m above the sea floor.

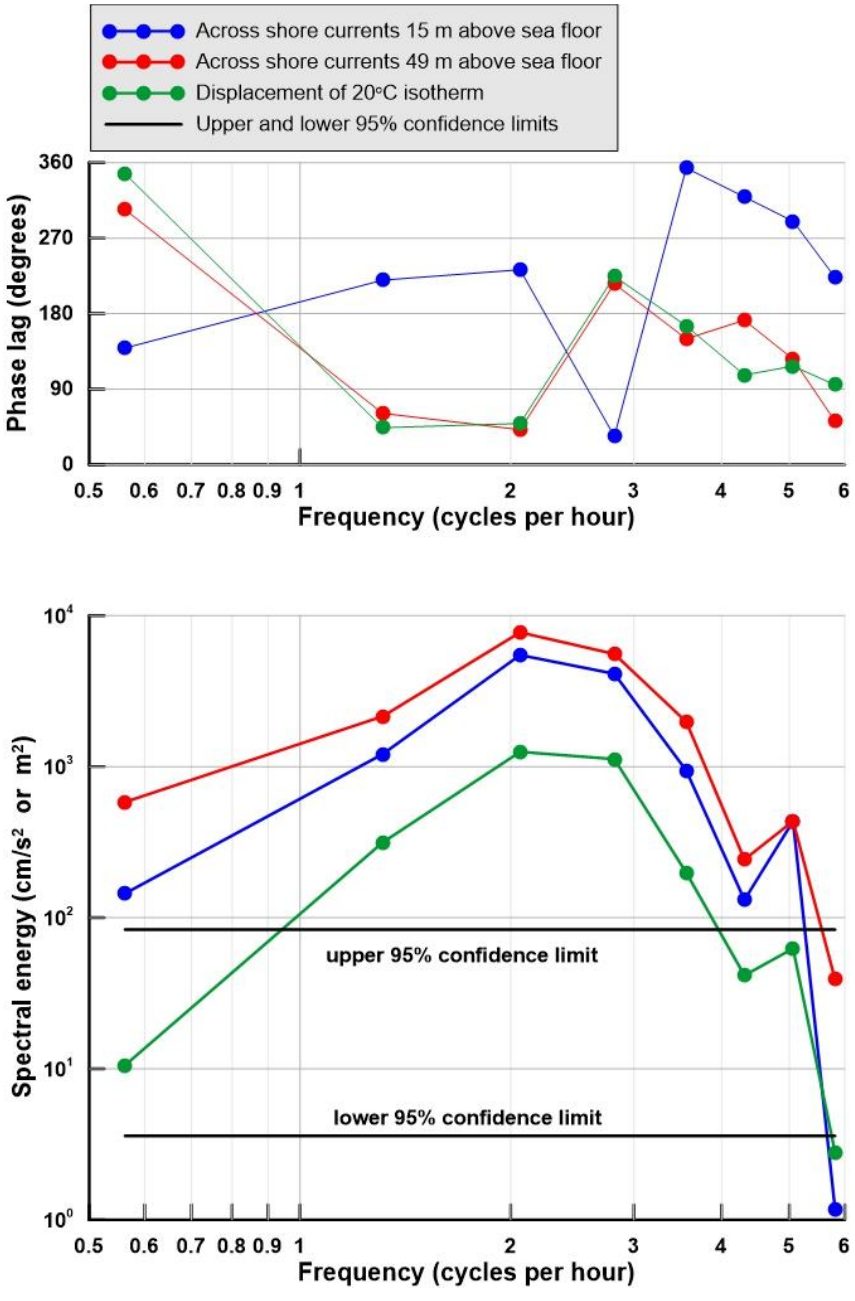


Figure 6–33. Spectral analysis of the across shore current components (15 m and 49 m above the sea floor) and the displacement of the 20°C isotherm.

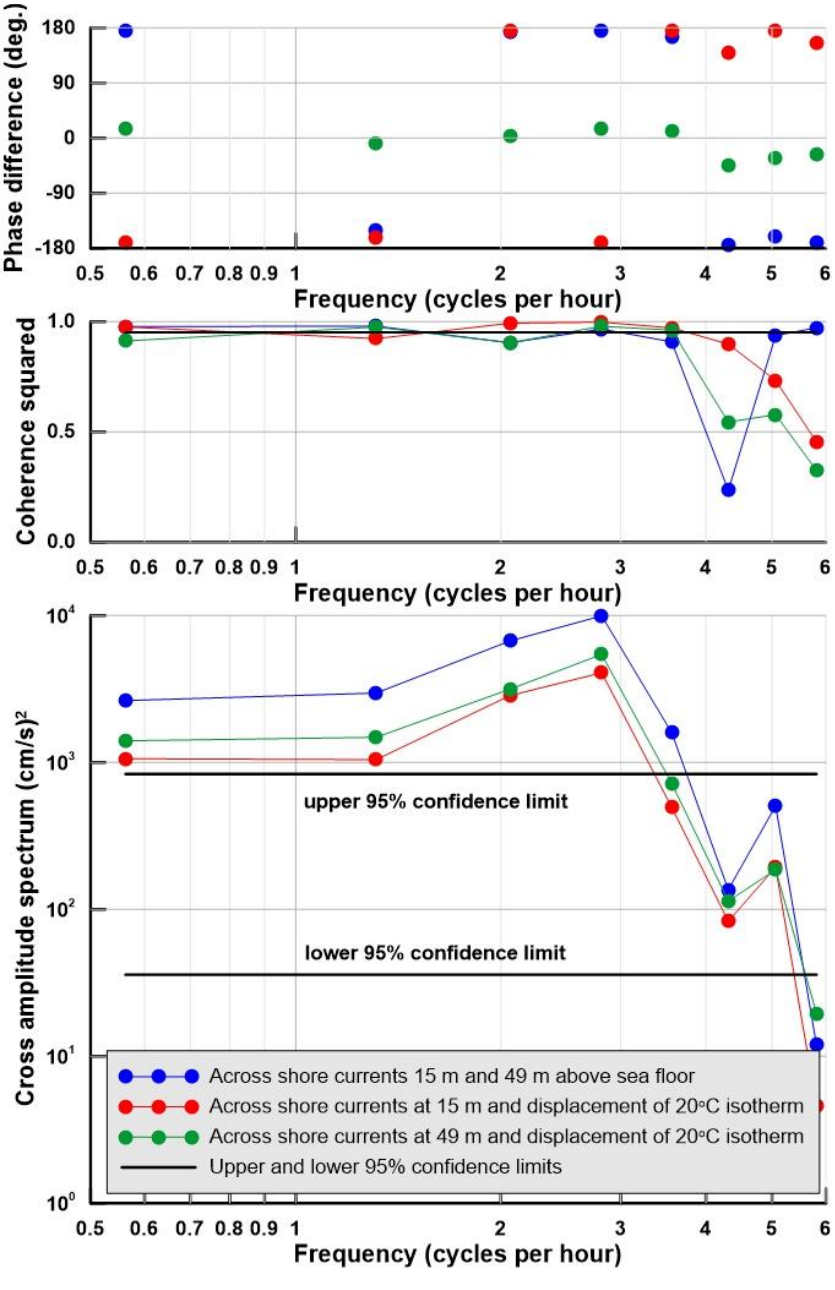


Figure 6–34. Cross amplitude (bottom), coherence squared (middle) and phase difference (top) spectra for combinations of the across shore current components, and displacement of the 20°C isotherm.

The cross spectral estimates for (a) the two across shore current components, (b) the across shore current component 15 m above the sea floor and the displacement of the 20°C isotherm and (c) the across shore current component 49 m above the sea floor and the displacement of the 20°C isotherm are all shown in Figure 6–34 (bottom plot). The cross amplitude spectra have a similar shape and structure to the spectral estimates shown in Figure 6–33 – a broad peak at about 2-3 cycles per hour. The size of this peak is less than the 95% confidence limits. The peak is not statistically significant. However, this does not necessarily mean that the peaks are unimportant. In this particular example, the peaks were probably smaller than otherwise anticipated because the periods of the individual internal waves are not the same. The waves are not stationary.

The coherence squared spectra all lie above (or very close to) the 95% confidence limits for frequencies less than about 4 cycles per hour (Figure 6–34, middle plot). At frequencies less than about 4 cycles per hour, there is a close relationship among all three pairs of data sets. This can probably be deduced from the time series plots – all three data sets showing essentially the same wave pattern.

The phase difference spectra (Figure 6–34, top plot) from the different pairs of data are either 180° out of phase (both the across shore currents at the different depths and the across shore currents 15 m above the sea floor and the displacement of the 20°C isotherm) or in phase (the across shore currents 49 m above the sea floor and the displacement of the 20°C isotherm). Again, this can be deduced by examining the phase lag spectra of the individual data sets (Figure 6–33).

Recalling, the inner rotary spectra compare vectors rotating in the same direction while the outer rotary spectra compare vectors rotating in opposite directions. The numerals '1' refer to the currented located 15 m above the sea floor, while the numerals '2' refer to the currents located 49 m above the sea floor.

The inner rotary energy spectra are plotted in Figure 6–35. All spectra show a peak (the sizes of which do not exceed the 95% confidence limits) between 2-3 cycles per hour. This peak is to be expected: from the time series plots (Figure 6–32) we note (a) that the across shore component of the currents dominate the along shore component, and (b) that there is a clear wave like structure in the across shore components of the currents at both levels in the water column. The fact that the peaks do not appear significant probably results from the individual waves having slightly different periods, which will have the effect of damping the amplitude of the dominant signal (and hence reducing the spectral energy).

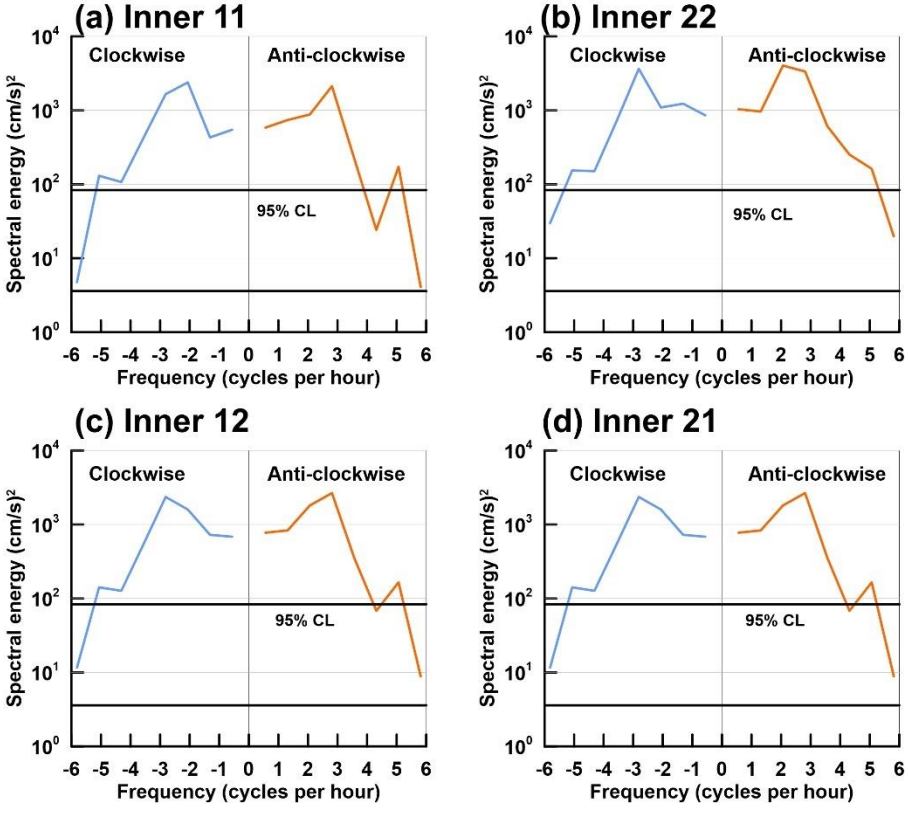


Figure 6–35. Inner rotary energy spectra of the currents 15 m and 49 m above the sea floor. The horizontal black lines represent the upper and lower 95% confidence limits (CL).

The inner rotary phase spectra (Figure 6–36, noting that only one of the four plots is non-trivial and unique) generally show the phase difference between the two current vectors to be out of phase (i.e. a phase difference of about 180°). Again, referring back to the time series (Figure 6–32), the dominant across shore components of the currents shows a crest in the current meter data 49 m above the sea floor corresponding to a trough in the current meter data 15 m above the sea floor, and vice versa. The currents in one layer move onshore, while the currents in the other layer simultaneously moves offshore. This is consistent with internal wave theory.

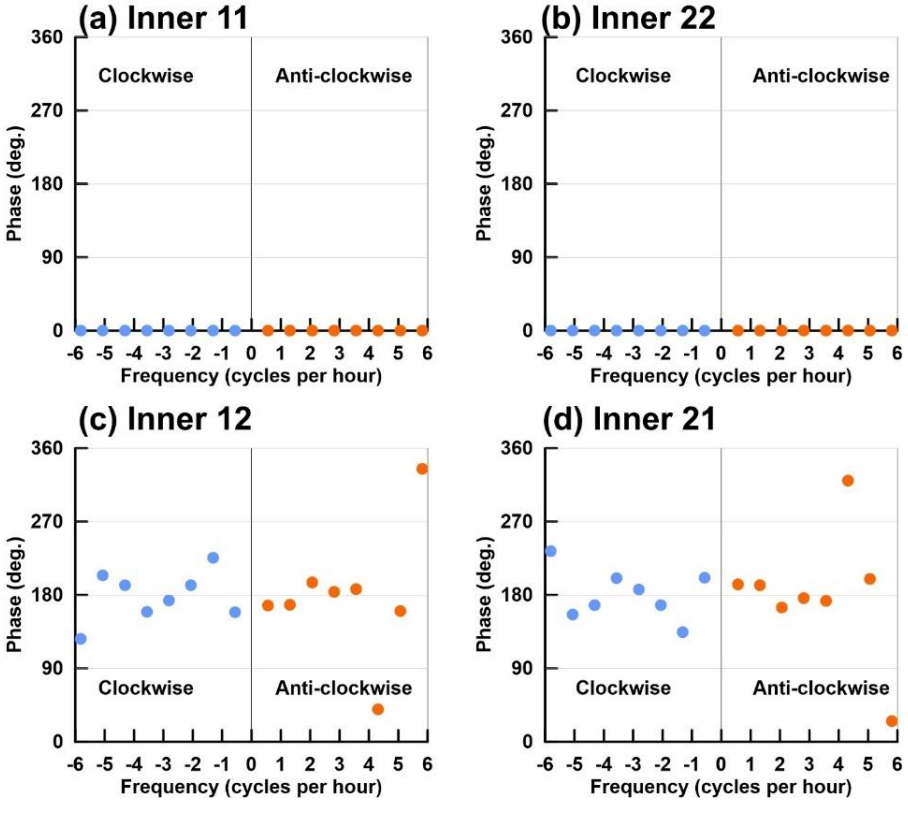


Figure 6–36. Inner rotary phase spectra of the currents 15 m and 49 m above the sea floor.

Inner rotary coherence squared (Figure 6–37) is high for frequencies less than about 4 cycles per hour. There is a high correlation between the currents from both depths in the water column whose vectors are rotating in the same direction. At high frequencies (greater than about 4 cycles per hour) this high correlation decreases – there is a de-coupling of the currents as the frequency increases.

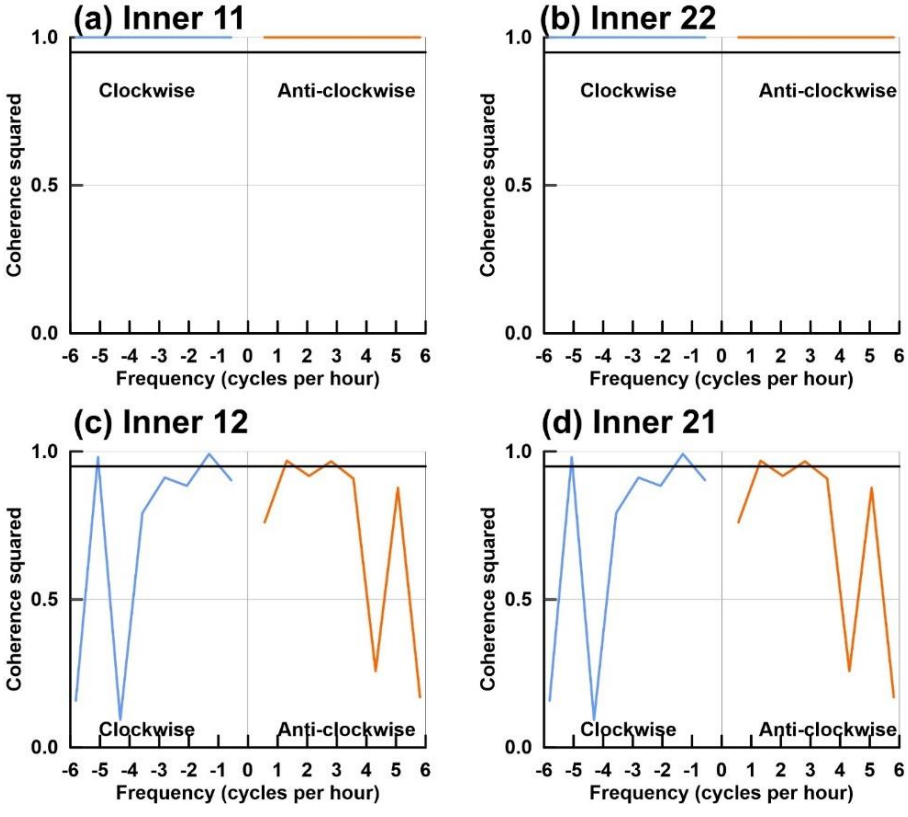


Figure 6–37. Inner rotary coherence squared spectra of the currents 15 m and 49 m above the sea floor. The horizontal black line represents the 95% confidence limit.

The structure of the outer rotary energy spectra (Figure 6–38) is basically the same as that of the inner rotary energy spectra – a spectral peak at 2-3 cycles per hour whose size is less than that of the 95% confidence limits. If the inner and outer rotary energy spectra and similar in structure and magnitude (as in this example) it suggests that the two data sets are (almost) the same, although there may be a phase difference between the two data sets.

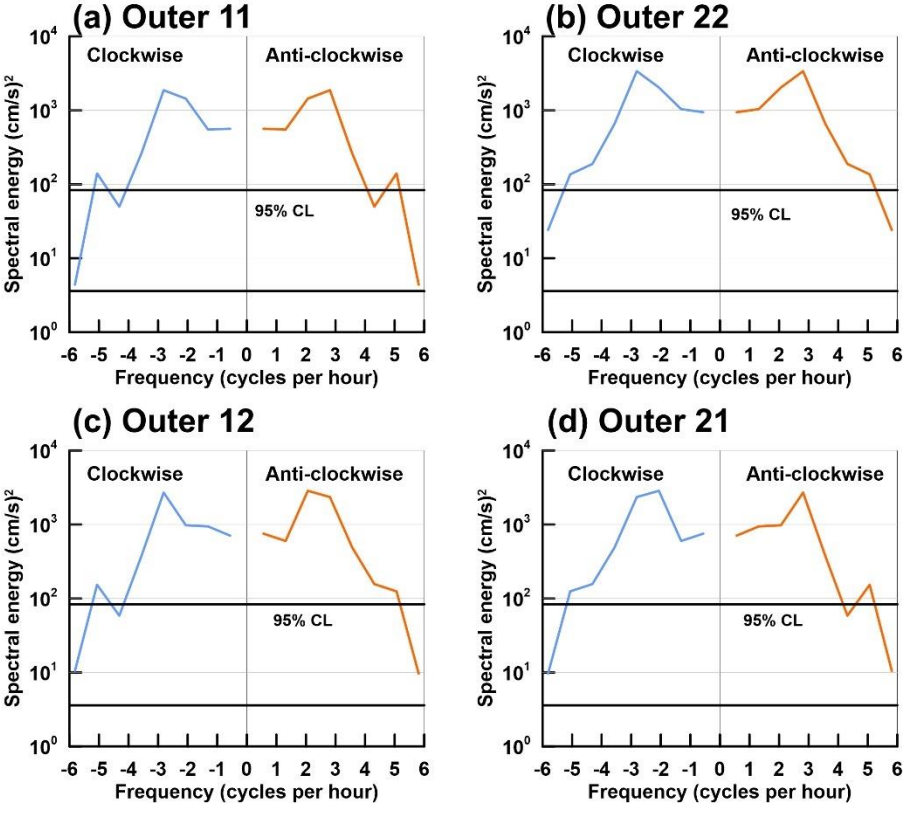


Figure 6–38. Outer rotary energy spectra of the currents 15 m and 49 m above the sea floor. The black, dotted lines represent the upper and lower 95% confidence limits (CL).

At low frequencies (less than about 4 cycles per hour) the outer rotary auto phase difference spectra (Figure 6–39 a, b) is close to 180° (i.e. out of phase). In contrast, at low frequencies, the outer rotary cross phase difference spectra (Figure 6–39 c, d) is close to being in phase.

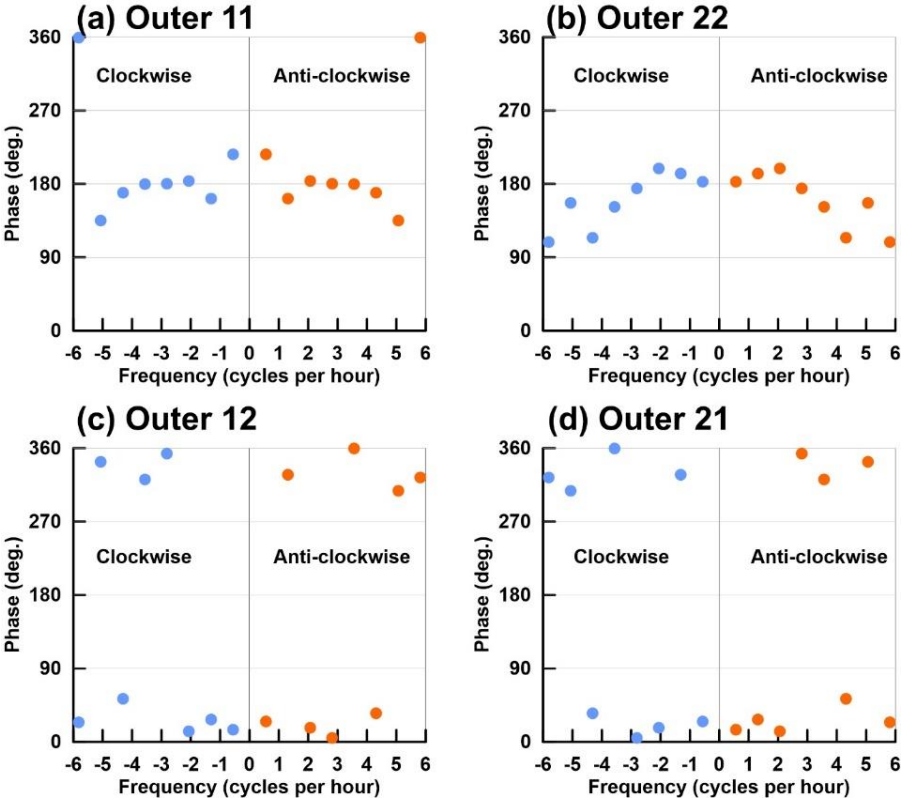


Figure 6–39. Outer rotary phase spectra of the currents 15 m and 49 m above the sea floor. Note, only one of the clockwise / anti-clockwise outer rotary phase spectra need be plotted.

The outer rotary coherence squared spectra (Figure 6–40) lies below, but close to, the 95% confidence limit for frequencies less than about 4 cycles per hour. The relationship between the clockwise rotating component of one vector and the anti-clockwise rotating vector of the other spectra is weaker than the relationship between the vectors when rotating in the likewise manner.

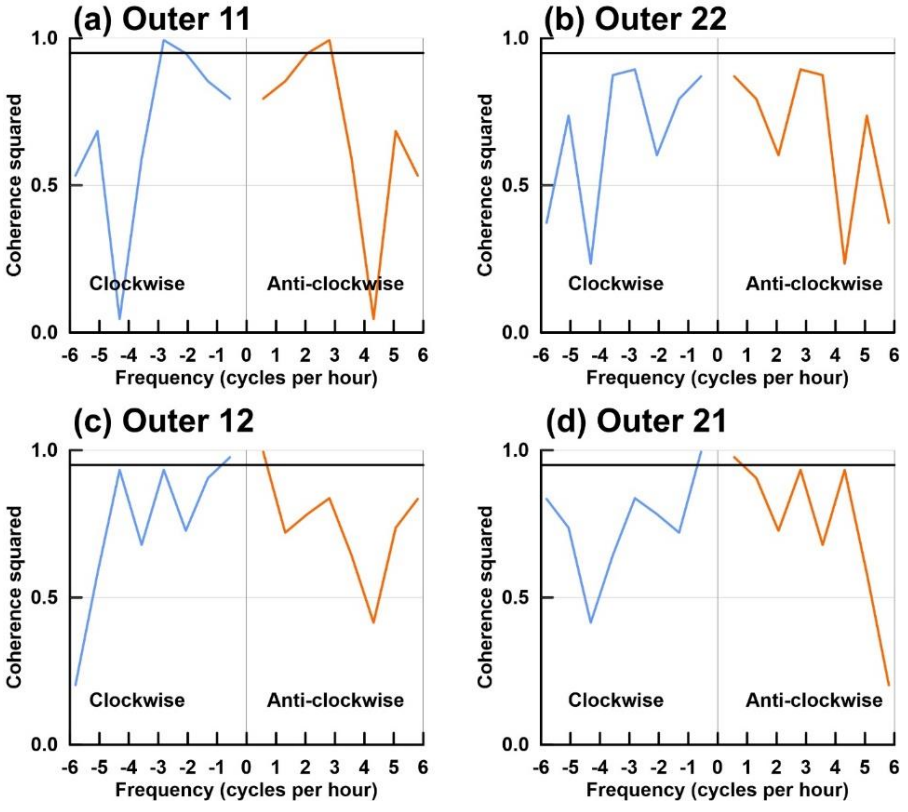


Figure 6–40. Outer rotary coherence squared spectra of the currents 15 m and 49 m above the sea floor. The black, dashed line represents the 95% confidence limit. Note, only one of the clockwise / anti-clockwise outer rotary coherence squared spectra need be plotted.

Chapter 7. Harmonic Analysis

The previous three chapters focused on different types of spectral analysis (auto, cross and rotary spectra). This chapter moves into harmonic analysis. The major difference between spectral analysis and harmonic analysis is that the former examines the data set in terms of all permissible frequencies of trigonometric functions that can fit a data set, while the latter examines specific frequencies, that are chosen *a priori*.

So why 'harmonic analysis' rather than 'spectral analysis'? There are three main reasons:

1. We may only be interested in a small number of spectral estimates, those at specific frequencies.
2. Frequencies in which we're interested may not be exactly represented in spectral analysis. Recall that spectral analysis only allows frequencies that are obtained from $f = \frac{p}{N\,\Delta t}\ for\ \ p = 1\ to\ N$ where p is an integer.
3. Spectral analysis may not allow a clear separation between frequencies that lie close together. For example, if we only have a few months of data, we cannot resolve (or distinguish between) the '2N2' and 'MU2' tidal constituents.

Spectral analysis ensures an exact representation of the data using a sum of sine and cosine curves, although the 'penalty' is that we are limited to the frequencies that can be resolved. Harmonic analysis enables us to solve the discrete Fourier expression at a specific frequency but the 'penalty' is that we can only achieve an approximate solution to the trigonometric equations that represent the data. Harmonic analysis finds the best fit (in a least squares sense) to the data using a smaller set number of sine and cosine curves of known frequency. The number of sine and cosine curves is always less than the 'all permissible frequencies of trigonometric functions that can fit a data set' that results from spectral analysis. The consequence is, that harmonic analysis can only ever achieve an approximate solution to the data set.

In short, spectral analysis gives us an exact trigonometric match to the data but there are limitations to the frequencies that are generated. Harmonic analysis allows us to select an exact frequency (or frequencies) but the fit to the data is only approximate.

Harmonic analysis is particularly useful when, for example, trying to identify, separate and/or remove the tidal signals from a data set. (Indeed, most of this chapter will use tides as the signals for analysis. Tidal theory is complex and is not considered here, we simply examine tidal data sets). The frequencies of the different tidal constituents are known. Therefore, the focus of harmonic lies in quantifying the relative amplitude and phase of the signals at each of these known frequencies.

Tides:

Harmonic analysis (and more generally spectral analysis) is a popular tool for the analysis of tidal signals and most of the work presented in this chapter focuses the analysis of tidal signals in a data set. Tides are the result of the relative position (and motions) of the earth, moon, sun and planets and the effect on water level can be well represented as the sum of a series of sine and cosine curves. However, the rotations of these celestial bodies are independent and this results in a complex set of interactions, in addition to the primary signals.

There are 45 major tidal constituents, although some 65 (or more) may be important in particular situations such as in rivers and estuaries (i.e. shallow waters). Apart from the presence of these celestial bodies, further complications arise because the individual rotations are ellipses, not circles, and the planes of rotation are not equatorial. In temperate regions, the change in water level due to the tides is usually in the order of 1-2 m.

Tides are not actually measured, it is water level that is measured. Apart from the tides, the water level may be affected by other processes including: atmospheric pressure (a change in atmospheric pressure of 1 hPa results in a change in water level of about 1 cm), storm surges, cyclones / typhoons / hurricanes, the land-sea breeze effect, tsunamis, bottom topography (shallow water), resonance in certain bays (e.g. water levels in the Bay of Fundy, Canada may exceed 15 m).

Such processes should be identified and, where possible, removed from the water level data prior to the analysis of tides. However, that is not the purpose of this book. Further details on the analysis of tides can be found in Godin (1972), Foreman (1996) and Parker (2007).

We demonstrate the differences between spectral and harmonic analyses below.

Let's assume we're only interested in the principal lunar semi-diurnal tide (the M2 component and the major tidal constituent). This component of the time has a period of 12.42 hours (or a frequency of 1.9324 cycles per day). Further, assume we have a data set comprising 6 weeks of hourly water levels ($N = 1{,}024$ data points and $\Delta t = 1/24$ days). We undertake our spectral analysis. Now, there are only certain permissible frequencies (f) that can be resolved. These are given by:

$f = \frac{p}{N\,\Delta t} \qquad for \;\; p = 1 \; to \; N$ (noting that p is an integer).

The p values that result in frequencies closest to that of the M2 tide, 1.9324 cycles per day are given by:

$p = 82$, then the corresponding frequency is $f = 1.9219$ cycles per day.

$p = 83$, then the corresponding frequency is $f = 1.9453$ cycles per day.

While both of these frequencies are close to the desired frequency, nether are exact. Some of the energy of the M2 tidal component is spread over both frequencies and the size of the M2 component may be underestimated, potentially compromising the subsequent interpretation of the spectrum (for the M2 tide).

Let's now assume that, for the same data set, we're interested in the 2N2 and MU2 tidal constituents (these have frequencies of $f = 1.8597\ and\ 1.8645$ cycles per day, respectively).

The closest p values to these frequencies are given by:

$p = 79$, and the corresponding frequency is $f = 1.8516$ cycles per day.

$p = 80$, and the corresponding frequency is $f = 1.8750$ cycles per day.

Frequencies associated with both the 2N2 and MU2 lie between these (consecutive) p values. Spectral analysis using $N = 1024$ records and $\Delta t = 1/24\ days\ (or\ 1\ hour)$ cannot separate these two constituents. They would appear as a single peak in the periodogram and their combined energies are spread over the frequencies $f = 1.8516$ and 1.8750 cycles per day.

Suppose we have 15 years of hourly data, then $N = 131{,}072$ data points and $\Delta t = 1/24$ days and the p values that result in frequencies closest to that of the M2 tide are given by:

$p = 10{,}553$, the corresponding frequency is $f = 1.9323$ cycles per day.

$p = 10{,}554$, the corresponding frequency is $f = 1.9325$ cycles per day.

These are much closer to the frequency of the M2 tide, but still not an exact match.

The p values closest to the frequencies associated with the 2N2 and MU2 are given by, respectively: $p = 10{,}156$ and $p = 10{,}183$.

This longer length of the data set allows us to separate the 2N2 and MU2 tidal constituents. If we have $N = 8{,}192$ and $\Delta t = \frac{1}{24}\ days$ (i.e. almost 1 year of hourly data) then we can just separate the 2N2 and MU2 tidal constituents.

The longer the data set, the closer we can get to an exact match of any desired frequency and the easier it becomes to separate the frequencies of two different signals. However, we may not have a long data set at our

disposal for the required analysis. Harmonic analysis enables us to select a specific frequency and determine the relative energy associated of features at that frequency with features at other, specific, frequencies.

7.1. Harmonic analysis using a scalar

From Section 2.3, the time series can be represented as follows:

$$x_t = a_o + \sum_{p=1}^{\frac{N}{2}-1} \left[a_p \, cos(\omega_p \, t) + \, b_p \, sin(\omega_p \, t)\right] + \, a_{N/2} \, cos(\pi t) \qquad for \quad t = 1, 2, 3, \ldots N$$

Where: $\omega_p = \frac{2\pi p}{N \, \Delta t} \;\; or \;\; f_p = \frac{p}{N \, \Delta t}$ and p is an integer. Hence $\omega_p \; (or \;\; f_p)$ takes on a specific set of values that, for a given Δt, is governed by the number of data points in the time series.

Let's assume that we're only interested in M such values of $f \;\; (or \;\; \omega)$ (i.e. M harmonics). At each harmonic there will be two (Fourier) coefficients and a total of $2M + 1$ harmonic coefficients. In general, the number of harmonic coefficients will be much less than the number of data points (i.e. $2M + 1 \ll N$). Mathematically, this means that we have an over-specified data set (i.e. we have more data points than we have equations) and, unlike spectral analysis, we cannot, in general, find an exact solution to the equation. We can optimise the values of the $2M + 1$ coefficients using a least squares method. Spectral analysis requires us to solve the usual equation, above.

However, instead of solving this equation for all values of p, we're going to split this equation in such a way that we have the M number of harmonics (each at a very specific frequency) plus all of the other information that is contained in the data set. We call all of this other information the 'residuals' or error (ε) which we want to minimise. At each time, t, we have:

$$x_t = A_o + \sum_{q=1}^{M}\left[A_q\, cos(\omega_q\, t) + B_q\, sin(\omega_q\, t)\right] + \left\{\sum_{\substack{p=1 \\ p\neq q}}^{\frac{N}{2}-1}\left[a_p\, cos(\omega_p\, t) + b_p\, sin(\omega_p\, t)\right] + a_{N/2}\, cos(\pi t)\right\} \qquad for \quad t = 1, 2, 3, \ldots N$$

$$x_t = A_0 + \sum_{q=1}^{M}\left[A_q\, cos(\omega_q\, t) + B_q\, sin(\omega_q\, t)\right] + residuals$$

Rearranging, we get:

$$residuals = x_t - A_0 - \sum_{q=1}^{M}\left[A_q\, cos(\omega_q\, t) + B_q\, sin(\omega_q\, t)\right]$$

$$(residuals)^2 = \left\{x_t - A_0 - \sum_{q=1}^{M}\left[A_q\, cos(\omega_q\, t) + B_q\, sin(\omega_q\, t)\right]\right\}^2$$

For all times, we then obtain:

$$\varepsilon^2 = \sum_{t=1}^{N}(residuals)^2$$

$$= \sum_{t=1}^{N}\left\{x_t - A_0 - \sum_{q=1}^{M}\left[A_q\, cos(\omega_q\, t) + B_q\, sin(\omega_q\, t)\right]\right\}^2$$

We then minimise the error term, such that:

$$\frac{\partial \varepsilon^2}{\partial A_0} = 0, \quad \frac{\partial \varepsilon^2}{\partial A_1} = 0 \quad and \quad \frac{\partial \varepsilon^2}{\partial B_1} = 0$$

And so on for other A_q's and B_q's.

This results in a set of simultaneous equations which can then be solved, using multiple linear regression, for the A_q's and B_q's.

Example: Assume $M = 1$. Then:

$$\frac{\partial \varepsilon^2}{\partial A_0} = 2\sum_{t=1}^{N}\{[x_t - A_0 - A_1\, cos(\omega_1\, t) - \, B_1\, sin(\omega_1\, t)][-1]\} = 0$$

$$\frac{\partial \varepsilon^2}{\partial A_1} = 2\sum_{t=1}^{N}\{[x_t - A_0 - A_1\, cos(\omega_1\, t) - \, B_1\, sin(\omega_1\, t)][-cos(\omega_1\, t)]\} = 0$$

$$\frac{\partial \varepsilon^2}{\partial B_1} = 2\sum_{t=1}^{N}\{[x_t - A_0 - A_1\, cos(\omega_1\, t) - \, B_1\, sin(\omega_1\, t)][-sin(\omega_1\, t)]\} = 0$$

Rearranging, step-by-step:

$$\sum_{t=1}^{N}\{[x_t - A_0 - A_1\, cos(\omega_1\, t) - \, B_1\, sin(\omega_1\, t)][-1]\} = 0$$

$$\sum_{t=1}^{N}\{[x_t - A_0 - A_1\, cos(\omega_1\, t) - \, B_1\, sin(\omega_1\, t)][-cos(\omega_1\, t)]\} = 0$$

$$\sum_{t=1}^{N}\{[x_t - A_0 - A_1\, cos(\omega_1\, t) - \, B_1\, sin(\omega_1\, t)][-sin(\omega_1\, t)]\} = 0$$

$$\sum_{t=1}^{N} x_t + \sum_{t=1}^{N} A_0 + \sum_{t=1}^{N} A_1\, cos(\omega_1\, t) + \sum_{t=1}^{N} B_1\, sin(\omega_1\, t) = 0$$

$$\sum_{t=1}^{N} x_t[-cos(\omega_1 t)] + \sum_{t=1}^{N} A_0 \, cos(\omega_1 t) + \sum_{t=1}^{N} A_1 \, cos^2(\omega_1 t) + \sum_{t=1}^{N} B_1 \, cos(\omega_1 t).sin(\omega_1 t) = 0$$

$$\sum_{t=1}^{N} x_t[-sin(\omega_1 t)] + \sum_{t=1}^{N} A_0 \, sin(\omega_1 t) + \sum_{t=1}^{N} A_1 \, cos(\omega_1 t).sin(\omega_1 t) + \sum_{t=1}^{N} B_1 \, sin^2(\omega_1 t) = 0$$

$$\sum_{t=1}^{N} A_0 + \sum_{t=1}^{N} A_1 \, cos(\omega_1 t) + \sum_{t=1}^{N} B_1 \, sin(\omega_1 t) = \sum_{t=1}^{N} x_t$$

$$\sum_{t=1}^{N} A_0 \, cos(\omega_1 t) + \sum_{t=1}^{N} A_1 \, cos^2(\omega_1 t) + \sum_{t=1}^{N} B_1 \, cos(\omega_1 t).sin(\omega_1 t) = \sum_{t=1}^{N} x_t[cos(\omega_1 t)]$$

$$\sum_{t=1}^{N} A_0 \, sin(\omega_1 t) + \sum_{t=1}^{N} A_1 \, cos(\omega_1 t).sin(\omega_1 t) + \sum_{t=1}^{N} B_1 \, sin^2(\omega_1 t) = \sum_{t=1}^{N} x_t[sin(\omega_1 t)]$$

$$NA_0 + \sum_{t=1}^{N} A_1 \, cos(\omega_1 t) + \sum_{t=1}^{N} B_1 \, sin(\omega_1 t) = \sum_{t=1}^{N} x_t$$

$$\sum_{t=1}^{N} A_0\, cos(\omega_1\, t) + A_1 \sum_{t=1}^{N} cos^2(\omega_1\, t) + B_1 \sum_{t=1}^{N} cos(\omega_1\, t).sin(\omega_1\, t) = \sum_{t=1}^{N} x_t[cos(\omega_1\, t)]$$

$$\sum_{t=1}^{N} A_0\, sin(\omega_1\, t) + A_1 \sum_{t=1}^{N} cos(\omega_1\, t).sin(\omega_1\, t) + B_1 \sum_{t=1}^{N} sin^2(\omega_1\, t) = \sum_{t=1}^{N} x_t[sin(\omega_1\, t)]$$

$$\begin{bmatrix} N & \sum_{t=1}^{N} cos(\omega_1\, t) & \sum_{t=1}^{N} sin(\omega_1\, t) \\ \sum_{t=1}^{N} cos(\omega_1\, t) & \sum_{t=1}^{N} cos^2(\omega_1\, t) & \sum_{t=1}^{N} cos(\omega_1\, t).sin(\omega_1\, t) \\ \sum_{t=1}^{N} sin(\omega_1\, t) & \sum_{t=1}^{N} cos(\omega_1\, t).sin(\omega_1\, t) & \sum_{t=1}^{N} sin^2(\omega_1\, t) \end{bmatrix} \begin{bmatrix} A_0 \\ A_1 \\ B_1 \end{bmatrix} = \begin{bmatrix} \sum_{t=1}^{N} x_t \\ \sum_{t=1}^{N} x_t[cos(\omega_1\, t)] \\ \sum_{t=1}^{N} x_t[sin(\omega_1\, t)] \end{bmatrix}$$

$$\begin{bmatrix} A_0 \\ A_1 \\ B_1 \end{bmatrix} = \begin{bmatrix} N & \sum_{t=1}^{N} cos(\omega_1 t) & \sum_{t=1}^{N} sin(\omega_1 t) \\ \sum_{t=1}^{N} cos(\omega_1 t) & \sum_{t=1}^{N} cos^2(\omega_1 t) & \sum_{t=1}^{N} cos(\omega_1 t).sin(\omega_1 t) \\ \sum_{t=1}^{N} sin(\omega_1 t) & \sum_{t=1}^{N} cos(\omega_1 t).sin(\omega_1 t) & \sum_{t=1}^{N} sin^2(\omega_1 t) \end{bmatrix}^{-1} \begin{bmatrix} \sum_{t=1}^{N} x_t \\ \sum_{t=1}^{N} x_t[cos(\omega_1 t)] \\ \sum_{t=1}^{N} x_t[sin(\omega_1 t)] \end{bmatrix}$$

And we can solve for A_0, A_1 and B_1. Clearly, this becomes more complex and certainly more tedious when $M \gg 1$.

There are many statistical packages available that do harmonic analysis for tidal signals, perhaps the most widely used is that of Foreman (1996). Foreman's package relates all time information back to 00:00 hours on 1 January 1900 Greenwich time (Coordinated Universal Time, UTC) and requires knowledge of much ancillary tidal information, such as astronomical arguments, Doodson numbers, locations and so on. The harmonic analysis used here dispenses with this information and all results are relative to the start of the record. This is useful for an initial examination of the data, but a more detailed analysis would warrant the use of Foreman's tidal analysis package (or equivalent).

7.1.1. Example 7-1: Water level using 6 weeks of hourly data

Let's assume that we're interested in the semi-diurnal and diurnal tidal constituents in a six-week data set comprising hourly water levels.

Question: How well does spectral analysis (using 6 weeks of hourly data) resolve the diurnal and semi-diurnal tides?

Data used:

- Hourly data over a 6-week period (1,024 hours).
- The Nyquist period is 2 hours (Nyquist frequency is ½ cycles per hour). This period is much less than the minimum period in which we're interested (12.42 hours or a frequency of 1/12.42 = 0.08 cycles per hour).
- The total length of the data set is 6 weeks or 42 days. Our rule of thumb requires about 5 cycles to resolve the longest period features in which we're interested (i.e. the diurnal period). So, we should be able to use spectral analysis to answer our question.

Pre-processing:

- A linear trend was removed from the data to produce a set of water level anomalies (Figure 7–1).

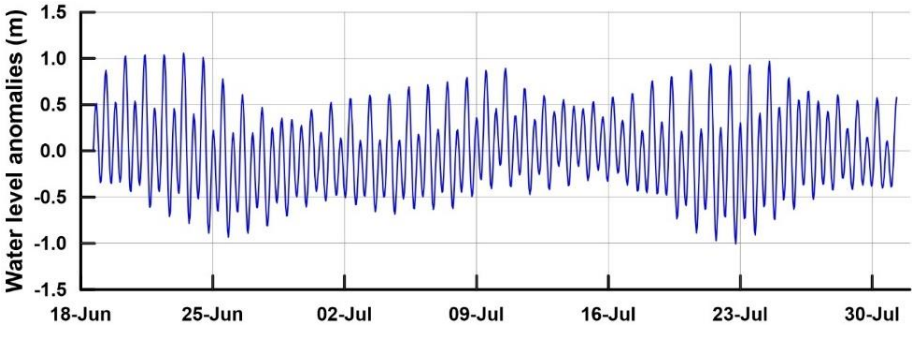


Figure 7–1. Six weeks of hourly water level data.

Testing:

- Stationarity: Results from the Dickey-Fuller test give Tau calculated values of -7.34 for the water level anomalies; the Tau critical value is - 2.86. As the calculated Tau is less than Tau critical for both data sets, we can reject the null hypothesis that the data sets are not stationary.
- All data anomalies (i.e. with the means removed) were normally distributed. Data were split into 16 groups each comprising 32 data points. Bartlett's test for homogeneity of variance indicated that the data sets comprised samples from the same population at the 5% level of significance. Similarly, ANOVA tests indicated no significant difference in the means of each data set at the 5% level of significance i.e. the different data sets are likely to have come from the same population. Therefore, the data are likely to be stationary and we proceed with the analysis.

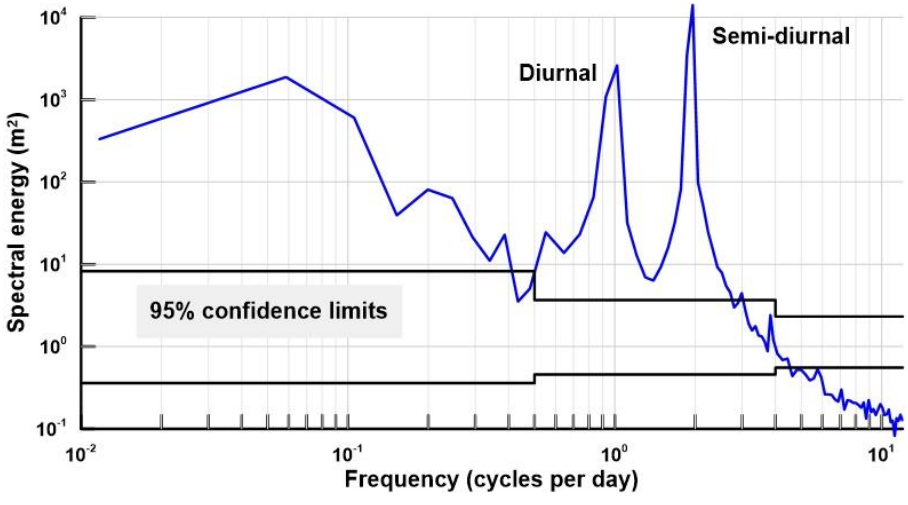


Figure 7–2. Spectral analysis of six weeks of hourly water level data.

Smoothing of the periodogram:

- The upper and lower 95% confidence limits were generated by averaging 2 spectral bands over the first 50 frequency values, 4 spectral bands averaged over the next 100 frequency values and 8 spectral bands averaged over the remaining frequency values.

Description:

- The amplitude (or energy) spectrum is shown in Figure 7–2.
- Two spectral peaks are apparent – one (a 'double peak') near the diurnal period and the other peak near the semi-diurnal period.

Returning to our question of interest.

(a) We can certainly identify spectral peaks that coincide with the frequencies of the diurnal and semi-diurnal tides.

So, it appears that the spectral analysis has achieved what it was set out to do: viz. resolve the diurnal and semi-diurnal tides. However, we know that there are 45 main tidal constituents, at least 10 of which are associated with the diurnal tides and at least 12 of which are associated with the semi-diurnal tides. Instead of the two observed diurnal spectral peaks, we should see 10 spectral peaks and instead of the one observed semi-diurnal spectral peak, we should see 12 spectral peaks.

Why can't we see all of these spectral peaks? The fundamental problem is that the allowable frequencies (which are based on the p values and the total length of the data set i.e. 6 weeks, $\omega_p = \frac{2\pi\, p}{N\, \Delta t}$) do not exactly match the tidal frequencies of interest. A second problem arises because the allowable frequencies cannot effectively resolve the tidal constituents within the diurnal (or semi-diurnal) frequency bands. To do so, we need to reduce the frequency for each p-value.

This can be done by increasing N or Δt. Recall that p is simply a counter which only takes on integer values and it cannot be reduced. We can increase Δt, but in so doing we will eventually run into the problem of the Nyquist frequency exceeding the frequencies of the tidal constituents of interest. To resolve the semi-diurnal tidal constituents, we are limited to a Δt value of 6 hours. Even at 6 hourly sampling, not all of the semi-diurnal tidal constituents can be resolved.

The only practical solution is to increase N, in other words, collect data over a much longer period of time. However, we simply may not have sufficient data to increase N, which may need to be very large to exactly match the frequency of interest.

One approach to solving this problem is to use harmonic analysis on the 6-week data set (Figure 7–1) to examine the specific frequencies of interest. Firstly, we need to determine which frequencies we want to examine i.e. which are the tidal constituents that we want to examine?

Let's use harmonic analysis to resolve the two main diurnal tidal constituents and the two main semi-diurnal tidal constituents. In this case, $M = 4$, corresponding to the following tidal constituents (note $\omega = 2\pi f$):

O_1 (period = 25.819 hours, f_1 = 24/25.819 = 0.9295 cycles per day),

K_1 (period = 23.934 hours, f_2 = 24/23.934 = 1.0028 cycles per day),

M_2 (period = 12.421 hours, f_3 = 24/12.421 = 1.9322 cycles per day), and

S_2 (period = 12.000 hours, f_4 = 24/12.000 = 2.0000 cycles per day).

From Section 7.1 we set up the equation we need to solve.

$$\varepsilon^2 = \sum_{t=1}^{N} \left\{ x_t - A_0 - \sum_{q=1}^{4} \left[A_q \cos(\omega_q t) + B_q \sin(\omega_q t) \right] \right\}^2$$

With four tidal constituents, we need to solve this equation for A_q and B_q ($q = 1$ *to* 4), plus A_0 by minimising the error term, ε^2. The set up is a 9x9 matrix, which is too large to show here. However, the matrix does have a similar structure to the simple matrix shown in Section 7.1 and it is

relatively easy to solve using, for example, Matlab, a Fortran program (using the Gauss-Jordan elimination routine in Numerical Recipes) or indeed Excel (although the coding in Excel is relatively long and complex compared with the other two packages). Results from the harmonic analysis are presented in Table 7–1. We have an amplitude (and phase) for each of the four tidal constituents, rather than the broad peaks near the diurnal and semi-diurnal periods that result from the spectral analysis (Figure 7–2).

Table 7–1. Amplitude and phase for four tidal constituents using the data shown in Figure 7–1.

q	A_q (m^2)	B_q (m^2)	Tidal constituent	Freq. (cycles / day)	Amp. (m^2)	Phase (deg.)
0	1.0332				1.033	
1	-0.1488	-0.4501	M2 - Principal lunar semidiurnal	1.9323	0.474	198
2	0.0746	-0.0897	S2 - Principal solar semidiurnal	2.0000	0.117	140
3	0.1104	-0.1381	K1 - Lunisolar diurnal	1.0027	0.177	141
4	-0.0645	-0.0869	O1 - Principal lunar diurnal	0.9295	0.108	217

Note: The phase is relative to the start of the record.

This harmonic analysis is repeated using eight tidal constituents and the results presented in Table 7–2. (Note: The S2 and K2 tidal constituents have similar frequencies, 2.0000 and 2.0055 cycles per day, respectively. For $N = 1{,}024$ and $\Delta t = \frac{1}{24}$ days the S2 and K2 tidal constituents can just be separated, using $p = 85$ and 86, respectively). A comparison of the tidal components estimated using four (Table 7–1) and eight (Table 7–2) harmonics reveals slight differences in the corresponding amplitudes and phases. This results from the different least squares optimisation schemes used. Using 17 variables i.e. 8 harmonics will provide a slightly better fit to the overall data set (average root mean square of the differences = 0.082), compared with 9 variables i.e. 4 harmonics (average root mean square of the differences = 0.164). Using 8 harmonics can be thought of as a 'fine tuning' of the values of the amplitude and phase.

Table 7–2. Amplitude and phase for eight tidal constituents using the data shown in Figure 7–1.

q	A_q (m)	B_q (m)	Tidal constituent	Freq. (cycles / day)	Amp. (m)	Phase (deg.)
0	1.0330				1.033	
1	-0.1330	-0.4645	M2 - Principal lunar semidiurnal	1.9323	0.483	196
2	0.0984	-0.0769	S2 - Principal solar semidiurnal	2.0000	0.125	128
3	-0.1019	0.0471	N2 - Larger lunar elliptic semidiurnal	1.8960	0.112	295
4	-0.0175	0.0313	K2 - Lunisolar semidiurnal	2.0055	0.036	331
5	0.1370	-0.0986	K1 - Lunisolar diurnal	1.0027	0.169	126
6	-0.0611	-0.0903	O1 - Principal lunar diurnal	0.9295	0.109	214
7	0.0464	-0.0174	P1 - Principal solar diurnal	0.9973	0.050	111
8	-0.0102	0.0213	Q1 - Larger lunar elliptic diurnal	0.8932	0.024	334

Note: The phase is relative to the start of the record.

Spectral analysis allows us to identify the 'exact' spectral energy, but only at selected values of the frequency. Harmonic analysis allows us to select exact values of the frequency, but the penalty is that we can only ever get

an approximate solution to our equation. The amplitude (analogous to the spectral energy) is only an approximate value.

7.1.2. Example 7-2: Water level using 15 years of hourly data

Let's reconsider this problem using a 15 year data set of hourly water levels (Figure 7–3). The linear trend has been removed. To enable the generation of confidence limits, we averaged 4 spectral bands at low frequencies, 8 spectral bands at intermediate frequencies and 16 spectral bands at high frequencies.

Question: How well does spectral analysis (using 15 years hourly data) resolve the diurnal and semi-diurnal tides?

Data used:

- Hourly data over 15 years (131,072 hours).
- The Nyquist period is 2 hours (Nyquist frequency is ½ cycles per hour). This period is much less than the minimum period in which we're interested (12.42 hours or a frequency of 1/12.42 = 0.08 cycles per hour).
- The total length of the data set is 15 years, much more than our rule of thumb of 5 cycles for resolving long period features. So, we should be able to use spectral analysis to answer our question.

Pre-processing:

- A linear trend was removed from the data to produce a set of water level anomalies (Figure 7–3).

Testing:

- Stationarity: Results from the Dickey-Fuller test give Tau calculated values of -87.05 for the water level anomalies; the Tau critical value is - 2.86. As the calculated Tau is less than Tau critical for both data sets, we can reject the null hypothesis that the data sets are not stationary.

- Based on the results from previous examples, it was assumed that the alternate method for stationarity would result in a fail. Particularly so, as we have a very large number of data points.

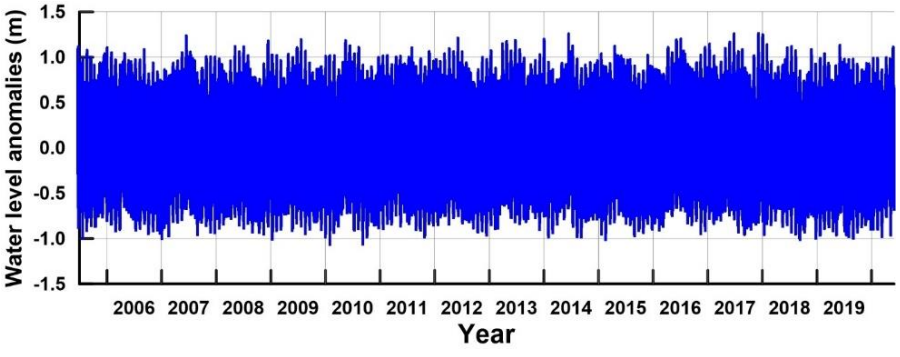


Figure 7–3. Fifteen years of hourly water level anomalies.

Smoothing of the periodogram:

- The upper and lower 95% confidence limits were generated by averaging 2 spectral bands over the first 50 frequency values, 4 spectral bands averaged over the next 100 frequency values and 8 spectral bands averaged over the remaining frequency values.

Description:

- Results from the spectral analysis are presented in Figure 7–4. The bottom plot (Figure 7–4, blue coloured line) represents the whole spectrum while the top plot (Figure 7–4, red coloured line) highlights the area of the spectrum that contains the major spectral peaks.
- For clarify, the 95% confidence limits are only shown on the former plot.
- The major tidal constituents have been labelled on the latter plot (only 19 of the 37 tidal constituents are shown).

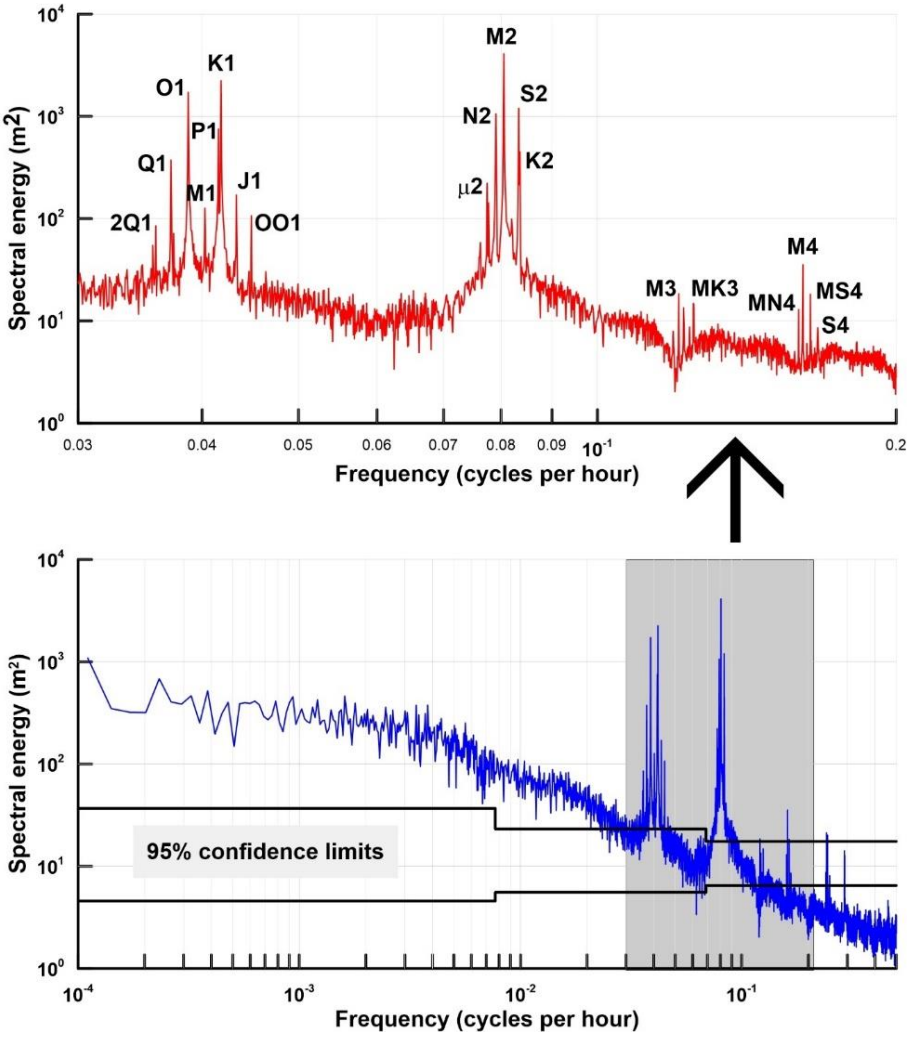


Figure 7–4. Spectral analysis of fifteen years of hourly water level data. The top plot is a subset of the shaded area in the bottom plot with the spectral peaks labelled with the relevant tidal constituents.

Returning to our question of interest.

(a) We can certainly identify multiple spectral peaks that coincide with the frequencies of different constituents associated with the diurnal and semi-diurnal tides. By using a much longer data set, we can better resolve the different frequency components (in this case, the tidal constituents).

Compared with the results from Example 7-1, there is a marked improvement in the resolution of the different diurnal and semi-diurnal tidal constituents. Eight of the 10 diurnal tidal constituents can be clearly identified, but the remaining two diurnal tidal constituents are unidentified. Five (and possibly seven) of the 12 semi-diurnal tidal constituents can be identified, leaving seven semi-diurnal tidal constituents unidentified. It is noted that some tidal constituents may not be important at the location that has been selected, which may explain why they cannot be clearly identified.

Results from the harmonic analysis for 37 tidal constituents are shown in Table 7–3.

Table 7–3. Amplitude and phase for 37 tidal constituents using the data shown in Figure 7–3. Cells shaded in yellow are those constituents identified in Figure 7–4.

q	Tidal constituent	Freq. (cycles / day)	Amp. (m)	Phase (deg.)
1	M8 - Shallow water eighth diurnal	7.7291	0.000	296
2	S6 - Shallow water over-tides of principal solar	6.0000	0.000	347
3	M6 - Shallow water over-tides of principal lunar	5.4419	0.000	228
4	S4 - Shallow water over-tides of principal solar	4.0000	0.001	94

q	Tidal constituent	Freq. (cycles / day)	Amp. (m)	Phase (deg.)
5	MS4 - Shallow water quarter diurnal	3.9323	0.002	294
6	M4 - Shallow water over-tides of principal lunar	3.8645	0.004	30
7	MN4 - Shallow water quarter diurnal	3.8283	0.001	60
8	MK3 - Shallow water terdiurnal	2.9350	0.000	11
9	M3 - Lunar terdiurnal	2.8984	0.002	345
10	2 MK3 - Shallow water terdiurnal	2.8618	0.001	44
11	2SM2 - Shallow water semidiurnal	2.0677	0.001	120
12	K2 - Lunisolar semidiurnal	2.0055	0.037	43
13	R2 - Smaller solar elliptic	2.0027	0.002	71
14	S2 - Principal solar semidiurnal	2.0000	0.124	37
15	T2 - Larger solar elliptic	1.9973	0.009	121
16	L2 - Smaller lunar elliptic semidiurnal	1.9686	0.014	128
17	Lamda2 - Smaller lunar evectional	1.9637	0.006	232
18	M2 - Principal lunar semidiurnal	1.9323	0.501	353
19	v2 - Larger lunar evectional	1.9008	0.020	266
20	N2 - Larger lunar elliptic semidiurnal	1.8960	0.113	28
21	MU2 - Variational	1.8645	0.021	127

q	Tidal constituent	Freq. (cycles / day)	Amp. (m)	Phase (deg.)
22	2 N2 - Lunar elliptical semidiurnal second-order	1.8597	0.020	236
23	OO1 - Lunar diurnal	1.0759	0.006	1
24	J1 - Smaller lunar elliptic diurnal	1.0390	0.010	236
25	K1 - Lunar diurnal	1.0027	0.148	84
26	S1 - Solar diurnal	1.0000	0.001	5
27	P1 - Solar diurnal	0.9973	0.043	245
28	M1 - Smaller lunar-elliptic diurnal	0.9661	0.003	297
29	O1 - Lunar diurnal	0.9295	0.097	5
30	rho - Larger lunar evectional diurnal	0.8981	0.004	273
31	Q1 - Larger lunar elliptic diurnal	0.8932	0.023	29
32	2Q1 - Larger elliptic diurnal	0.8570	0.004	39
33	Mf - Lunisolar fortnightly	0.0732	0.004	23
34	Msf - Lunisolar synodic fortnightly	0.0677	0.005	283
35	Mm - Lunar monthly	0.0363	0.005	28
36	Ssa - Solar semi-annual	0.0055	0.030	260
37	Sa - Solar annual	0.0027	0.044	342

Note: The phase is relative to the start of the record.

7.2. Harmonic analysis with a vector

The previous section used harmonic analysis with a scalar – water level. In much the same way that cross-spectral analysis (Section 5.1) can be used to examine a vector, harmonic analysis can also be used with a vector. Likewise, this generates much more information than just the magnitude and phase lag of each harmonic.

We proceed using a similar approach to that shown in Section 6.1.1 (rotary spectra for a single time series). The result is a 'tidal ellipse' associated with each harmonic. The characteristics that emerge are, for each harmonic: the magnitude of the major and minor semi-axes, the orientation of the ellipse, the phase (or starting point) of the vector that rotates around the ellipse and the direction in which the vector rotates. This is analogous to the output from a cross-spectral analysis. Strictly, this can be done for any harmonic although the interpretation of the results may be difficult. Hence, we'll restrict ourselves to harmonics associated with the tidal constituents.

Firstly, remove the mean and linear trend from the data sets, as usual.

Let d and δ be the amplitude and phase of the wth harmonic from the first time series (or one component of the vector). Let e and ε be the amplitude and phase of the wth harmonic from the second time series (or second component of the vector).

These are represented as rotary components R_1 and R_2 (anti-clockwise rotating components) and R_3 and R_4 (clockwise rotating components), see Section 6.1.1 where $R_1 = A + D;\ R_2 = B - C;\ R_3 = B + C\ and\ R_4 = A - D$.

$$R_1 = d\ \cos(\delta)\ +\ e\ \sin(\varepsilon) \qquad and \qquad R_2 = d\ \sin(\delta) -\ e\ \cos(\varepsilon)$$

$$R_3 = d\ \sin(\delta)\ +\ e\ \cos(\varepsilon) \qquad and \qquad R_4 = d\ \cos(\delta) -\ e\ \sin(\varepsilon)$$

The amplitude (A_{amp}) and phase (σ) of the anti-clockwise rotating component are given by:

$$A_{amp} = 0.5\sqrt{R_1^2 + R_2^2} \qquad and \qquad \sigma = \tan^{-1}\left(\frac{-R_2}{R_1}\right)$$

The amplitude (C_{amp}) and phase (γ) of the clockwise rotating component are given by:

$$C_{amp} = 0.5\sqrt{R_3^2 + R_4^2} \qquad and \qquad \gamma = \tan^{-1}\left(\frac{-R_4}{R_3}\right)$$

The amplitudes of the major and minor semi-axes are given by, respectively:

$$MAJOR = A_{amp} + C_{amp} \qquad and \qquad MINOR = A_{amp} - C_{amp}$$

If the sign of the $MINOR$ semi-axis is positive, the vector is rotating in an anti-clockwise direction, while if the sign is negative, the vector is rotating in a clockwise direction. `

The orientation of the ellipse (α) is given by (with respect to the positive real axis):

$$\alpha = \frac{(\sigma + \gamma)}{2}$$

There is an ambiguity (of 180°) in the orientation of the ellipse.

The phase (ϕ, with respect to the orientation of the major semi-axis) is given by:

$$\phi = \frac{(\gamma - \sigma)}{2}$$

The phase is relative to the start of the data set and there is also a 180° ambiguity in the phase (resulting in the ambiguity in the orientation of the ellipse).

7.2.1. Example 7-3: Currents with weak tidal signals

The first example uses currents that have a weak tidal signal. The time series of these data is shown in Figure 7–5, in stick vector form. No clear tidal signals can be identified. The trend has not been removed in the time series plot, but has been removed for the subsequent processing.

Question: What are the four dominant tidal signals in the spectrum? What are their harmonic characteristics? What do their tidal ellipses look like?

Data used:

- Current speed and direction (in east-west and north-south component form) from mid-water data, water depth is approximately 65 m.
- The data set is 12 months long (8,760 records).
- The time interval is 1 hour.
- The Nyquist period is 2 hours (Nyquist frequency of ½ cycles per hour).
- The longest period that we can practically resolve is about 1 month. Noting that our rule of thumb is that, to effectively resolve long period features, we need about 5-10 cycles in the data.
- The length of the data set and the value of delta-t is sufficient to effectively resolve the diurnal and semi-diurnal tidal signals.

Pre-processing:

- Data were truncated to 8,192 records, which enables the radix 2 FFT algorithm to be used.
- The linear trend was removed from both data sets.

Testing:

- Stationarity: Results from the Dickey-Fuller test give Tau calculated values of -26.03 for the east-west current component and -9.62 for the north-south current component; the Tau critical value is - 2.86. As the calculated Tau is less than Tau critical for both data sets, we can reject the null hypothesis that the data sets are not stationary.
- Based on the results from previous examples, it was assumed that the alternate method for stationarity would result in a fail. Particularly so, as we have a very large number of data points.

Smoothing of the periodogram:

- The upper and lower 95% confidence limits were generated by averaging 4 spectral bands over the first 400 frequency values, 8 spectral bands averaged over the next 800 frequency values and 16 spectral bands averaged over the remaining frequency values.

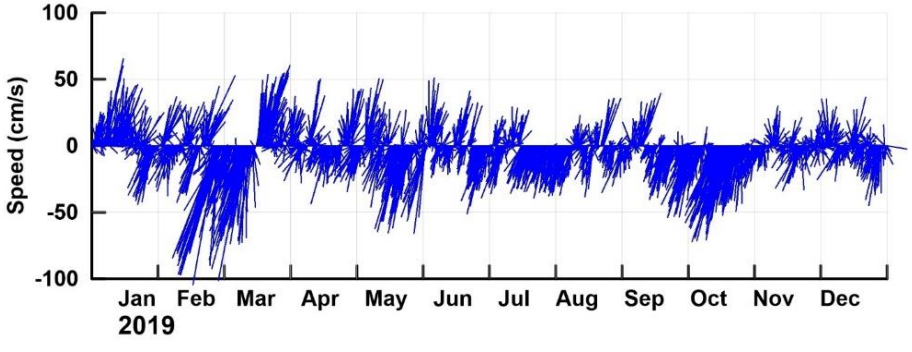


Figure 7–5. Time series of currents in stick vector form.

Description:

- The spectral analyses of the two current components are shown in Figure 7–6.
- Most of the energy occurs at low frequencies, despite the linear trend being removed from the data prior to analysis.

- The north-south component dominates the spectral energy, about one order of magnitude greater than the east-west component for frequencies less than about 0.5 cycles per day.
- Spectral peaks can be observed near the diurnal and semi-diurnal frequencies, the former being greater than the latter. However, these peaks are at least one order of magnitude less than the low frequency signal.
- Combined, the total energy in the diurnal and semi-diurnal bands represents about 5% of the total spectral energy (from Parseval's theorem).
- In contrast, the low frequency signals (those less than 0.5 cycles per day) represent about 75% of the total spectral energy.
- The tidal signal is apparent, but it is weak compared with other signals in the data set.
- Focus is on four tidal constituents: M2, S2, K1 and O1. Results of a harmonic analysis at the frequencies associated with these four tidal constituents are presented in Table 7–4. These results are also presented graphically in Figure 7–7.
- The K1 tidal component has the greatest magnitude of the tidal signals, although it is only about 8 cm/s. This component was also the maximum signal (at the tidal frequencies) in the spectrum.
- Three of the four constituents (M2, K1 and O1) rotate in an anti-clockwise direction, following the paths indicated by the red lines in Figure 7–7. The scale for all figures is the same. The black circles indicate the starting point for each component. It takes 12.42 hours for the black circle associated with the M2 plot to complete one revolution around its ellipse, 24.00 hours for the S2, 23.94 hours for K1 and 25.82 hours for O1.

Table 7–4. Ellipse characteristics for four tidal constituents using the data shown in Figure 7–5.

Tidal constituent	Freq. (cycles / day)	Period (hrs)	Major semi-axis (cm/s)	Minor semi-axis (cm/s)	Ellipse orient-ation (deg.)	Phase (deg.)
M2 - Principal lunar semidiurnal	1.9323	12.42	3.23	1.06	70	-23
S2 - Principal solar semidiurnal	2.0000	12.00	1.36	-0.16	-140	-35
K1 - Lunisolar diurnal	1.0027	23.94	7.32	1.95	82	-82
O1 - Principal lunar diurnal	0.9295	25.82	1.23	0.29	-1	152

Note: The phase is relative to the start of the record. There is a 180° ambiguity in the ellipse orientation, which results in it appearing negative on occasions.

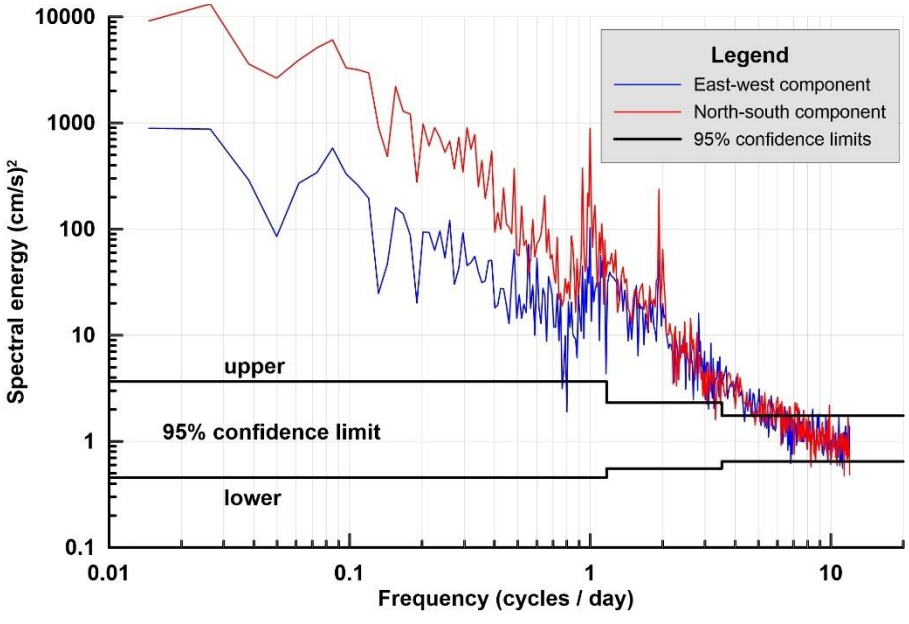


Figure 7–6. Spectral analysis of the current components using the data from Figure 7–5.

Returning to the questions asked:

(a) The four dominant tidal signals in the spectrum are the K1, O1, M2 and S2 constituents (Figure 7–6). While these signals are clearly identified in the data set, they remain one order of magnitude less than the low frequency end of the spectrum. This is despite the linear trend being removed from the data. This suggests additional low frequency elements in the data set that need to be investigated.

(b) Their harmonic characteristics are presented in Table 7–4. The largest tidal currents are associated with the K1 tides (about 8 cm/s). The average current speed over the whole data set is 20 cm/s. The three dominant tidal constituents (O1, K1 and M2) rotate in an anti-clockwise direction, as expected at this location. Motion of the S2 tidal constituent is clockwise, although its motion is almost rectilinear.

(c) Tidal ellipses for these four constituents are presented in Figure 7–7. The dominant tidal signal is K1, which is also the highest tidal peak in the spectrum (Figure 7–6).

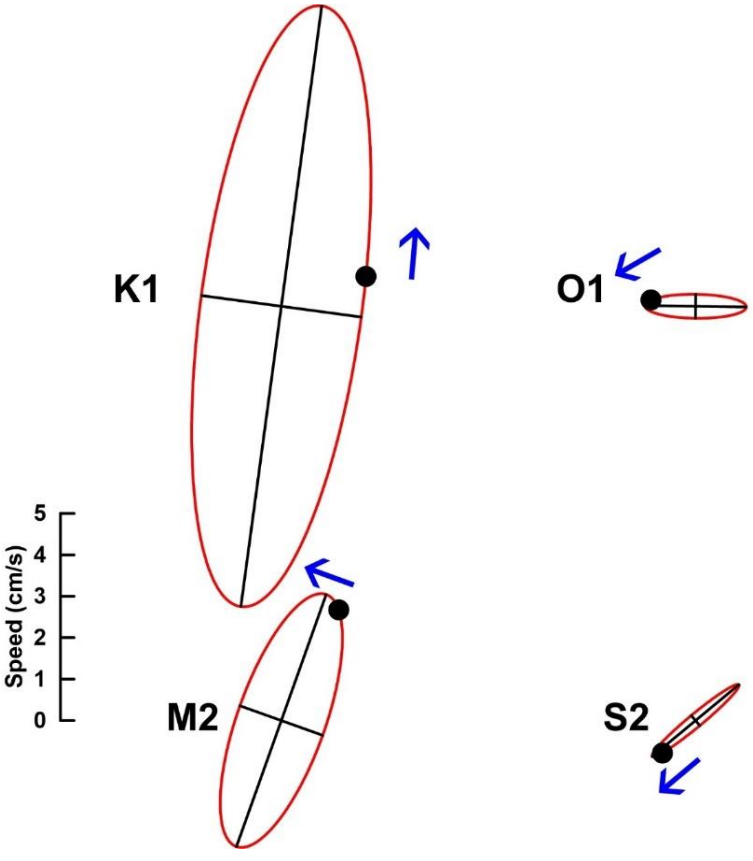


Figure 7–7. Tidal ellipses for four constituents. Black circles indicate the starting point for the phase, while the blue arrows indicate the direction of rotation of the currents around the ellipse.

7.2.2. Example 7-4: Currents with strong tidal signals

The second example uses currents (near the bottom of an estuary) that have a strong tidal signal. The time series of these data is shown in Figure 7–8, in component form. Oscillations in the currents (particularly in the east-west component) occur at the tidal frequencies. The trend has not

been removed in the time series plot but has been removed for subsequent processing.

Question: What are the four dominant tidal signals in the spectrum? What are their harmonic characteristics? What do their tidal ellipses look like?

Data used:

- Current speed and direction (in east-west and north-south component form) from mid-water data, water depth is approximately 65 m.
- The data set is about 57 days long (8,192 records).
- The time interval is 10 minutes.
- The Nyquist period is 20 minutes (Nyquist frequency of 1/3 cycles per hour).
- The longest period that we can practically resolve is about 1 week. Noting that our rule of thumb is that, to effectively resolve long period features, we need about 5-10 cycles in the data.
- The length of the data set and the value of delta-t is sufficient to effectively resolve the diurnal and semi-diurnal tidal signals.

Pre-processing:

- The linear trend was removed from both data sets.

Testing:

- Stationarity: Results from the Dickey-Fuller test give Tau calculated values of -6.25 for the east-west current component and -31.32 for the north-south current component; the Tau critical value is - 2.86. As the calculated Tau is less than Tau critical for both data sets, we can reject the null hypothesis that the data sets are not stationary.
- Based on the results from previous examples, it was assumed that the alternate method for stationarity would result in a fail. Particularly so, as we have a very large number of data points.

Smoothing of the periodogram:

- The upper and lower 95% confidence limits were generated by averaging 4 spectral bands over the first 400 frequency values, 8 spectral bands averaged over the next 800 frequency values and 16 spectral bands averaged over the remaining frequency values.

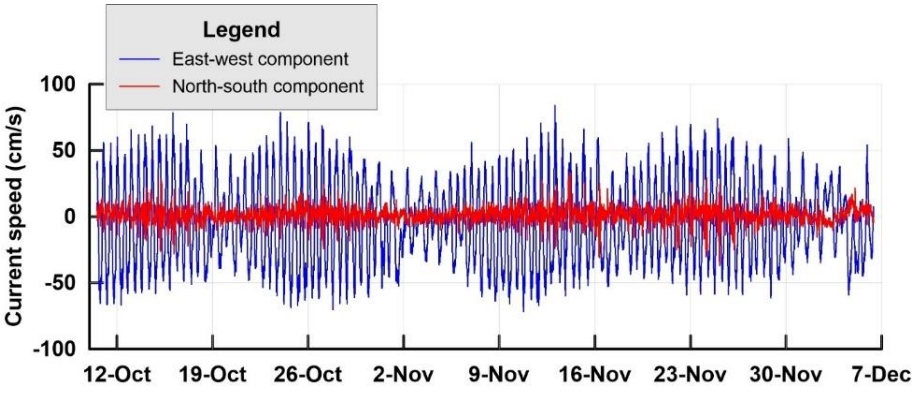


Figure 7–8. Time series of currents in component form.

Description:

- The spectral analyses of the two current components are shown in Figure 7–9. Spectral peaks can be observed at the tidal frequencies. Energy in the spectra (particularly the spectrum associated with the east-west currents) is greatest at the semi-diurnal and diurnal frequencies.
- The dominant flow occurs in the east-west direction. From Parseval's theorem, spectral energy near the diurnal band (for the east-west currents) accounts for about 10% of the variance in the original data set. Spectral energy near the semi-diurnal band (for the east-west currents) accounts for almost 80% of the variance in the original data set. The tidal signal is strong, accounting for about 90% of the variance in the data set.

- Focus is on four tidal constituents: M2, S2, K1 and O1. Results of a harmonic analysis at the frequencies associated with these four constituents are presented in Table 7–5 and in Figure 7–7.

Table 7–5. Ellipse characteristics for four tidal constituents using the data shown in Figure 7–8.

Tidal constituent	Freq. (cycles / day)	Period (hrs)	Major semi-axis (cm/s)	Minor semi-axis (cm/s)	Ellipse orient-ation (deg.)	Phase (deg.)
M2 - Principal lunar semidiurnal	1.9323	12.42	37.3	-1.3	42	84
S2 - Principal solar semidiurnal	2.0000	12.00	15.7	-1.1	13	-34
K1 - Lunisolar diurnal	1.0027	23.94	9.6	-0.4	-33	-79
O1 - Principal lunar diurnal	0.9295	25.82	5.7	-0.0	-5	38

Note: The phase is relative to the start of the record.

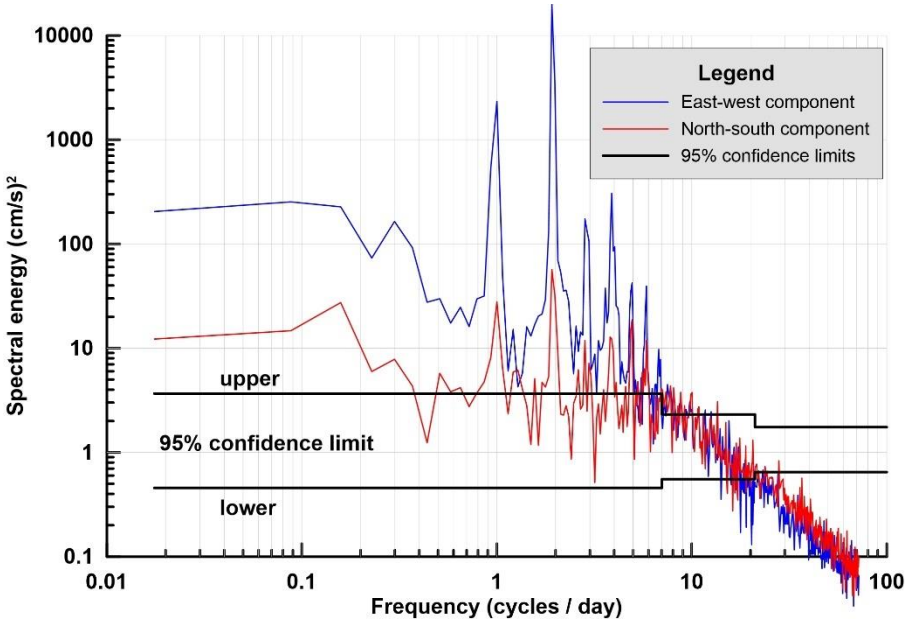


Figure 7–9. Spectral analysis of the current components using the data from Figure 7–8.

Returning to the questions asked:

(a) The dominant tidal signals in the spectrum are at the diurnal and semi-diurnal periods. However, the data set is of insufficient length the differentiate between the K1 and O1 diurnal constituents and the M2 and S2 semi-diurnal constituents (Figure 7–9). While these two signals dominate the spectral energy, other significant spectral peaks are also observed. The diurnal and semi-diurnal signals are substantially greater than the low frequency signals, in contrast to the previous example.

(b) The harmonic characteristics of the O1, K1, M2 and S2 tidal constituents are presented in Table 7–5. The largest tidal currents are associated with the M2 tides (about 45 cm/s). The average current speed over the whole data set is 20 cm/s. The three dominant tidal constituents (O1, K1 and M2) rotate in an anti-

clockwise direction, as expected at this location. Motion of the S2 tidal constituent is clockwise.

(c) Tidal ellipses for these four constituents are presented in Figure 7–10. The dominant tidal signal is M2. All four tidal ellipses have large eccentricity – their motion is almost rectilinear. This location is in a tidal river in the Northern Hemisphere. The water is constrained by the boundaries of the river, hence the tidal ellipses are (almost) bidirectional. Given this high eccentricity, the direction or rotation of the tidal currents loses its meaning. The scale for all figures is the same. The black circles indicate the starting point for each component. It takes 12.42 hours for the black circle associated with the M2 plot to complete one revolution around its ellipse, 24.00 hours for the S2, 23.94 hours for K1 and 25.82 hours for O1.

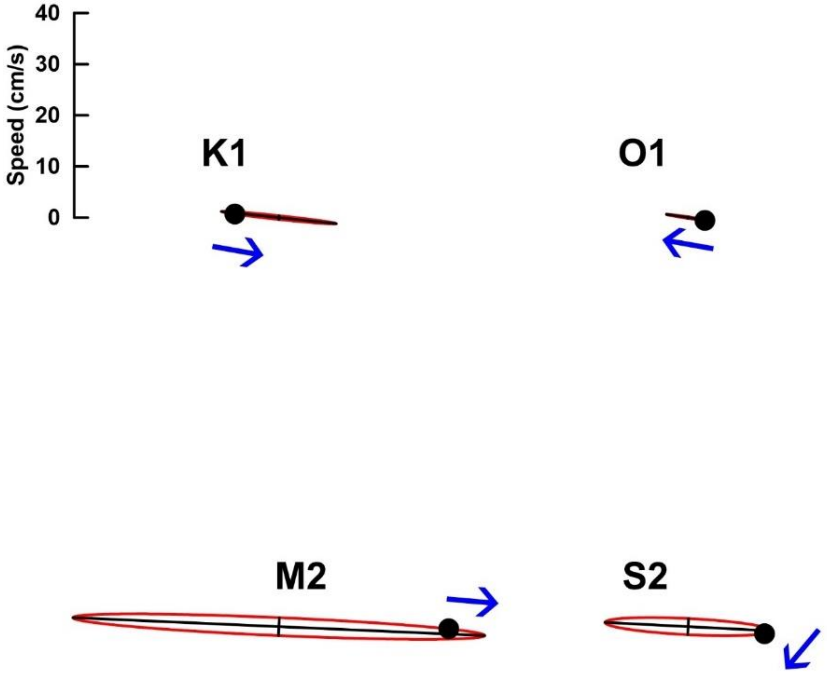


Figure 7–10. Tidal ellipses for four constituents. Black circles indicate the starting point for the phase, while the blue arrows indicate the direction of rotation of the currents around the ellipse.

7.3. Harmonic analysis in space

This section actually describes spectral (rather than harmonic) analysis in space. Strictly, it should form part of Chapter 4. However, it is included here because we feel that the interpretation of the results can be better visualised in terms of harmonics.

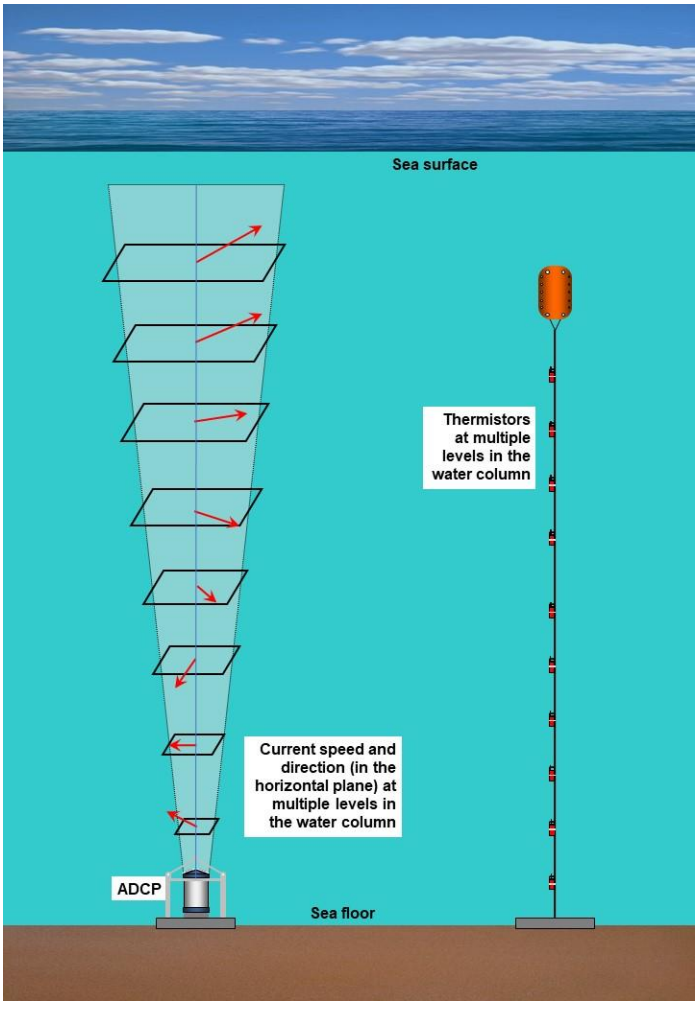


Figure 7–11. Current speed and direction and temperature from multiple vertical levels in the water column.

Thus far, we've focused on harmonic analysis (and indeed spectral analysis) in the time / frequency domain, but there is nothing to prevent us from undertaking the same analyses in the space / wave number domain. The analysis relies on data collected from different locations in space. Temperature data from a thermistor string or current components from an Acoustic Doppler Current Profiler (ADCP), both provide data from multiple levels in the water column. Similarly, harmonic analysis could be performed from a series of any sensors aligned (preferably with uniform spacing between the sensors) in the horizontal plane. Examples might include a vertical array of current speed and direction or temperatures (Figure 7–11) or an across-shelf array of uniformly spaced thermistors or current meters (Figure 7–12).

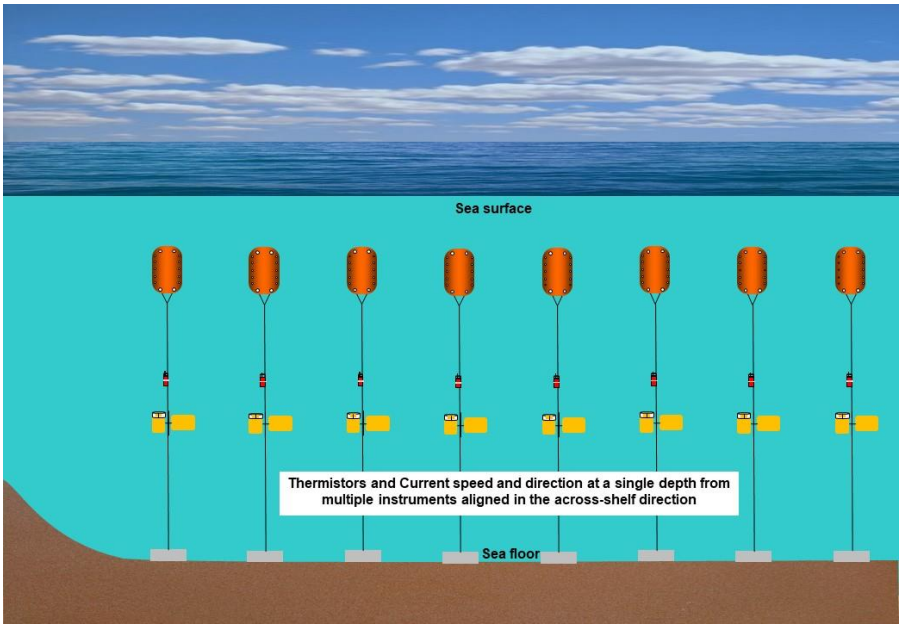


Figure 7–12. Thermistors and Current speed and direction at a single depth from multiple instruments, uniformly spaced, across the continental shelf.

The concept is visualised in Figure 7–13, which shows the fundamental frequency and the first three harmonics. Here, space is assumed to lie in

the vertical direction, as indicated by the mooring configuration presented in Figure 7–11. Let's assume that we're working with currents. The 'fundamental' indicates currents throughout the whole water column are moving in the same direction. The first harmonic may be thought of as a two-layered system, with the currents flowing in one direction in the top layer and in the opposite direction in the bottom layer. The second harmonic is a three-layered system with the top and bottom currents flowing in the same direction and the middle part of the water column flowing in the opposite direction, and so on for higher order harmonics.

The same concept holds for other variables e.g. temperature. In such cases, the concept of 'direction' is relative to some mean value.

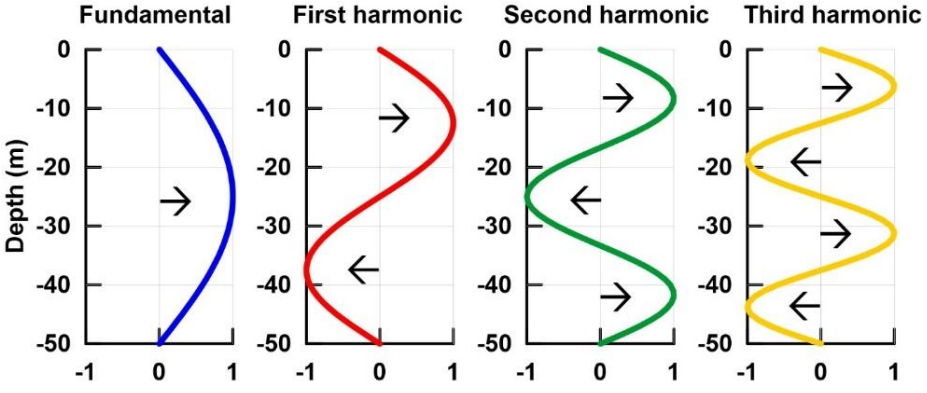


Figure 7–13. Visualisation of the fundamental and the first three harmonics in space. Arrows indicate the direction of flow. The scales are immaterial – the vertical scale representing space and the horizontal scale representing spectral energy.

The mechanics of spatial harmonics is identical to that for spectral analysis. That is, we can exactly fit a set of N data points with N trigonometric functions. Again, let's assume that we're working with currents. Assume we have an ADCP that produces current speed and direction at N levels in the water column. The spatial representation of the data is similar to that using time, with time t being replaced with space, s.

$$x_s = a_o + \sum_{p=1}^{\frac{N}{2}-1} \left[a_p \cos\left(\frac{2\pi ps}{N}\right) + b_p \sin\left(\frac{2\pi ps}{N}\right) \right] + a_{\frac{N}{2}} \cos(\pi s)$$

$$for \quad s = 1, 2, 3, \ldots N$$

Where:

$$a_o = \frac{1}{N} \sum_{s=1}^{N} x_s = \langle x \rangle$$

$$a_{N/2} = \frac{1}{N} \sum_{s=1}^{N} (-1)^s x_s$$

$$a_p = \frac{2}{N} \sum_{s=1}^{N} \left[x_s \cos\left(\frac{2\pi ps}{N}\right) \right] \quad for \quad p = 1, 2, 3, \ldots . \left(\frac{N}{2}\right) - 1$$

$$b_p = \frac{2}{N} \sum_{s=1}^{N} \left[x_s \sin\left(\frac{2\pi ps}{N}\right) \right] \quad for \quad p = 1, 2, 3, \ldots . \left(\frac{N}{2}\right) - 1$$

Unlike the harmonic analysis associated with the tides (above), the spatial harmonic analysis used here contains no error term. This is an exact trigonometric representation of the data, as in spectral analysis.

At each time step, the fundamental is given by a_o, while the amplitude of the harmonic p is given by $R_p = \sqrt{a_p^2 + b_p^2}$. As for spectral analysis, a phase is also generated, although the primary output that is used is the amplitude. The process can then be repeated for all time steps in the data set. Hence, the spectral breakdown of the spatial data (usually in the vertical) can be produced as a function of time. Often, it is more convenient to plots the results in terms of the percentage contribution of each harmonic to the total spectral energy.

7.3.1. Example 7-5: Across shore currents in the vertical

The first example examines the harmonics associated with the vertical spatial representation of the across shore current component. The time frame is one day (on 1 January 2019), with a sampling interval of 5 minutes. This will result in 288 sets (i.e. one every 5 minutes for 24 hours) of spatial harmonics.

Question: What is the relative importance of the fundamental and the first three harmonics?

The processes (data pre-processing, stationarity, smoothing of the periodogram) that we used for the spectral analysis don't really fit well with harmonic analysis, which usually contain a relatively small number of data points. Their interpretation in the spatial domain is more difficult to address than it is in the time domain. They are not presented in this section.

Data used:

- The vertical distribution of the across shore component of the currents.
- The data set is 24 hours long, with data collected every 5 minutes.
- We're interested in the spatial harmonics, so the temporal Nyquist period and frequency are not important.
- Data are spatially separated (in the vertical, by 2 m).
- A contour plot of the across shore current component as a function of depth and time is shown in Figure 7–14. Areas shaded blue represent onshore currents, while those shaded red/yellow/green represent offshore currents.

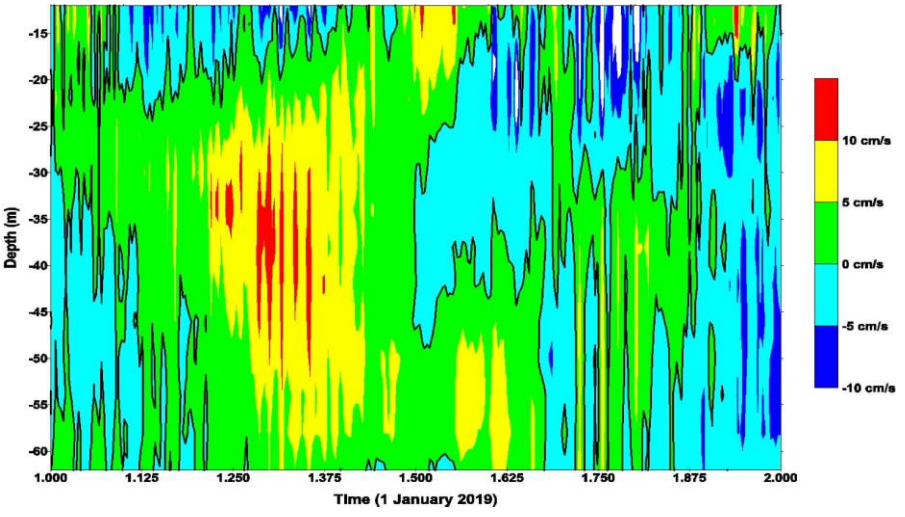


Figure 7–14. Contour plots of the across-shore current component on 1 January 2019 as a function of depth and time. Blue shades indicate onshore currents while red shades indicate offshore currents.

Description:

- Examples of the currents at three times (10:00, 13:00 and 15:00 on 1 January 2019) are shown in Figure 7–15.
- The plot in Figure 7–15 at 10:00 shows the currents throughout the whole water column moving offshore. This is similar to the fundamental frequency shown in Figure 7–13.
- The middle plot in Figure 7–15 (at 13:00) indicates that the currents are flowing offshore in the top and bottom layers and onshore in the middle layer. This is similar to the second harmonic shown in Figure 7–13.
- The final plot in Figure 7–15 (at 15:00) indicates a two-layer system, currents flowing onshore in the top layer and offshore in the bottom layer. This is similar to the first harmonic shown in Figure 7–13.

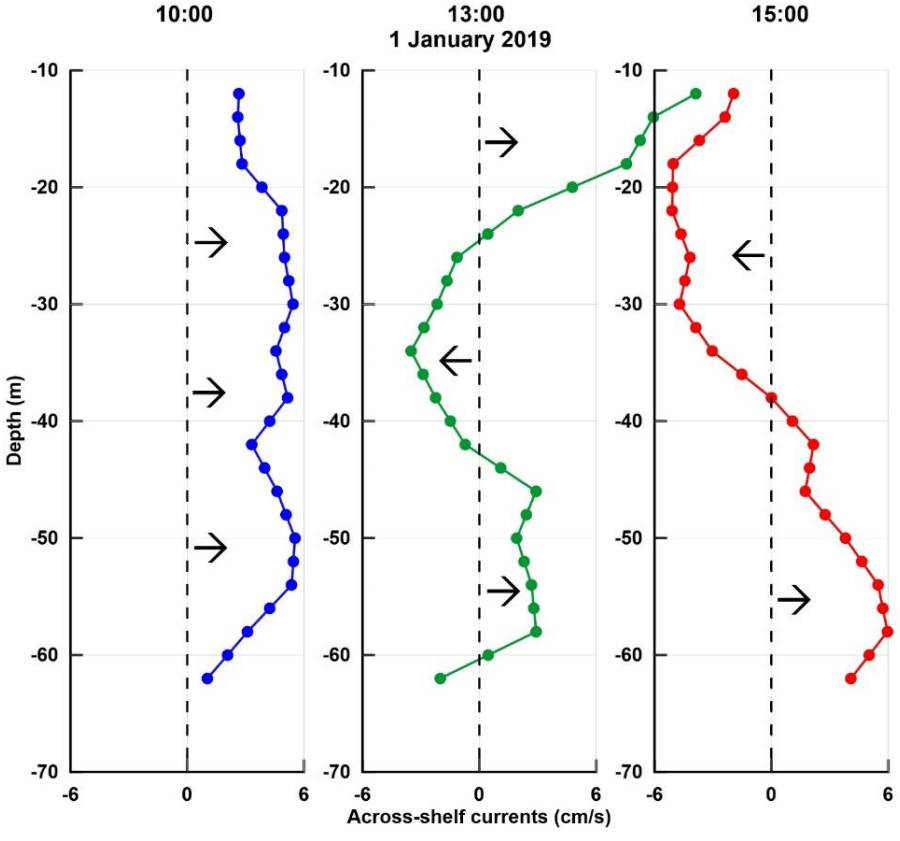


Figure 7–15. Examples of the across-shelf current component at three times on 1 January 2019. Arrows indicate the direction of flow at different levels in the water column (offshore direction arrow towards the right; onshore direction arrow towards the left). The filled circles indicate the locations in the water column from which the currents data were obtained.

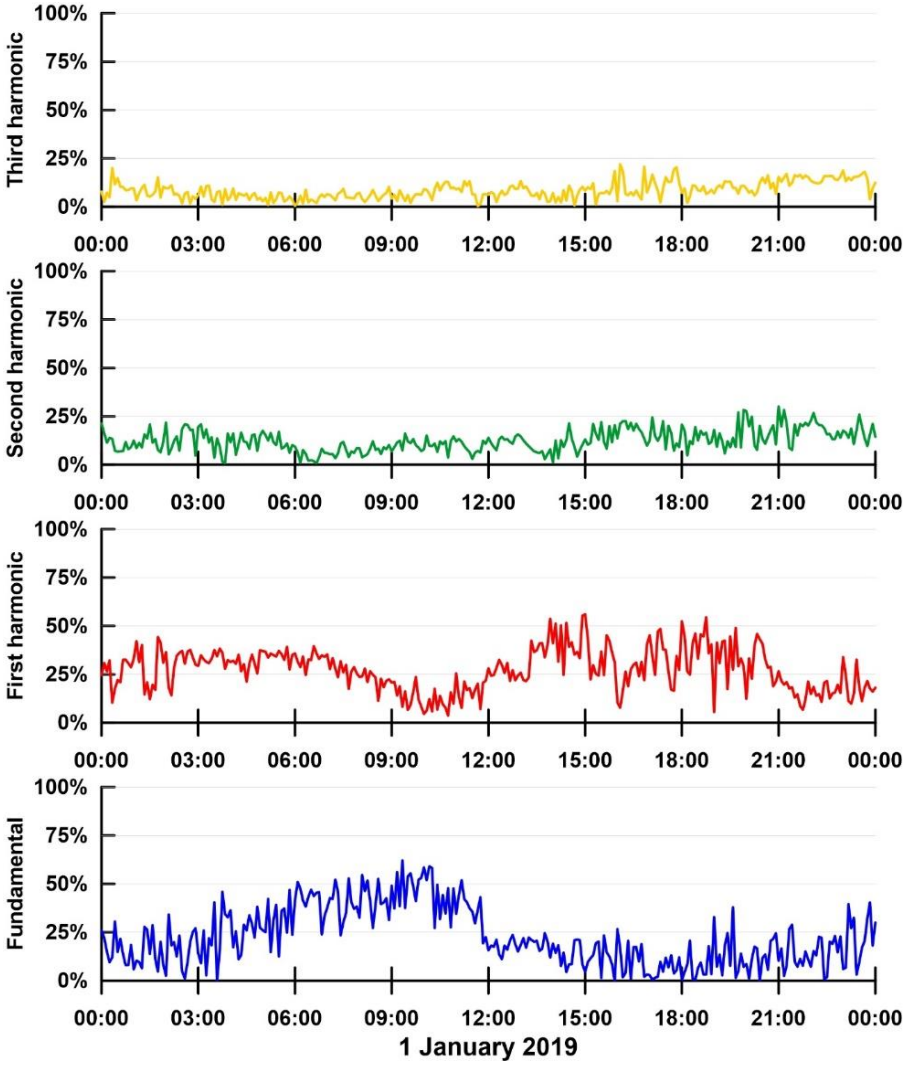


Figure 7–16. Percentage contribution of each harmonic to the total spectral energy at each (5-minute) time step on 1 January 2019.

Return to the question we originally asked.

A time series of the spatial harmonics (the fundamental and the first three harmonics) is shown in Figure 7–16. Each plot represents the percentage contribution of each harmonic (and the fundamental) to the total spectral

energy at each time step. The relative importance of the different harmonics changes over time.

(a) Between about 06:00 and 11:00, the fundamental represents about 50% of the total spectral energy. The three harmonics presented contribute, in total, about 40-50% of the total spectral energy, while the sum of the remaining harmonics contribute <10% to the total spectral energy. Between about 13:00 and 20:00 the first harmonic contains most of the spectral energy (about 30-50%). The second and third harmonics each contain, proportionally, considerably more spectral energy than in the 06:00-11:00 time period. The fact that the first harmonic contains more spectral energy than the fundamental suggests it is a simple mass balance of the across shelf water movement, or the presence of internal waves.

7.3.2. Example 7-6: Water temperature

The second example considers the vertical temperature distribution over time two days: 24 January 2019 and 24 June 2019 (Figure 7–17, bottom and top plots respectively). They represent examples from summer and winter seasons, respectively.

Question: How do the spatial (in the vertical) harmonics of water temperature differ between summer and winter?

Again, with relatively few data points, the processes (data pre-processing, stationarity, smoothing of the periodogram) that we used for the spectral analysis don't really fit well with harmonic analysis. Their interpretation in the spatial domain is more difficult to address than it is in the time domain. They are not presented in this section.

Data used:

- Two sets of 5-minute water temperature data, each taken over 24 hours. One data set is from summer and the other from winter.

- At each time, water temperature data are collected from 14 depths in the water column. The vertical spatial separation of the data is 4 m.

Description:

- Data from 24 June show very little thermal structure throughout the water – the system is (approximately) isothermal.
- Data from 24 January show considerable thermal stratification in the water column, together with, what appears to be, considerable internal wave activity (between 00:00 and 04:00 and again between 12:00 and 18:00.
- Results from the harmonic analysis of the temperature data are presented in Figure 7–18 – results from 24 January on the left hand side of the figure and those from 24 June on the right hand side of the figure. Results are presented as the percentage of the total spectral energy from each of the fundamental and the first three harmonics. Note that the vertical scale for the fundamental differs from those for the harmonics.
- On 24 January 2019, about 70% of the total energy is contained in the fundamental, about 10% in the first harmonic, 5% in the second harmonic and slightly less than 5% in the third harmonic. (The remaining spectral energy is contained in the higher order harmonics).
- During the periods with high internal wave activity, spectral energy in the first harmonic is slightly greater than at other times.
- In contrast, results from 24 June 2019 indicate almost 100% of the energy resides in the fundamental and virtually nothing in any of the harmonics.
- High energy in the fundamental indicates the direction of thermal change is the same throughout the entire water column.
- High energy in the harmonics indicates that the direction of thermal change is different at different locations in the water column.

Returning to the question asked.

(a) In winter, when there is little thermal variability throughout the water column (e.g. 24 June 2019, Figure 7–18, top), almost all of the energy is contained in the fundamental. In summer, there is substantial thermal variability throughout the water column (e.g. 24 January 2019, Figure 7–18, bottom) and the energy in the fundamental is considerably reduced, compared with the results from winter. Variability in the higher order harmonics increases. During periods of high internal wave activity, the relative importance of the first harmonic increases.

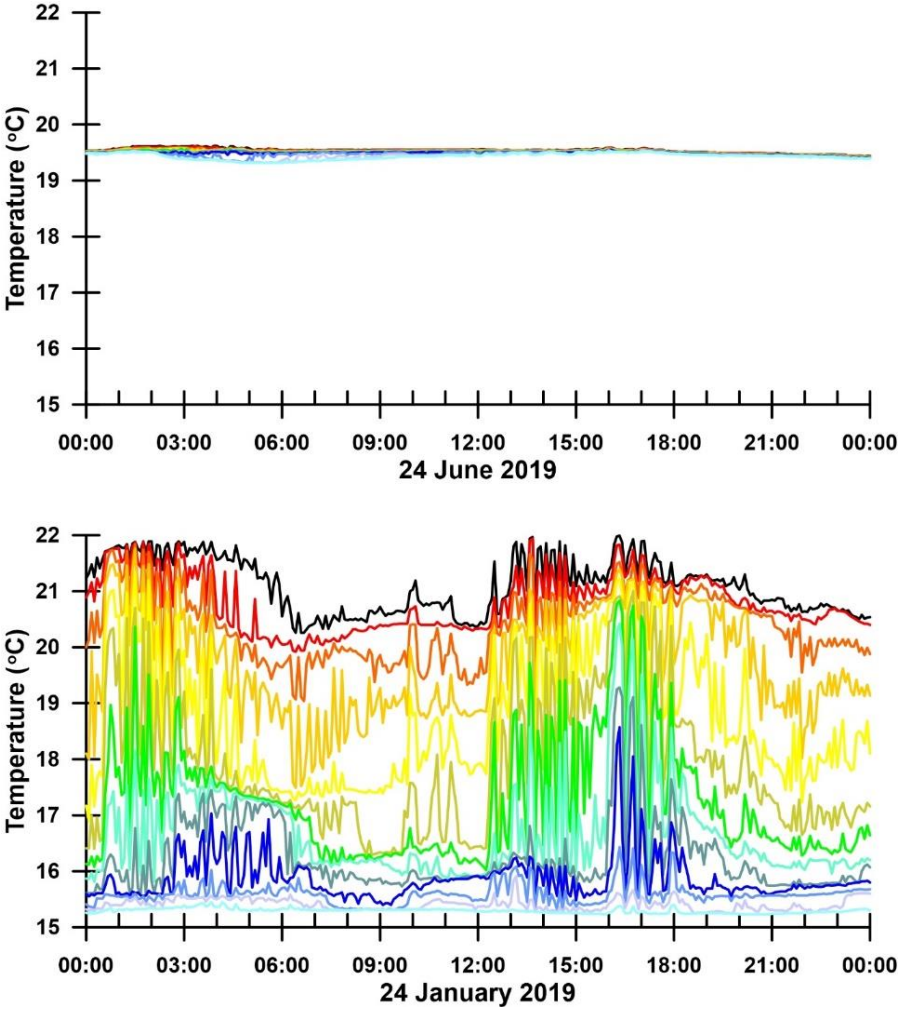


Figure 7–17. Time series plots of water temperature every 4 m throughout the water column on 24 June 2019 (top) and 24 January 2019 (bottom). Data are from a location in the Southern Hemisphere.

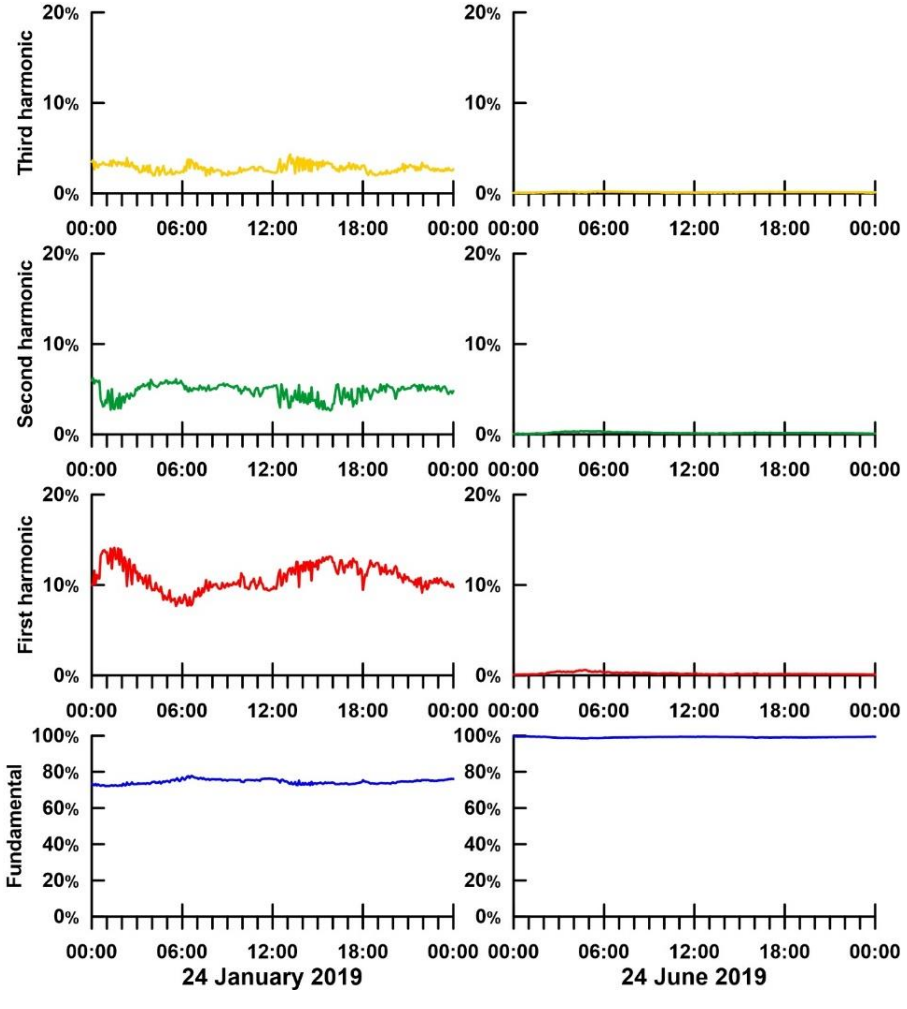


Figure 7–18. Percentage contribution of each harmonic to the total spectral energy at each (5-minute) time step on 24 January 2019 and 24 June 2019. The vertical scale for the fundamental differs from that of the harmonics.

7.4. Summary

The fundamental difference between spectral and harmonic analysis is that the former provides an exact fit to the data but only allows certain frequencies, while the latter is only an approximate fit to the data but at specific (user defined) frequencies. Other than that, they are essentially the same operation.

The equation representing the time series is solved (approximately) using a selected number, M, of desired frequencies. The unwanted information is combined into a 'residuals' term that is minimised in a least squares sense.

$$x_t = A_0 + \sum_{q=1}^{M} \left[A_q \, cos(\omega_q \, t) + \, B_q \, sin(\omega_q \, t) \right] + \, residuals$$

The advantage of harmonic analysis over spectral analysis is that it can target very specific frequencies. So, if our data set contains regularly occurring features in which we're interested (the tides are a good example), then we can focus on the frequencies associated with those features. It is for this reason that harmonic analysis is a popular tool for coastal engineers and oceanographers.

Chapter 8. Short Time Fourier Analysis

The final part of this work concerns wavelet analysis, associated with time series data. Wavelet analysis is a relatively new subject, gaining popular momentum as an analysis tool from about the 1990s onwards.

Many types of wavelets exist and all have features that are advantageous for different types of problems. For example, Haar wavelets are 'good for edge detection and reconstructing binary pulses' (Fugal, 2009). Examples of the more comment types of wavelets include: Haar, Daubechies, derivative of Gaussian (DOG, which includes the Mexican hat wavelet), Morlet, Paul and many others. Here, we will focus on the complex Morlet wavelet for several reasons, including:

- It is complex (hence amplitude and phase information can be produced).
- It smoothly tapers to zero at it ends.
- The wavelet has no discontinuities or sharp edges in either the time or frequency domains.
- It has a structure that is similar to that of the data sets often encountered in the physical sciences.

Wavelets can be classified as continuous or discrete (in much the same way as the Fourier transform). Continuous wavelets are more generally applied to the processing of time series data, while discrete wavelets are more generally applied to image processing. The focus here will be on continuous wavelet analysis. However, what we actually implement is a discretised approximation to the continuous form of the wavelet equation. Like Fourier transforms, wavelet transforms may have different forms, although the basic function will always have a similar structure. Therefore, it is important to specify the definition used for both the Fourier and wavelet transform.

Much of the early work on wavelets we've found overly complex and confusing. Polikar (2006) states 'wavelet transforms … are written my math people, for the math people, in a language which even most of the

math people themselves cannot understand…'. While there is a little bit of mathematics used here, we've tried to avoid terms like: kernels, basis functions, orthonormal sets etc. While these are critical to, and lie behind, wavelet analysis, we do not think it is necessary to know of these terms in detail prior to applying wavelet analysis. We leave this the pure mathematicians.

There are multiple approaches to explaining wavelet analysis. One of the better tools that we found to help our understanding of wavelet analysis is 'The wavelet tutorial' (Polikar, 2006). Perhaps this is because Polikar starts with Fourier analysis as a basis for understanding the origins of wavelet analysis, explaining the inherent problem with Fourier (or spectral) analysis and showing how this problem can be largely overcome using wavelet analysis.

We'll start here in much the same way as Polikar (2006) and develop our understanding from there.

There are also many 'You-tube' wavelet videos that are useful. We find those presented by Michael X. Cohen particularly informative. Much of Cohen's videos focus on neurological sciences, but we find the work presented is clear and provides many insights to wavelet analysis.

Two further references provided good practical information: Torrence and Compo (1998) and Grinsted et al (2004). Both of these references contain directions to Matlab scripts for undertaking auto wavelet analysis and cross wavelet analysis, respectively. The Matlab scripts are relatively easy to implement, although like all computer programs, they contain a few idiosyncrasies. While Excel contains Fourier analysis as part of its statistics package, we are not aware of a wavelet package in Excel. (One of the authors has developed an Excel macro for wavelet analysis. However, at best, that macro is very time consuming to implement, specific to each problem and cumbersome to use. Using the Matlab wavelet analysis tools is far more efficient).

8.1. Limitations of spectral analysis

There are two major limitations to spectral (or Fourier) analysis. They are: (a) stationarity, and (b) the trade-off between time and frequency.

8.1.1. Stationarity

Recurring events may not always appear regularly in a data set. The amplitudes of recurring events may not always be the same (e.g. the maximum temperature each year will likely differ) and the frequency of recurring events may not be consistent (e.g. maximum temperatures, in the Southern Hemisphere, may occur in December one season, in February the next season and in January the following season). The time series is not stationary, even for this dominant signal. The amplitude or frequency of a recurring event may not be consistent over time.

One of the fundamental assumptions in spectral analysis is that the time series is 'stationary', usually approximated by 'weak stationarity'. However, even weak stationarity may not always exist. The best that we can really hope for is that the mean and variance of segments of the time series are not statistically different or at least these statistics 'don't vary too much'. Often, we simply remove the linear trend from the time series and hope that the deviations from the de-trended data provide a sufficiently good approximation to stationarity.

That approach works reasonably well for many applications. However, there are also many other applications for which this approach is sub-optimal. Some examples include: a packet of internal waves that occur only during certain times in a time series (Figure 8–1), time series that span many years or decades (the Southern Oscillation Index, an indicator of El Nino, is a common example, Figure 8–2). The problem with spectral (or Fourier) analysis is that is assumes that the same signal is always present in the time series. More than that, spectral analysis actually forces the best fitting set of trigonometric functions to apply to the whole time series (noting that this set of trigonometric functions is actually an exact fit to the data points).

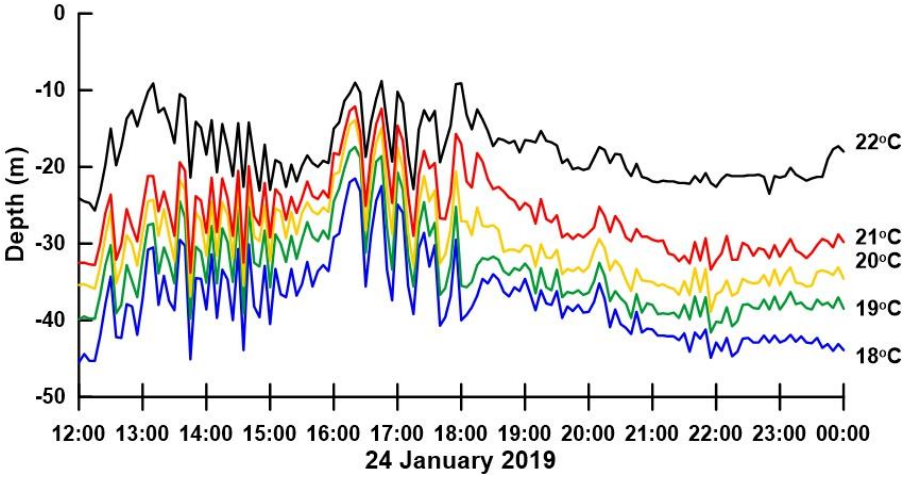


Figure 8–1. Isotherm displacement during the afternoon of 24 January 2019. Large vertical isotherm displacements are observed in the first half of the time series, while only small such displacements are observed in the second half of the time series.

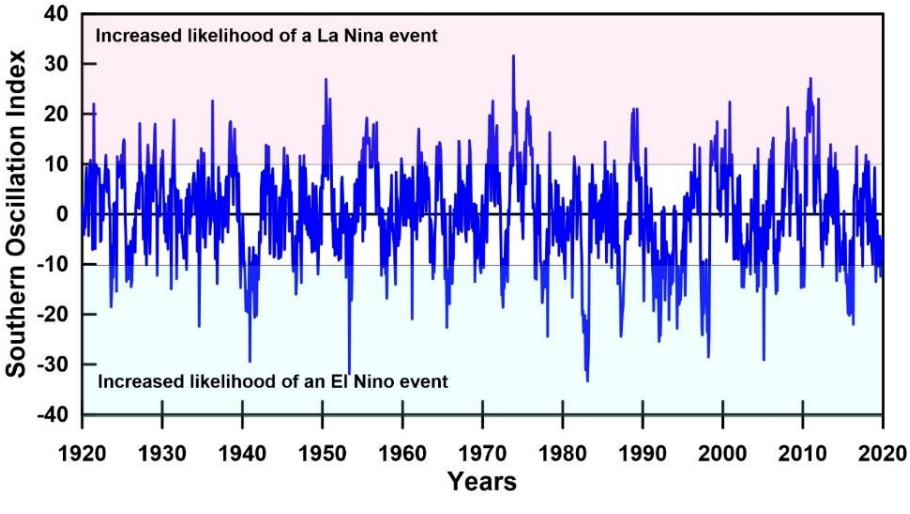


Figure 8–2. The monthly Southern Oscillation Index over the last 100 years.

8.1.2. Time and frequency resolution

A time series allows us to easily identify when an 'event' (e.g. the maximum water temperature) occurs in the data set, Figure 8–3. However, a time series contains no information on the frequency of this event (i.e. how regularly the event repeats itself).

Spectral analysis identifies how frequently certain events are and the relative importance of that type of event in relation to other information contained in the data set, Figure 8–3. However, it contains no information on when the events occur. This single spectrum applies to all times.

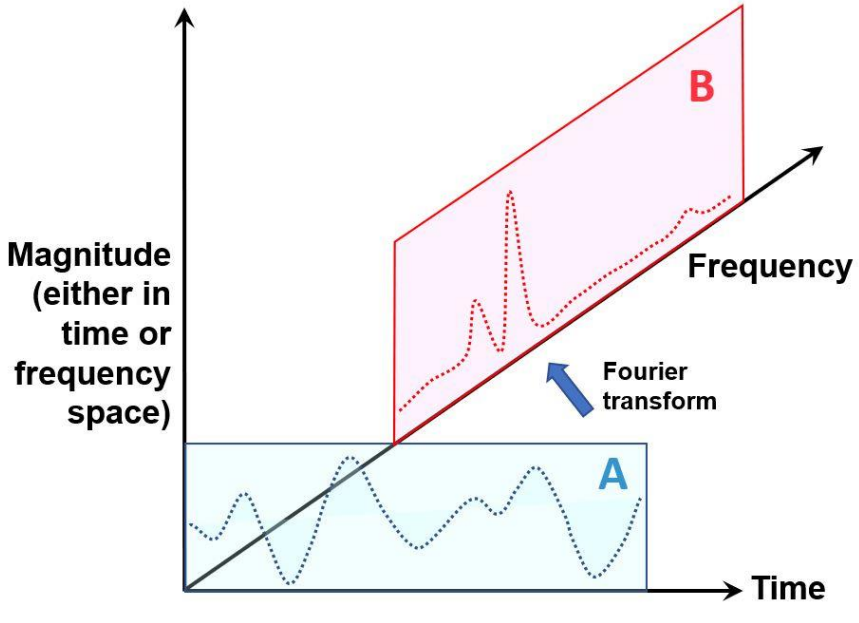


Figure 8–3. View of information as a time series (A) and after a spectral analysis (B). The Fourier transform moves us from the time domain to the frequency domain.

Ideally, we'd like to identify a transform that can cope with changing amplitudes and frequencies and also retains resolution in both time and frequency – we'd like to have information on both time and frequency. Obviously, the wavelet transform is such a function.

However, before we discuss wavelets, we'll show a few more examples where Fourier analysis breaks down and one early attempt to solve that problem.

8.1.3. Example 8-1 – A chirp

The first example examines a 'chirp' signal – one in which the frequency continually changes. (Think of how the sound of a siren changes when an emergency vehicle passes us). The chirp is compared with a sine curve. The amplitudes of both signals are the same and do not change over time. Mathematically, a chirp can be represented as:

$$x(t) = sin\left[2\pi\left(\frac{f_1 - f_0}{T}\frac{t^2}{2} + f_o t\right)\right]$$

Where:

t is time,

f_1 and f_0 are the frequencies at the end and start of the time series, respectively, and

T is the total duration of the signal.

We'll compare this with a sine curve with a frequency that is the average of the start and end frequencies above. Viz.

$$y(t) = sin\left[2\pi\left(\frac{f_1 + f_0}{2}\ t\right)\right]$$

Questions: (a) What does the spectrum of the chirp look like? (b) How does it differ from the spectrum of a sine curve?

Data used:

- Both the chirp and the sine curves are mathematical constructs. Respective data are generated from their mathematical expressions.
- Time series of the chirp $x(t)$ and the sine $y(t)$ curves are plotted in Figure 8–4 ($\Delta t = 0.01\ sec$). Both data sets have six peaks / troughs. The peaks for the sine curve are uniformly spaced at

exactly 1.7 s apart, while the peaks for the chirp become closer together as time progresses.

Description:

The corresponding periodograms are given in Figure 8–5.

- The spectrum of the sine curve shows a single, dominant peak at a frequency of about 0.6 cycles per sec (i.e. a period of about 1.7 sec).
- The spectrum of the chirp shows a very broad 'peak' between, maybe, 0.3 and 1 cycles per sec.
- The total spectral energy of the two data sets is the same.

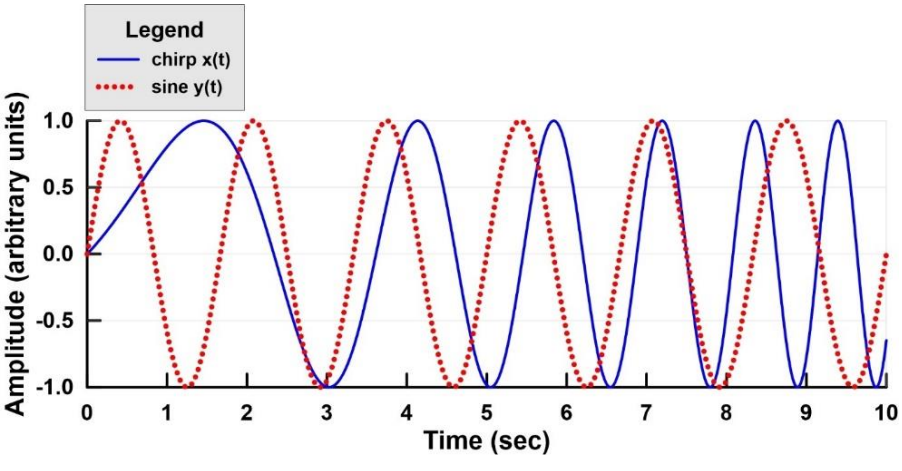


Figure 8–4. Time series of the chirp (blue curve) and sine (red curve) signals.

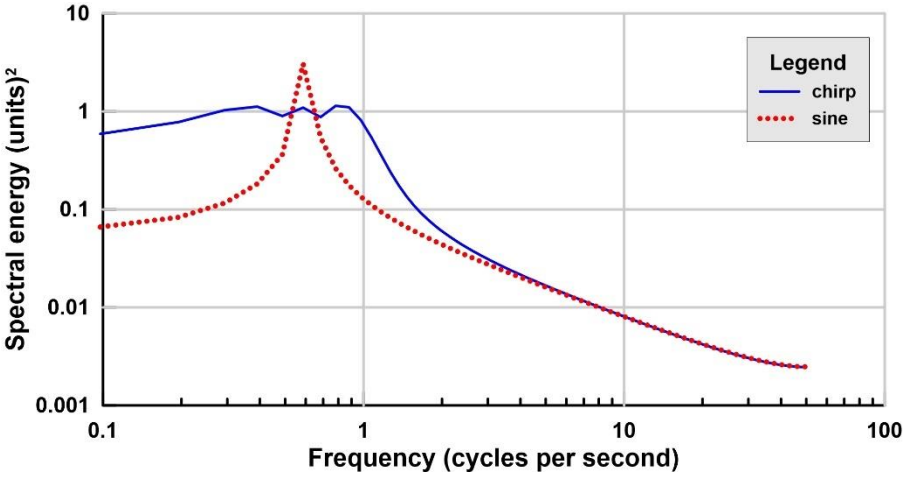


Figure 8–5. Spectral forms of the chirp (blue line) and sine (red line) signals.

Returning to the question.

(a) The chirp signal contains a frequency that is always changing. The spectrum cannot find a single, dominant frequency. The best it can do is to find a range of 'important' frequencies and spreads the energy over these frequencies. The result is a broad, flat (and probably not significant) peak in the lower frequency end of the spectrum.

(b) For the sine curve, spectral energy is focused on a frequency of 0.6 cycles per sec. This frequency is always present in the data set and does not change over time. It results in a relatively sharp spectral peak, in contrast to the very broad region of high spectral energy observed in the spectrum of the chirp.

8.1.4. Example 8-2 – Sine curve variation 1

This is a slight modification of the chirp example (the time series is shown in Figure 8–6 and the corresponding spectral analyses in Figure 8–7). The first plot (a), $x(t)$, comprises a sine curve with frequency $\omega_1 = 0.8$ cycles per second for its first half and a sine curve with frequency $\omega_2 = 1.2$

cycles per second for the second half of the data set. The second plot (b), $y(t)$, comprises a single sine curve with frequency 1.0 cycles per second. That is:

$$x(t) = \sin(2\pi\,\omega_1 t) \qquad for\ t = 1\ to\ 512$$

$$x(t) = \sin(2\pi\,\omega_2 t) \qquad for\ t = 513\ to\ 1024$$

$$y(t) = \sin\left(2\pi\,\frac{\omega_1 + \omega_2}{2}t\right) \qquad for\ t = 1\ to\ 1024$$

Question: What is the difference between the spectra of a sine curve with a single frequency and one in which the frequency suddenly changes half way through the data set?

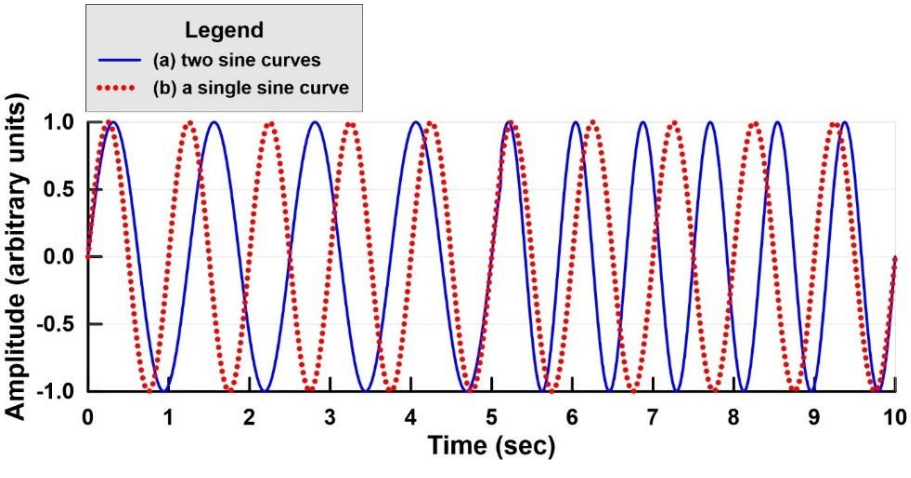


Figure 8–6. Two time series: (a) two separate frequencies and (b) a single frequency.

Data used:

- Both the sine curves are mathematical constructs. Respective data are generated from their mathematical expressions (above).
- Time series of the two sine curves, $x(t)$ and $y(t)$ curves are plotted in Figure 8–6 ($\Delta t = 0.01\ sec$).

- Close examination of Figure 8–6 shows, for the single sine curve (the red, dotted line), 10 peaks, uniformly spaced in time. For the 'two sine curves' (the blue solid line), examination shows 4 peaks in the first half of the data set and 6 peaks in the second half of the data set.

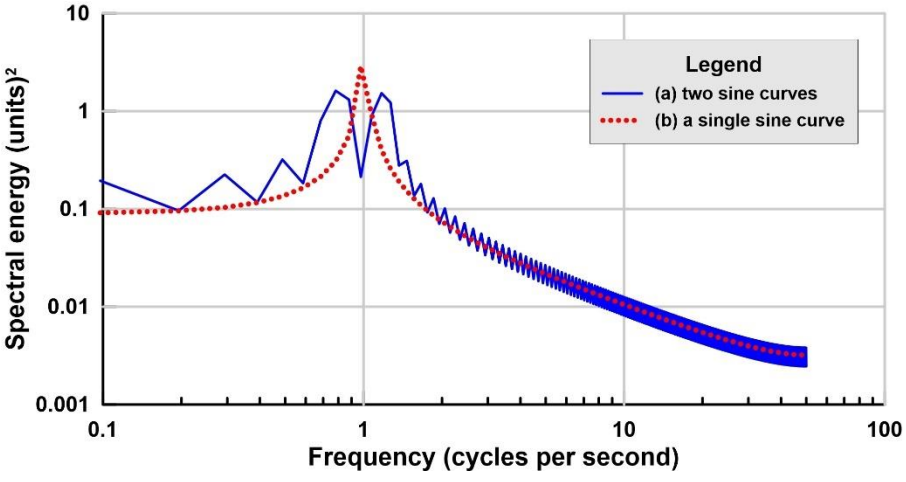


Figure 8–7. Spectral forms of the time series for (a) a sine curve with two separate frequencies and (b) a sine curve with a single frequency.

Description:

- The spectral plot for the single sine curve (Figure 8–7, the red dotted line) shows a single, sharp spectral peak at 1 cycle per second.
- The spectral plot for the 'two sine curves' (Figure 8–7, the blue solid line), shows two spectral peaks – at 0.83 and 1.25 cycles per second. However, these two spectral peaks are smaller and broader than the single sine curve.
- Further, the 'two sine curves' is considerably noisier than the single sine curve at the high frequency end of the spectrum.
- The spectral analysis tries to fit each sine curve across the whole data set, resulting in a 'smearing' of the features and a reduction in the size of the spectral peaks.

Returning to our question:

(a) The spectral plot for the single sine curve shows a single, sharp spectral peak at 1 cycle per second, while the spectral plot for the second spectrum shows two broad spectral peaks at 0.83 and 1.25 cycles per second. There is considerably more noise in the second spectrum, particularly at higher frequencies.

8.1.5. Example 8-3 – Sine curve variation 2

This is a variation on the previous example in which we have a sine curve, $x(t)$, with one frequency in the first half of the data set and a different frequency in the second half of the data set, Figure 8–8 (a), and a second time series, $y(t)$, comprising two superimposed sine curves with both the above frequencies, Figure 8–8 (b). That is:

$$x(t) = a \ \sin(2\pi\omega_1 t) \qquad for\ t = 1\ to\ 512$$

$$x(t) = a \ \sin(2\pi\omega_2 t) \qquad for\ t = 513\ to\ 1024$$

$$y(t) = b \ \sin(2\pi\omega_1 t) + b \ \sin(2\pi\omega_2 t) \qquad for\ t = 1\ to\ 1024$$

The amplitude of the second time series, $y(t)$, was scaled to ensure that the total energy for both data sets was the same ($b = 0.7\ a$).

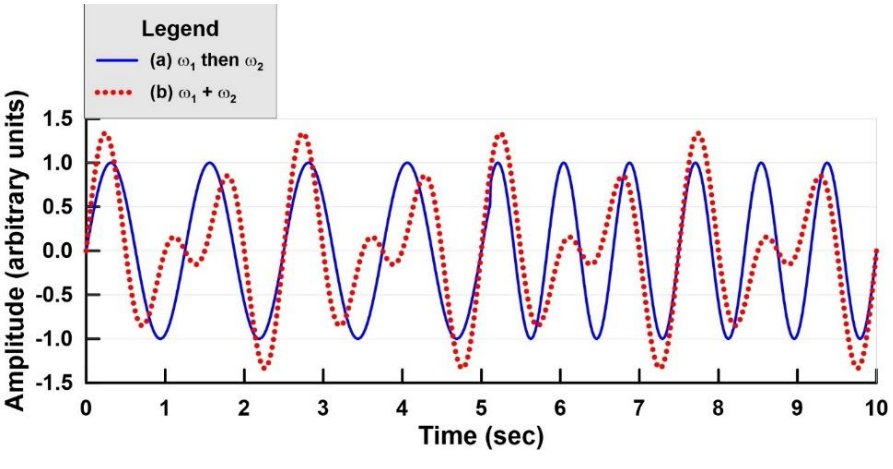


Figure 8–8. Two sine time series: (a) two separate frequencies and (b) two combined frequencies.

Question: What is the difference between the spectra of a sine curve in which the frequency suddenly changes half way through the data set and the sum of these two sine curves but with a different amplitude?

Data used:

- Both the sine curves are mathematical constructs. Respective data are generated from their mathematical expressions (above).
- Time series of the two sine curves, $x(t)$ and $y(t)$ curves are plotted in Figure 8–8 ($\Delta t = 0.01\ sec$).
- Close examination of Figure 8–8 shows, for the first function, $x(t)$, (the blue solid line), examination shows 4 peaks in the first half of the data set and 6 peaks in the second half of the data set. The second function, $y(t)$, contains a total of 12 peaks in the data, eight of which are larger peaks and four of which are smaller peaks.

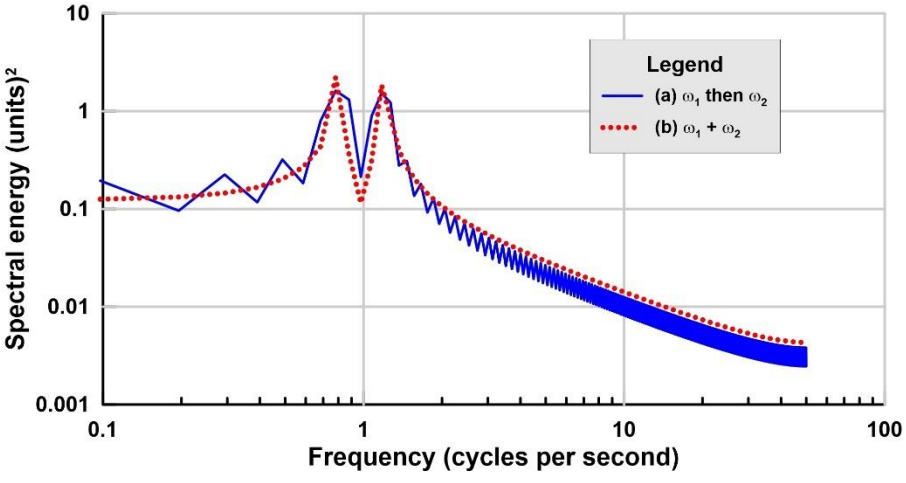


Figure 8–9. Spectral forms of the two sine time series for (a) a curve with two separate frequencies and (b) a curve with two combined frequencies.

Description:

- From the spectral plots (Figure 8–9), we find a similar result to the previous example.
- Using the data set with the combined frequencies, there are two, relatively sharp spectral peaks at frequencies of 0.83 and 1.25

cycles per second, which correspond to the two frequencies in the combined data set.

- The result from the two sequential frequencies shows two broad, slightly smaller spectral peaks at frequencies of 0.83 and 1.25 cycles per second and a noisy tail at the high frequency end of the spectrum.

Returning to our question:

(a) The spectral plot for $x(t)$ (Figure 8–9, blue curve) shows two, broad spectral peaks at 0.83 and 1.25 cycles per second. This spectrum also shows considerable noise, particularly at higher frequencies. In contrast, the spectrum of $y(t)$ (Figure 8–9, red curve) shows two sharp peaks at these same frequencies, with virtually no noise at the high frequency end of the spectrum. The spectral analysis of $x(t)$ tries to fit each sine curve across the whole data set, resulting in a 'smearing' of the features and a reduction in the size of the spectral peaks.

8.1.6. Example 8-4 – An internal wave

This example examines the effect of a real, non-stationary time series on the spectrum associated with a temperature record that includes a period of time during which internal waves are present. The time series used is that of the yellow curve in Figure 8–1.

Question: How does the spectrum change when internal waves are included in, and excluded from, the analysis?

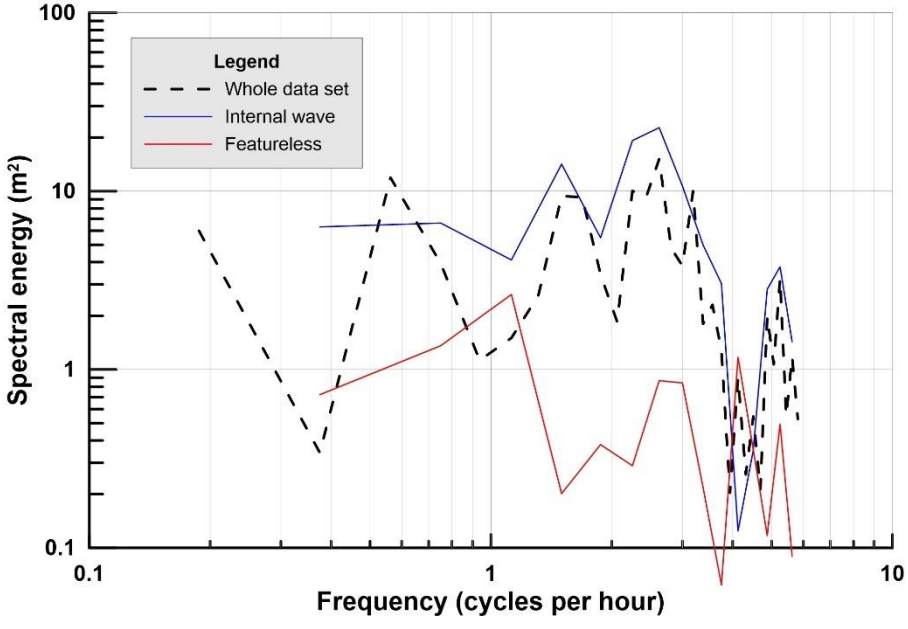


Figure 8–10. Spectral analysis of the whole time series, the first half of the time series (with large internal wave activity) and the second half of the time series (with little internal wave activity).

Data used:

- Input data is the displacement of the 20°C isotherm (the time series is shown in Figure 8–1, yellow curve).
- All data are taken at 5 minute intervals.
- Spectral analysis is undertaken using the whole data (64 records long), the first half of the data (32 records long) where there are large displacements of the depth of the 20°C isotherm (i.e. high internal wave activity) and the second half of the data (32 records long) where there is little internal wave activity.

Note: because the size of the data sets used here are small, we have not presented the results from our preliminary testing, nor have we undertaken any smoothing of the periodograms, hence confidence limits have not been determined. This does not change the basic message that this example is designed to highlight.

Description:

- Maximum spectral energy occurs during the period of high internal wave activity. There may be a small, broad spectral peak at about 2-3 cycles per hour. (The broadness of the peak suggests that the frequency of the internal wave is not constant).
- The period with high internal wave activity contains spectral energy that is about one order of magnitude greater than the spectrum that contains little internal wave activity.
- The energy of the spectrum obtained using the whole data set lies between these two extremes.

Traditional spectral analysis does not allow us to discriminate between the two halves of the time series (i.e. the time periods with and without the internal wave) in the data set. Spectral analysis 'gets confused', trying to fit a single set of sine / cosine curves to the whole data, when clearly, there are different processes (or signals) occurring in the first and second halves of the time series.

Returning to our question:

(b) Spectral analysis using the whole data set results in an underestimate of the period containing the internal waves (the first half of the data set) and an overestimate of the energy in the second half of the data set. (Clearly, if we're only interested in the internal wave signal, then we'd only analyse the first half of that data set i.e. those times during which internal waves are clearly present).

8.1.7. Example 8-5 – The Southern Oscillation Index

The Southern Oscillation Index (SOI) is a surrogate for El Nino events – the cycle of warming and cooling of eastern Pacific Ocean waters. It is essentially the difference in the average atmospheric pressure between Darwin and Tahiti. If the SOI remains strongly positive (>6) for several months or longer, there is an increased likelihood of there being an El Nino event (drier than normal conditions along the east coast of Australia). Conversely, if the SOI remains strongly negative (<-6) for several months or longer, there is an increased likelihood of a La Nina event (wetter than normal conditions along the Australian east coast).

The SOI causes a problem for spectral analysis because the El Nino / La Nina cycle of events does not occur at regular intervals. One event may occur two years after the previous event, while the next event may not occur for, say, five years. (An analogy occurs with temperature: one year we might find the lowest Southern Hemisphere temperatures in July, the next year it is in June and the following year it is in August). That is, El Nino events are not uniformly spaced in time, nor are their amplitudes the same for each event. Spectral analysis tries to 'optimise' the period between events – it forces the period to a single value that is, in some sense, the average period between El Nino events.

Question: How does the spectrum of the SOI behave: it is a very long time series (circa 85 years) which contains irregular signals?

Data used:

- Monthly values of the SOI.
- Length of the data set is about 85 years (1,024 data points).

Description:

- Spectral analysis of the Southern Oscillation Index (SOI) is presented in Figure 8–11, based on the time series data in Figure 8–2.
- The spectrum is largely featureless.
- There are no statistically significant spectral peaks and the slope of the spectrum does not follow the $-5/3^{rds}$ power law.

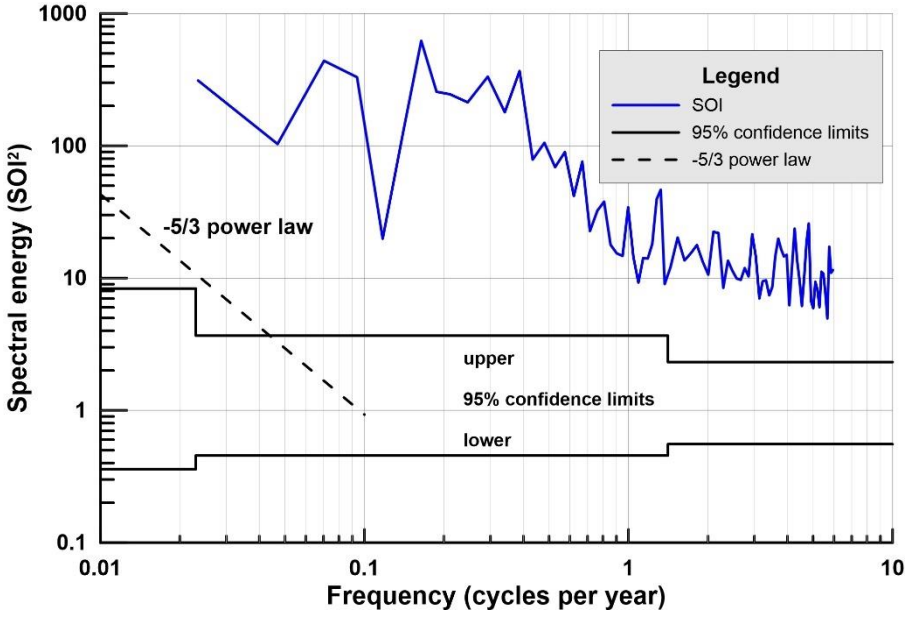


Figure 8–11. Spectral analysis of the Southern Oscillation Index (SOI) using 85 years of monthly SOI data.

Returning to our questions:

(a) From this spectrum, we would conclude that there is nothing of note in the SOI signal. However, we would also note that, because the spectrum does not follow the -5/3^{rds} power law, there may be some hidden information in the data set that spectral analysis cannot identify. What we'll eventually find is that there are significant peaks in the (wavelet) spectrum, we just can't resolve them using the traditional spectral techniques.

8.1.8. Summary

So, the problem (or limitation) we have with spectral analysis centres around the lack of stationarity in the time series. That is: frequencies (or amplitudes) do not remain constant throughout the whole data set. Spectral analysis assumes that the frequency (and amplitude) of signals in the data set remain constant throughout the whole data set – stationary

(at least approximately so). As previously noted, in dealing with a practical data set, stationarity is unlikely to occur.

The example using the data in Figure 8–8 (the blue line) clearly demonstrates this. The frequency ω_1 does NOT occur for all times – it only occurs in the first half of the time series. Similarly, the frequency ω_2 only occurs in the second half of the time series. Neither frequency occurs over the whole data set.

Spectral analysis tells us whether a particular frequency exists and tells us how important is the signal at that frequency (compared with other signals in the data set at different frequencies). However, it does not tell us when that frequency occurs.

This was noted many years ago and one attempt to overcome (or at least mitigate) this problem was to use a 'short time Fourier transform' (STFT). Some authors use the expression 'short term Fourier transform'.

8.2. Short time Fourier transform

8.2.1. The time – frequency resolution trade-off

Ideally, what we'd like to produce is a spectrum that contains frequency information for every point in our time series. The time / frequency information shown in Figure 8–3 would then look more like that shown in Figure 8–12. We obtain frequency information at each time, B_i, in the data set (for clarity, the spectra for five times, B_1 to B_5, only are shown).

Our objective is to obtain frequency information for each time point in the data set. In other words, undertake a spectral analysis at each point in time. (For the remainder of this chapter, we assume that the spectral analysis is undertaken using a FFT).

So, how might this be achieved? Let's assume we have an hourly data set that is 1,024 hours long. We could take 128 FFTs each of 8 hours duration.

This process is gives us a set of 'short time Fourier transforms' (STFT). We are faced with two 'problems'.

(a) The smaller that each of our STFTs is, the larger the number of FFTs that is required. This is not a major problem *per se*, computers can rapidly process many FFTs.
(b) Each time we reduce the number of data points in the FFT, we lose low frequency information. There is much less low frequency information in a data set that is 8 hours long compared with a data set that is 1,024 hours long.

We could broaden this approach in the following way.

- Undertake a FFT of 1,024 hours and the resulting spectrum is 'centred' on time 512 in the data set. Only one such FFT can be produced.
- Split the data set into two halves, then undertake two FFTs each of length 512 hours. The first FFT is centred on time 256 in the data set and the second FFT is centred on time 768 (256+512) in the data set.
- Repeat this process using four data sets, each 256 hours; eight data sets, each 128 hours long and so on, eventually reaching 512 FFTs each of length 2 hours. This does mean undertaking 1,023 FFTs!

This process provides a set of 'short time Fourier transforms' (STFT). Each time we reduce the number of data points in the FFT, the resulting spectra become more localised in time i.e. we improve the temporal resolution. However, each time we do this, we lose low frequency information. So, our ideal representation of the time / frequency information (as per Figure 8–12) will actually look more like that shown in Figure 8–13. As we use smaller and smaller time-based data sets (i.e. improve the temporal resolution), the frequency information contracts to the high frequency end of the spectrum (i.e. we reduce the frequency resolution).

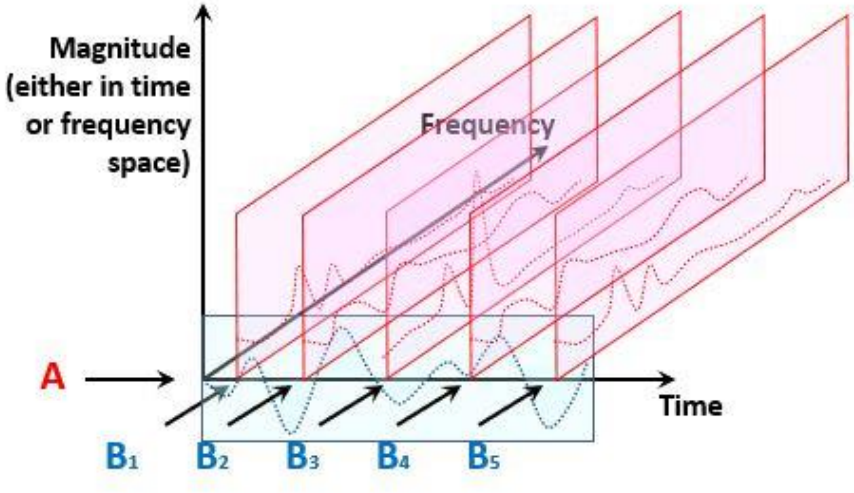


Figure 8–12. At each time B_i in the data set, we'd like to produce a complete spectrum.

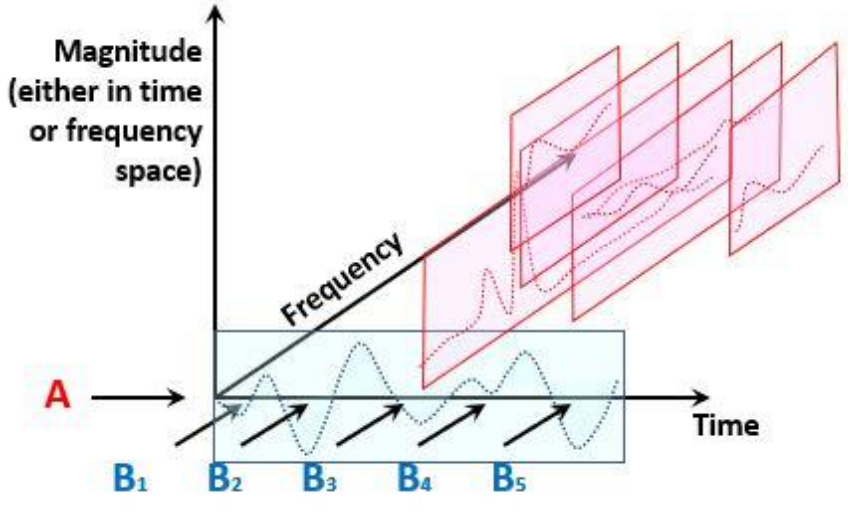


Figure 8–13. View of information: when we reduce the size of the FFT, we lose low frequency information.

The type of representation often used to describe this is shown in Figure 8–14. Assume we have 1,024 data points in our time series. Using all of the data points in the time series, we obtain one FFT, centred in the middle of the time series (the bottom, blue rectangle). In this FFT there is no temporal resolution, the FFT applies to all times in the data set. As we reduce the size of the FFT, we get better temporal resolution (i.e. the distance between the ticks along the time axis becomes smaller) but lose frequency information (i.e. the distance between ticks along the frequency axis become farther apart).

We know that the frequency decreases as the number of data points increases (Chapter 2) i.e.

$$\omega_p = \frac{2\,\pi\,p}{N\,\Delta t} \qquad for \qquad p = 1, 2, 3, \ldots . \left(\frac{N}{2}\right) - 1$$

So, as we improve the temporal resolution, we decrease N, the number of data points in the FFT. By decreasing N, the frequency ω_p increases. Therefore, we lose low frequency information. There is a trade-off – either having good resolution in time or having good resolution in frequency.

Figure 8–14 also provides another piece of information important to the interpretation of results. Again, assume we have a time series that is 1,024 hours long. If we undertake a FFT, the result is centred on time 512 (the middle of the time series). Let's assume that we undertake a FFT using 1,024 data points but we want it centred on hour 256. This would mean that the FFT is undertaken using data from hours -256 to +768. There is no data when time is less than zero i.e. all 'data' points for $t < 0$ are 'zero'. So, this FFT is undertaken using 256 data points that are zero and 768 data points that are, in general, non-zero. Clearly the results from the FFT are not entirely valid, or at the very least they will be biased.

Applying this process to FFTs of other sizes, there will be a region on the time / frequency plot in which the FFT will be impacted by the 'zeroes' that lie outside the time series i.e. the boundaries affect the results. The impacted region is defined by a 'cone of influence' and is shown as the shaded area of the plot in Figure 8–15.

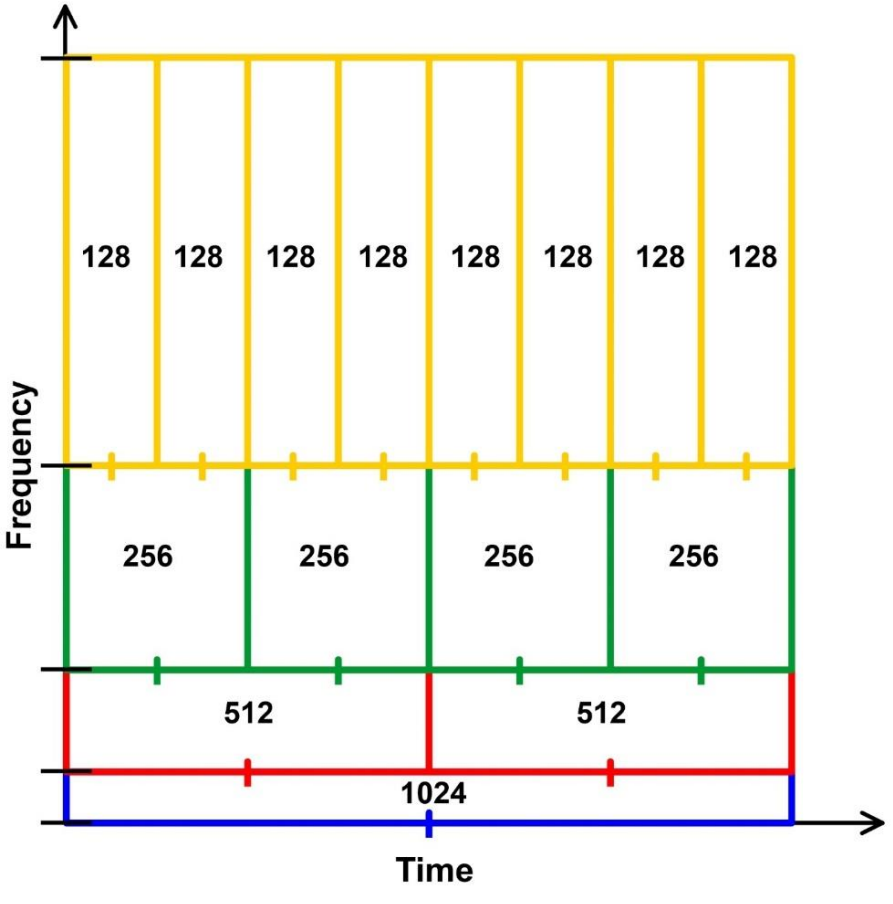


Figure 8–14. Time / frequency view of the STFT. In this example, the original length of our time series was 1,024 time units. The numbers in each rectangle indicate the size of the FFT, while the vertical tick marks indicate the central point, in time, of each FFT.

If the size of the FFT is N, then edge of the cone of influence lies $N/2$ time units from the vertical (frequency) axis. We cannot make any valid inferences about the results that lie in the shaded region.

The better we try to resolve the time at which we can produce frequency information, the more frequency information we lose (starting from the low frequency end of the spectrum). This is analogous to Heisenberg's

uncertainty principle: *the momentum and position of an object cannot be quantified at the exact same time*. We can't 'simultaneously' resolve all of the time and frequency information. The best that we can do is to say that 'this range of frequencies is present in this time interval'.

So, while in principle the STFT can solve much of our problem, (a) it is not particularly efficient and (b) there is a trade-off between time and frequency resolution. Wavelet analysis is a substantial improvement to the STFT, and is much more efficient.

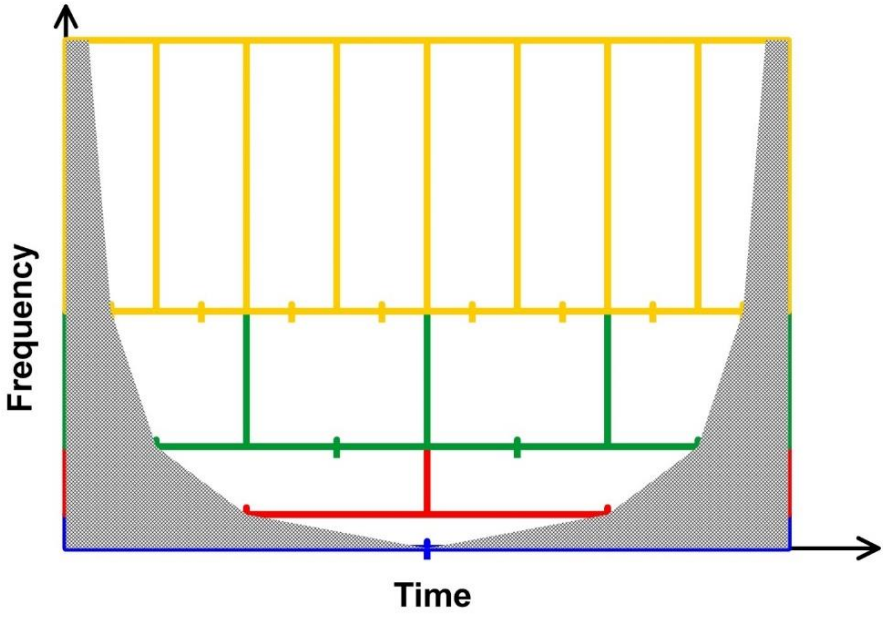


Figure 8–15. Time / frequency view of the STFT and the 'cone of influence' – the shaded region.

Before we move on to wavelet analysis, let's have a look at the STFT in a bit more detail for the time series, $x(t)$. After all, the STFT is a valid spectral analysis technique. Often, it is reasonable to assume that low frequency features are present through the whole time series (and the Fourier transform of the whole data set will identify such features) and that high frequency features are short-lived (and the STFT will identify

these features, when they are present in the time series). So, the STFT will likely produce useful information about our data.

8.2.2. The STFT and Convolution

Let's revisit our problem from a slightly different perspective. We have a time series $x(t)$ and when we take its Fourier transform, we lose all temporal information. The problem is: we want to retain as much temporal information as possible AND we want as much frequency information as possible. (Noting that there is a limit to how much of both time and frequency information that we can simultaneously retain).

One possible solution that comes to mind is to apply the STFT. We can represent the STFT as a rectangular window (which we will pass through the time series). For example, in Figure 8–16 we have constructed a rectangular 'window' of height unity and length $\tau = 8$ time units long, which we represent by the function $h(t)$. We pass (or move) $h(t)$ through the time series $x(t)$. The data comprising $h(t)$ changes after each move on its journey though $x(t)$. After each move, we undertake a FFT. Each move results in one spectrum, using the data in $h(t)$, which is centred on the middle time in $h(t)$.

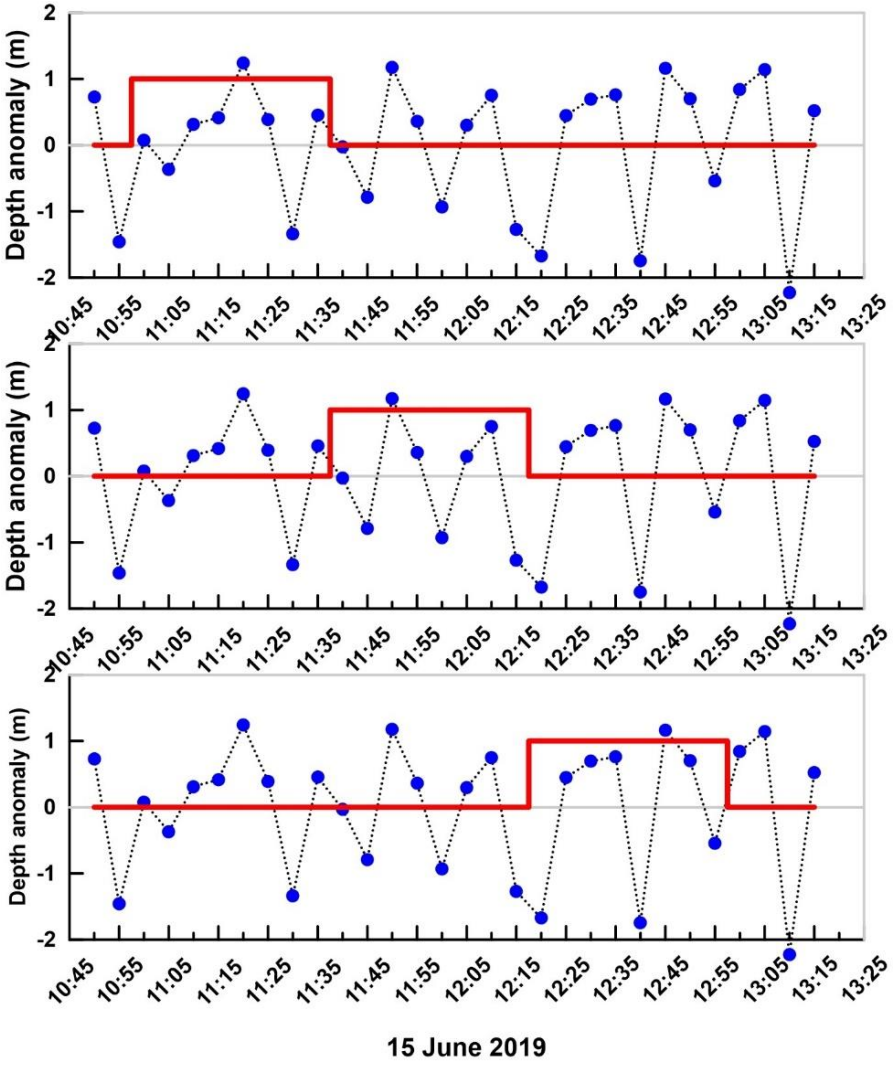


Figure 8–16. Time series of depth anomalies with a rectangular window of length 8 time units passing through the time series.

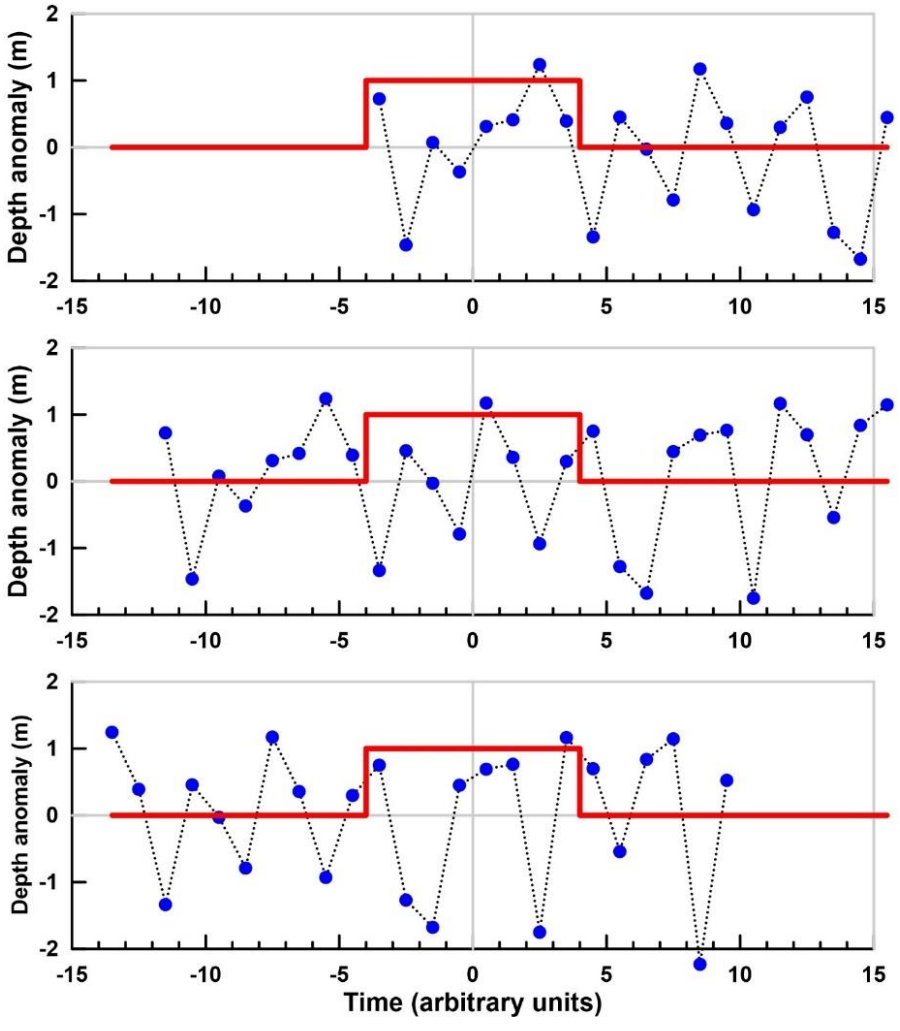


Figure 8–17. Time series of depth anomalies with a rectangular window of length 8 time units. The rectangular window is fixed in time and the time series passes through the window.

Without loss of generality, we can centre $h(t)$ on $t = 0$ and pass the time series $x(t)$ through the window $h(t)$ as shown in Figure 8–17 (where the time series is moving from right to left through the window). Therefore, $h(t)$ can be defined as:

$$h(t) = \begin{cases} 1 & for \quad -\tau/2 \le t \le \tau/2 \\ 0 & elsewhere \end{cases}$$

Note that $h(t)$ is symmetric (i.e. an even function) about $t = 0$, hence $h(-t) = h(t)$.

This process of passing one function through another (in this case passing a window, $h(t)$, through a time series, $x(t)$, or conversely, passing a time series through a window) is known as convolution, which is defined as:

$$(x * h)(t) = \int_{-\infty}^{\infty} x(\tau)\, h^*(t-\tau)\, d\tau \equiv \int_{-\infty}^{\infty} x(t-\tau)\, h^*(\tau)\, d\tau$$

- Convolution uses the complex conjugate (indicated by a superscripted asterisk) to ensure the product of x and h always remains positive. If we don't do this then the length of the vector produced by the product of x and h may become negative.
- If h is real, then the complex conjugate of h is equivalent to h itself.
- It is a little unfortunate that the symbol for the convolution operation is also, traditionally, indicated by an asterisk.

Swapping the t and τ variables, the convolution can be rewritten as:

$$(x * h)(\tau) = \int_{-\infty}^{\infty} x(t)\, h^*(\tau-t)\, dt \equiv \int_{-\infty}^{\infty} x(\tau-t)\, h^*(t)\, dt$$

Now $h^*(t)$ is an even function so that $h^*(\tau - t) = h^*(t - \tau)$, and

$$(x * h)(\tau) = \int_{-\infty}^{\infty} x(t)\, h^*(t-\tau)\, dt$$

Let $h^*(t-\tau) = g^*(t-\tau)\, e^{-i\omega t}$, then:

$$(x * h)(\tau) = \int_{-\infty}^{\infty} x(t)\, g^*(t-\tau)\, e^{-i\omega t}\, dt$$

The convolution of $x(t)$ and $h(t)$ can now be seen as the Fourier transform of $[x(t)\ g^*(t-\tau)]$.

If the height of the window is unity (and its length is τ), then:

$$g^*(t-\tau) = \begin{cases} 1 & for \quad -\tau/2 \le t \le \tau/2 \\ 0 & elsewhere \end{cases}$$

Where we've fixed the window in time and centred it on time $t = 0$ and we pass the time series through the window. (This is equivalent to moving the window through the time series). Then:

$$(x * h)(\tau) = \int_{-t/2}^{t/2} x(t)\ e^{-i\omega t}\ dt \ = STFT_{\tau,\omega}(t)$$

In other words, each convolution is equivalent to the short time Fourier transform over that time interval.

If we set $\tau = t$, the length of the window spans the entire time series, then

$$(x * h)(\tau) = \int_{-\infty}^{\infty} x(t)\ e^{-i\omega t}\ dt \ = \mathcal{F}(t)$$

Therefore, we can think of the usual Fourier transform as a windowed transform where the window is of unit height and its length spans the whole time series. Similarly, the STFT is a windowed transform where the height of the window is unity and the length spans the length of the window.

This is all very nice, but how does it help our problem? The whole convolution process seems complicated, time consuming and appears to be of no real benefit. Well, it turns out that convolution in the time domain is equivalent to multiplication in the frequency domain. (Conversely, multiplication in the time domain is convolution in the frequency domain). So, if we're already using Fourier (or spectral) analysis as part of our analytical tools, we can apply convolution by using multiplication in the frequency domain. Recall that convolution is

equivalent to the STFT, hence implementation of the STFT then becomes much easier and quicker. This is demonstrated in the box below.

Convolution and the Fourier transform

The convolution operation is computationally expensive. There is a way in which this operation can be sped up substantially using the Fourier transform.

The convolution of two functions $y(t)$ and $g(t)$ is defined as:

$$(y * g)(t) = \int_{-\infty}^{\infty} y(\tau)\, g^*(t-\tau)\, d\tau$$

Take the Fourier transform:

$$\mathcal{F}[(y * g)(t)] = \int_{-\infty}^{\infty} \left\{ \int_{-\infty}^{\infty} y(\tau)\, g^*(t-\tau)\, d\tau \right\} e^{-i\omega t}\, dt$$

As $e^{-i\omega t}$ is independent of τ it can be brought inside the integral with respect to τ.

$$\mathcal{F}[(y * g)(t)] = \int_{-\infty}^{\infty} \left\{ \int_{-\infty}^{\infty} y(\tau)\, g^*(t-\tau)\, e^{-i\omega t}\, d\tau \right\} dt$$

Reverse the order of the integrals, so that:

$$\mathcal{F}[(y * g)(t)] = \int_{-\infty}^{\infty} \left\{ \int_{-\infty}^{\infty} y(\tau)\, g^*(t-\tau)\, e^{-i\omega t}\, dt \right\} d\tau$$

We can bring $y(\tau)$ outside of the integral with respect to t as it is independent of t.

$$\mathcal{F}[(y * g)(t)] = \int_{-\infty}^{\infty} y(\tau) \left\{ \int_{-\infty}^{\infty} g^*(t-\tau)\, e^{-i\omega t}\, dt \right\} d\tau$$

Let $T = t - \tau$, then $\frac{dT}{dt} = 1$ or $dT = dt$, then:

$$\mathcal{F}[(y * g)(t)] = \int_{-\infty}^{\infty} y(\tau) \left\{ \int_{-\infty}^{\infty} g^*(T)\, e^{-i\omega(T+\tau)}\, dT \right\} d\tau$$

$$= \int_{-\infty}^{\infty} y(\tau)\, e^{-i\omega\tau} \left\{ \int_{-\infty}^{\infty} g^*(T)\, e^{-i\omega T}\, dT \right\} d\tau$$

$$= \int_{-\infty}^{\infty} y(\tau)\, e^{-i\omega\tau}\, d\tau \int_{-\infty}^{\infty} g^*(T)\, e^{-i\omega T}\, dT$$

$$\equiv \mathcal{F}[y(t)]\ \mathcal{F}[g^*(t)]$$

Where we've changed the T to t in the second integral, which we can do as it is a dummy variable.

The Fourier transform of the convolution of two functions is the product of the Fourier transforms of the individual functions. So, the convolution of two functions in time can be written as the inverse Fourier transform of the product of the Fourier transforms of the individual functions, viz.:

$$(y * g)(t) = \mathcal{F}^{-1}\{\mathcal{F}[y(t)]\ \mathcal{F}[g^*(t)]\}$$

The implication of this is, if we choose $g^*(t)$ such that it has an analytic Fourier transform, then we only need to take one Fourier transform (of $y(t)$), multiply it by the function that is the analytic Fourier transform of $g^*(t)$ and take the inverse Fourier transform of this product. While this appears to be a longer route to solving the problem, it is substantially quicker.

For example, assume that

$$g^*(t-\tau) = \begin{cases} 1 & for \quad -\tau/2 \le t \le \tau/2 \\ 0 & elsewhere \end{cases}$$

Then the Fourier transform of $g^*(t-\tau)$ is:

$$\mathcal{F}[g^*(t-\tau)] = \int_{-\infty}^{\infty} g^*(t-\tau)\, e^{-i\omega t}\, dt = \int_{-\tau/2}^{\tau/2} 1.e^{-i\omega t}\, dt$$

$$= \frac{1}{-i\omega} e^{-i\omega t}\Big|_{-\tau/2}^{\tau/2}$$

$$= \frac{1}{-i\omega}\left[e^{-i\omega\tau/2} - e^{i\omega\tau/2}\right] = \frac{2}{\omega}\left[\frac{e^{i\omega\tau/2} - e^{-i\omega\tau/2}}{2i}\right]$$

$$= \frac{2}{\omega} sin\left(\omega\tau/2\right) = \tau\, sinc\left(\omega\tau/2\right)$$

Hence, to evaluate the Fourier transform of the rectangular window, we only need to specify the values of ω and τ. The convolution problem then simplifies to:

$$(y * g)(t) = \mathcal{F}^{-1}\left\{\mathcal{F}[y(t)] \;\; \tau\, sinc\left(\omega\tau/2\right)\right\}$$

In summary, for our rectangular window:

(a) In the time domain, we run one convolution then take the Fourier transform of the result. This is repeated for each time step.
(b) In the frequency domain, we take one Fourier transform of the original data and multiply it by the *sinc* function for selected values of ω and τ. Then repeat for different values for ω and τ.

The process is demonstrated in Figure 8–18, which is based on a real data set. We start with a signal (i.e. a time series) and a window as a function of time (Figure 8–18, top). From there we can choose to follow the time path or the frequency path.

The time path uses convolution of the signal and window. We then take the Fourier transform of the resulting convolution.

Alternatively, we can follow the frequency path. We take the Fourier transform of the whole signal and that of the window. We then (piecewise) multiply the FFT of the signal by the FFT of the window. The result is exactly the same as the FFT of the convolution. We can also

return to the original convolution signal by taking the inverse FFT of this product.

We saw above that if we chose a rectangular window, then its Fourier transform is the *sinc* function with a single parameter (τ). If we are clever in the choice of our type of window (i.e. its Fourier transform is analytic), then we can greatly simplify the number of computations involved in following the frequency path. This is the lead-in to the wavelet transform.

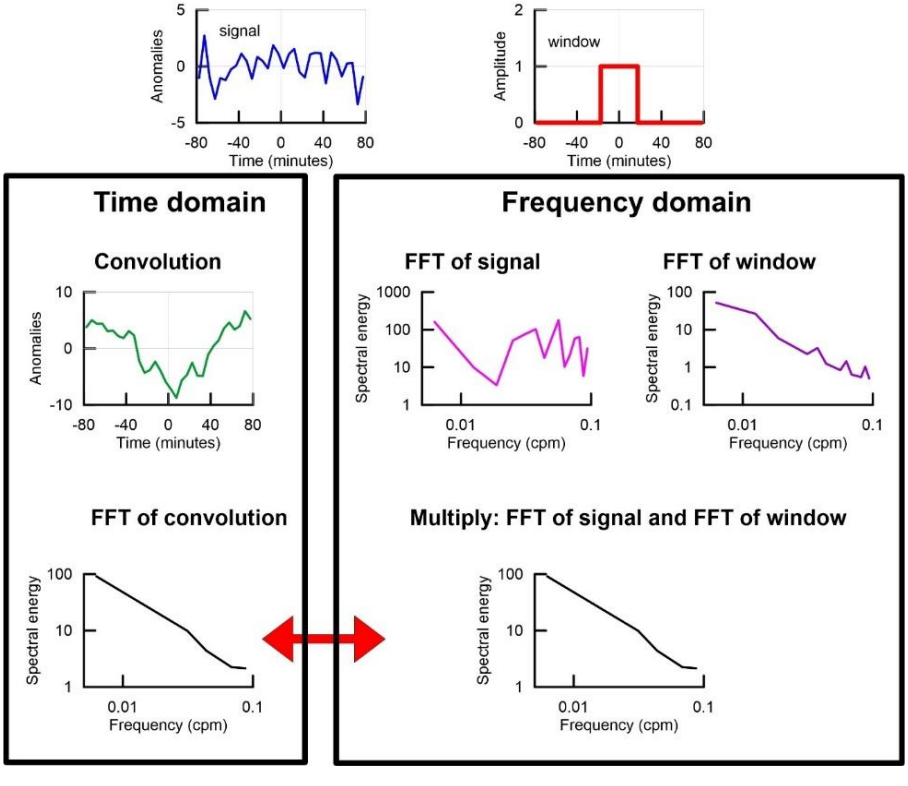


Figure 8–18. Convolution in the time domain versus multiplication in frequency domain.

We'll demonstrate the use of the STFT using similar examples to those used in Section 8.1. In these examples we dispense with the questions and

processing that we would normally undertake. It won't substantially affect the points that are being demonstrated in these examples.

8.3. Examples using the STFT

8.3.1. Example 8-6 – A pair of sine curves

Consider a time series (1,024 data points at 5 minute intervals) in which the data are from a sine curve of constant amplitude with frequency f in the first half of the time series and frequency $f/2$ in the second half of the time series (Figure 8–19, bottom plot).

The top plot in Figure 8–19 is the spectrogram of the whole data set. The spectrum is centred on time 2,560 minutes, but applies to the entire time series.

Clearly seen in the spectrum are two peaks, corresponding to the two different frequencies. These spectral peaks are relatively broad indicating that neither frequency is present over the whole time series – the time series is not stationary over the entire time domain. Observed at the high frequency end of the spectrum are slight 'wobbles', another indication that neither frequency is present over the whole time series.

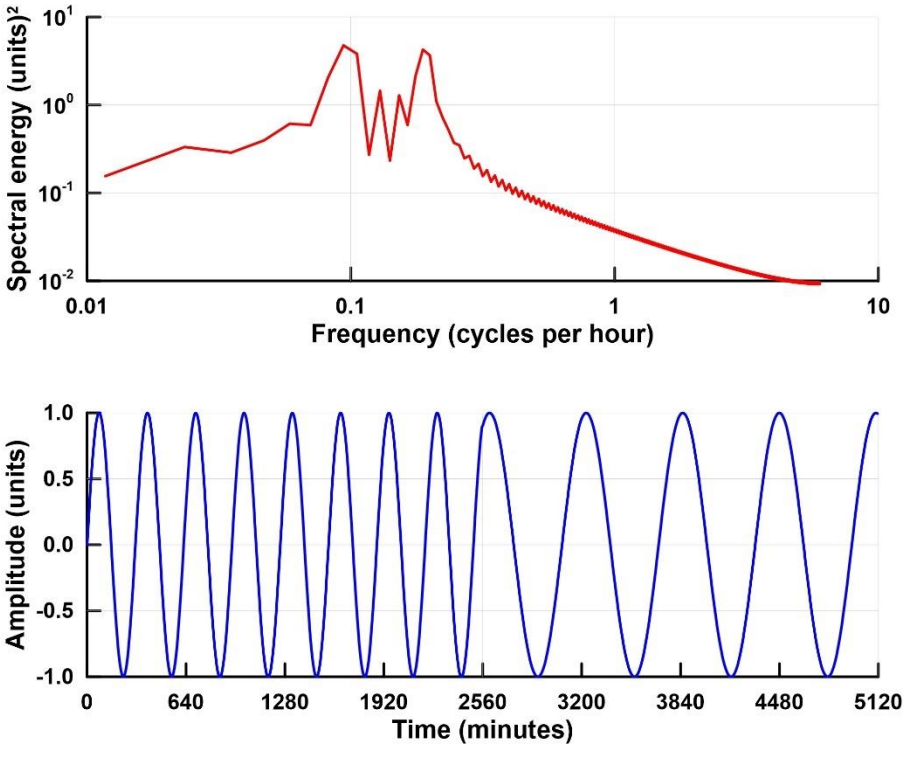


Figure 8–19. Bottom panel: Time series of a sine wave with one frequency for the first half of the record and half of that frequency in the second half of the record. Top panel: spectral (or Fourier) analysis of the whole time series.

In this example we know that there are two different frequencies in each the two halves of the time series. Let's reanalyse the data using the STFT, where the size of each window is $\tau = 512$ records.

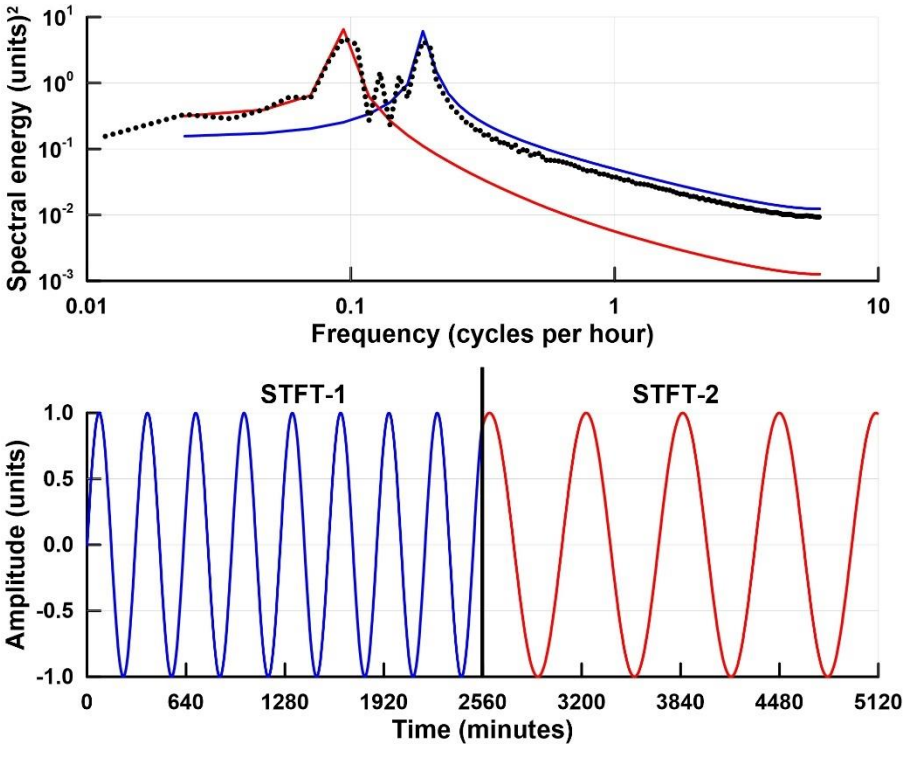


Figure 8–20. Bottom panel: Two time series, each a sine wave with frequency f for the first half of the record and $f/2$ in the second half of the record. Top panel: the corresponding spectral analyses of the two time series.

The time series is plotted in Figure 8–20 (bottom panel) where the first time series is coloured blue and the second time series is coloured red. Also included in Figure 8–20 (top panel) are the respective spectra for the two time series. Superimposed on this plot is the spectrum resulting from analysing the whole time series (black dotted line). Each of the STFT spectra have sharp peaks (at about 0.09 cycles per hour and 0.19 cycles per hour). The first spectrum is centred on time 1,280 minutes but represents the first 2,560 minutes of the time series. The second spectrum is centred on time 3,840 minutes but represents the second 2,560 minutes of the time series. There are no 'wobbles' at the high frequency end of either spectra. Of note is the loss of information at the low frequency end

of the spectrum of the two STFT spectra compared with the single spectrum of Figure 8–19, although we have (slightly) improved our temporal resolution.

So, it might appear that the STFT solves the problem – we have improved our temporal resolution, but this is at the expense of low frequency information. In fact, we have solved the problem for this simple example. However, in most applied data sets we won't unambiguously know the number of, or the locations where, the changes in frequency (or amplitude) occur – although simply examining the time series data may provide some clues.

8.3.2. Example 8-7 – A chirp

We revisit the chirp signal, as defined in Section 8.1.3. A chirp signal is characterised by a continuously changing frequency. The chirp is plotted in Figure 8–21 (bottom panel). The different colours of the four sections in the time series plot represent the time periods over with the four STFT were performed. The length of the data set is 1,024 data points (arbitrarily set at 5 minute intervals). Three sets of spectral analyses are undertaken:

(a) The first spanning the whole data set (the corresponding spectrum is the left hand side spectrogram).
(b) The second comprises two STFTs, each of length 512 records (no overlap) and the corresponding spectra are in the centre spectrogram.
(c) The third comprises four STFTs, each spanning 256 records (no overlap) and the corresponding spectra are in the right hand spectrogram.

The spectrum of the whole data set has a broad 'peak' between about 0.08 and 0.28 cycles per hour (Figure 8–21, left hand spectrogram). At frequencies less than about 0.3 cycles per hour the spectrum is relatively flat (similar to a random signal). However, in this case, the Fourier breakdown of the signal cannot determine a single dominant frequency and it spreads the energy over a very broad frequency range.

The spectra from each of the two half-data sets includes two peaks, one near 0.08 and one near 0.28 cycles per hour (Figure 8–21, centre spectrogram). Although these two peaks are still relatively broad, this is a substantial improvement on the whole-of-data spectrum. We appear to be making progress.

However, the spectra from each of the four quarter-data sets now has four peaks, each of roughly the same amplitude (Figure 8–21, right hand spectrogram). The spectral peaks have sharpened. If we split the data set into 8 sets of equal length, we'll end up with 8, even sharper, spectral peaks and so on. Further, as we use windows of smaller size, we lose low frequency information, as we are starting to see in the right hand spectrogram (Figure 8–21). So, the STFT has not really solved the problem for the chirp signal. The more we improve the time resolution the more spectral peaks we'll see in the range 0.08 to 0.28 cycles per hour. However, we will start to lose low frequency information and the 'peaks' may start to disappear, as observed in the green data set. Eventually the spectra will appear as if the trend has not been removed.

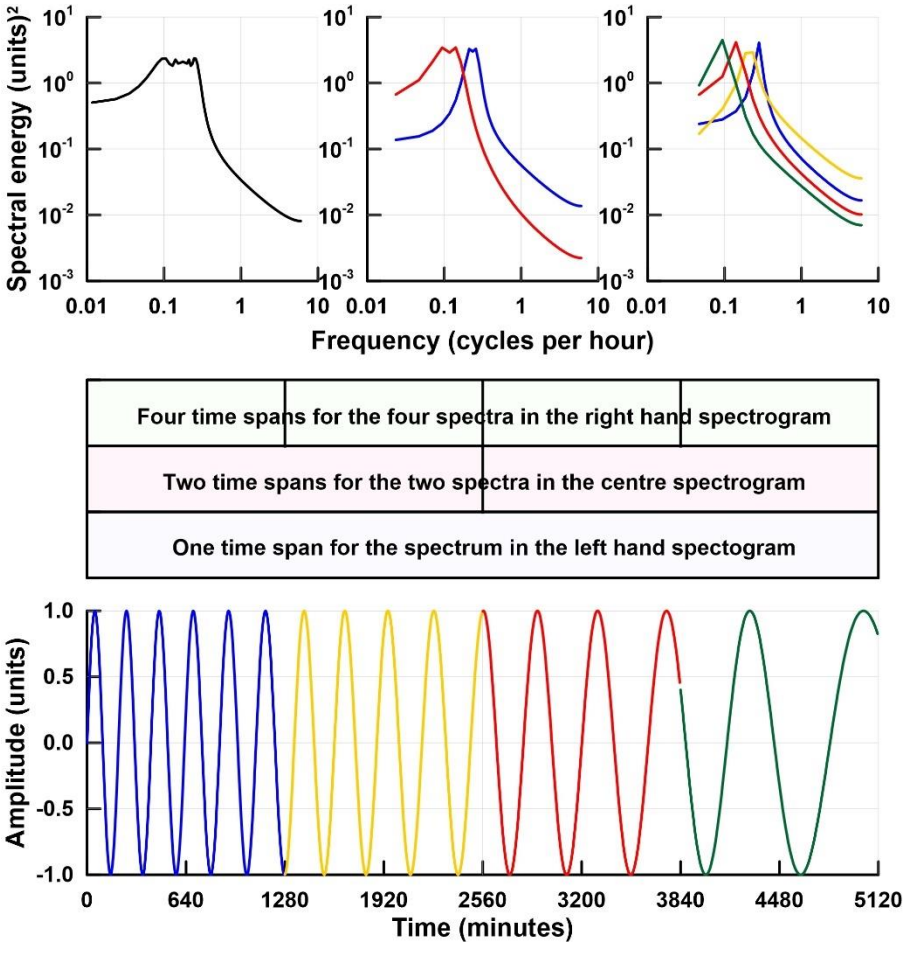


Figure 8–21. Bottom panel: Time series of a chirp signal. Top panels: Spectral analyses of (a) the whole data set (left hand spectrum), (b) two data sets each half the length of the first data set (spectra in the centre plot), and (c) four data sets each of length one quarter of the original data set (spectra in the right hand plot).

8.3.3. Example 8-8 – Internal waves

The final example of the STFT is a data set that includes periods of internal waves. An internal wave is similar to the ordinary surface wave we see at the beach, which propagates along the boundary between the air and the water. Internal waves propagate inside the water, along an interface formed by water bodies of different density at different depths in the water column. In many cases we can approximate the density by its thermal signal. In Figure 8–22 (bottom panel) the internal wave is characterised by the vertical displacement of the 20°C isotherm. The time span is slightly more than 21 hours and the time between readings is 5 minutes.

From Figure 8–22 (bottom panel) we immediately notice that there are periods of no internal wave activity and periods of substantial internal wave activity (large rapid displacements of the 20°C isotherm). Clearly, the data set is non-stationary. The spectrum of this data set is shown in Figure 8–22 (top panel). There are no spectral peaks, although a slight raising of the spectral energy near 3-4 cycles per hour. So, we might think that there is nothing particularly interesting in the data set – no significant signals. The spectrum falls within the inertial subrange, hence we expect that the spectrum would follow the -5/3rds power law (in the absence of any energy sources), which is observed in the figure.

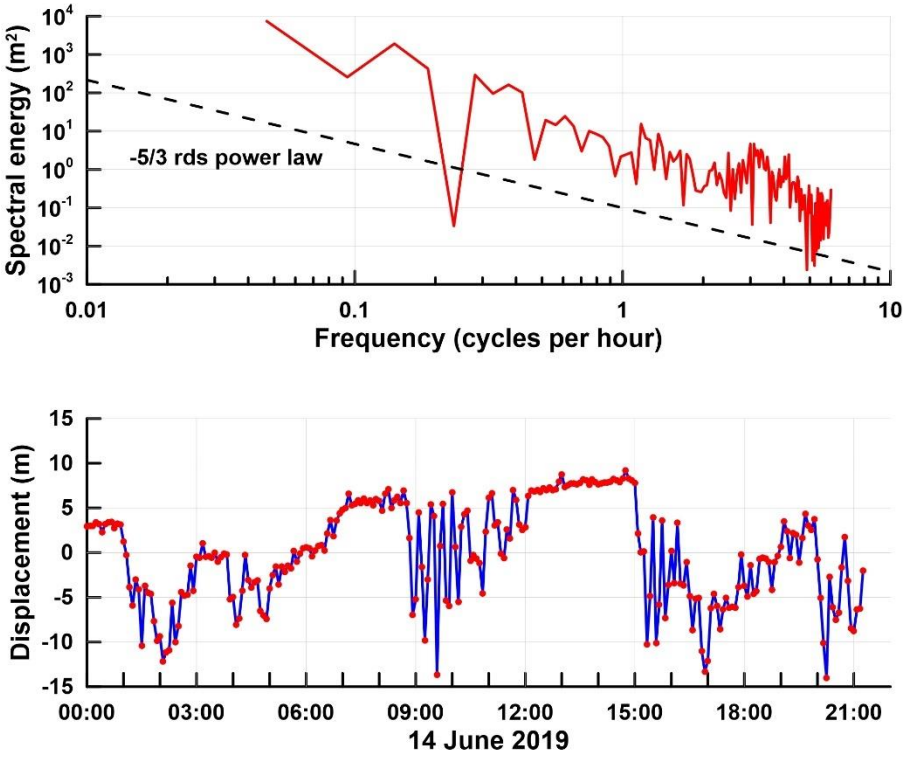


Figure 8–22. Bottom panel: Time series of the vertical displacement of the 20°C isotherm. Top panel: Spectral analyses of the whole data set, together with the -5/3rds power law line.

Let's re-examine this time series using the STFT. Prior to undertaking the STFT, several decisions need to be made. Some are outlined below.

1. What is the size of the window we want to use? Often a window of constant size will contain both stationary and non-stationary elements, and we may use a window of variable length.
2. The starting point of the window may be critical. This is particularly so if we choose a window of fixed length.
3. Do we use a DFT or an FFT? Most FFT algorithms are 'radix 2' (i.e. they require the length of the window to be a power of two), which can be restrictive. However, many Fourier transforms may be required in which case the speed associated with the FFT is

advantageous. Given the speed of computers nowadays, use of the FFT is, perhaps, less of a factor, while the DFT provides more flexibility.

4. Do we remove the trend from the individual windows or just the overall trend? Ideally, we'd treat each the FFT from each window individually i.e. we remove the trend from each individual window. However, the different windows will have different slopes and offsets. Further, removal of a linear signal may not be appropriate – the signal simply may not be linear. This will make the interpretation of the overall results more complex.

Based on an examination of the time series plot in Figure 8–22, the following decisions were made:

1. A window of size 32 data points (160 minutes) was selected. This covered most of the data set. However, to ensure we have an approximately stationary data set, the end of one window does not necessarily align with the start of the next window.
2. The start of the first window occurs 45 minutes after the start of the original data set.
3. The FFT was chosen as the spectral analysis tool.
4. Trends were removed from the individual windows resulting in a set of displacement anomalies. The removed trends were not always linear. The types of trends removed are indicated on Figure 8–23 (bottom panel).

A FFT was undertaken for each of the seven windows shown in Figure 8–23 using the anomalies. The number of data points in each FFT was 32.

The result from the seven STFTs is shown in Figure 8–24, presented as a contour plot (similar to the presentation of the output from a wavelet analysis). Clearly, the spectra between about 09:00 and 12:00 and again between 15:00 and 18:00 contain large peaks (high spectral energy) coincident with the internal wave activity. The dominant frequency is between about 3 and 3.5 cycles per hour.

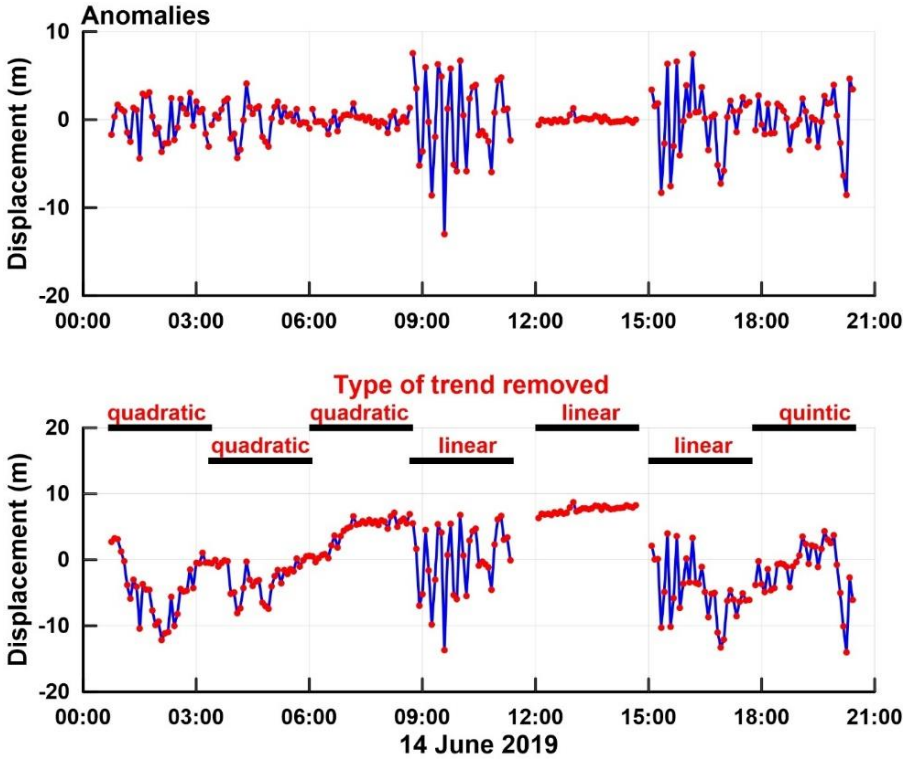


Figure 8–23. Bottom plot: Time series of the vertical displacement of the 20°C isotherm. The overbars at the top of the plot indicate the length and location of each window. Top plot: Time series of the anomalies for each window.

This is a substantial improvement on the spectral plot shown in Figure 8–22. However, we have gone to considerable effort to arrive at this plot: we exercised considerable judgement on the size of the FFT, the starting points and the start and end times did not always align. These decisions were largely based on a visual examination of the data set. While this was manageable for the relatively short time series examined, other data sets may provide easy decision making. Further, the problem of temporal resolution verses loss of low frequency information remains.

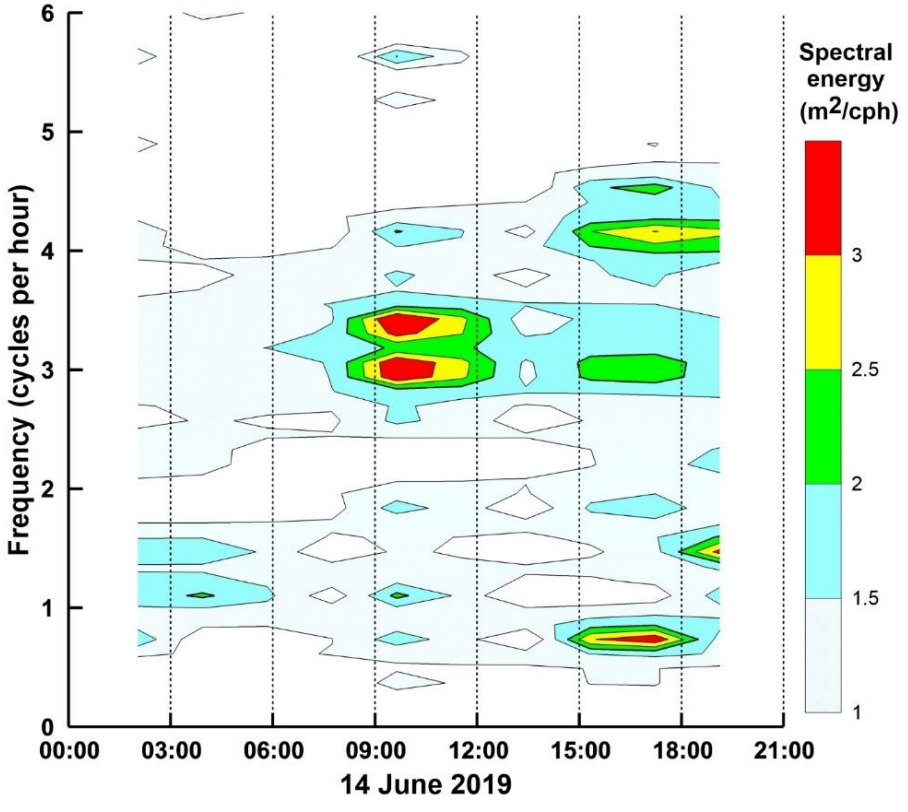


Figure 8–24. Contour plot of the STFT of the vertical displacement of the 20°C isotherm using seven FFTs each of size 32 records.

The analysis is repeated using 32 FFTs, each of 8 records in length. A separate quadratic trend was removed from each of the 32 windows. The results are presented in Figure 8–25. Compared with Figure 8–24, there is improved temporal resolution (although it is a bit difficult to see for this data set), particularly at each end of the data set. There is a clear reduction in the frequency resolution – much of the frequency information is lost. Specifically, there is no information at frequencies less than about 1 cycle per hour.

To interpret the spectral results associated with this data set, there is better / clearer information provided in Figure 8–24 than in Figure 8–25.

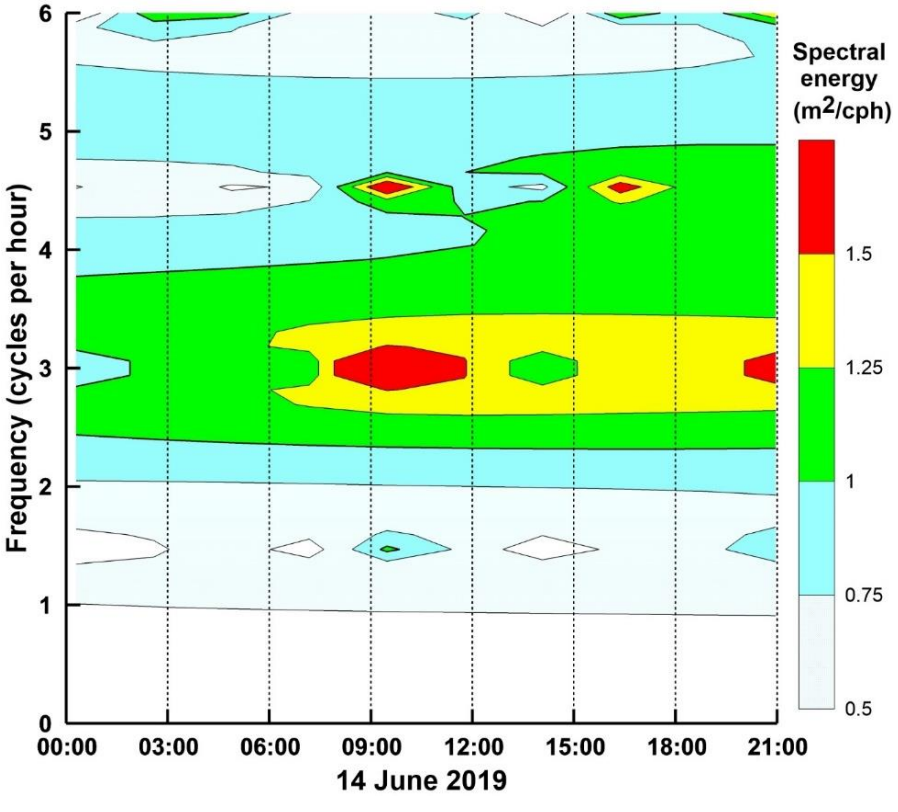


Figure 8–25. Contour plot of the STFT of the vertical displacement of the 20°C isotherm using 32 FFTs each of size 8 records.

Chapter 9. The Morlet wavelet

9.1. Theory of the Morlet wavelet

In the previous chapter, we passed our time series through a rectangular window. This is a mathematical process known as convolution and is one way in which we can improve the temporal resolution in a spectral plot. In the time domain, implementing convolution is, computationally, relatively expensive. This can be substantially reduced by undertaking a Fourier transform of the data and the window and multiplying the results in a piecewise manner. Convolution in the time domain is equivalent to multiplication in the frequency domain (and vice versa).

If the window is designed such that it has an analytic Fourier transform, then the computations become faster again. The Fourier transform of the rectangular window is a $sinc$ function, which depends on the parameter τ. To evaluate the Fourier transform of the rectangular window, we only need to specify the width of the window (τ) for each value of the frequency, ω. This is the basis of the STFT.

Convolution of the data (x) and the window (h) is defined as:

$$(x * h)(\tau) = \int_{-\infty}^{\infty} x(t)\, h^*(t-\tau)\, dt$$

Let $h^*(t-\tau) = g^*(t-\tau)\, e^{-i\omega t}$, then

$$(x * h)(\tau) = \int_{-\infty}^{\infty} x(t)\, g^*(t-\tau)\, e^{-i\omega t}\, dt$$

To generate the STFT, we chose $g^*(t-\tau)$ to be a rectangular window of height unity and length τ:

$$g^*(t-\tau) = \begin{cases} 1 & for \quad -\tau/2 \le t \le \tau/2 \\ 0 & elsewhere \end{cases}$$

Then,

$$(x * h)(\tau) = \int_{-\tau/2}^{\tau/2} x(t)\ e^{-i\omega t}\, dt$$

The rectangular window places equal emphasis on the data points within that window. Perhaps a neater option is to construct a window that places greater emphasis on data at the centre of the window and less emphasis on the data at the ends of the window. There are many options available to achieve this, but one possible option is to choose $g^*(t-\tau)$ such that it is Gaussian-like.

From statistics, the Gaussian (or normal) distribution is given by:

$$\frac{1}{\sigma\sqrt{2\,\pi}}\ e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$$

Where μ is the mean and σ is the standard deviation of the data.

Let's redefine these parameters such that: $x = t$, $\mu = \tau$ and $\sigma = s$. The rectangular window is specified by a single parameter, τ. The Gaussian has introduced a second parameter s (or σ) and we have a two-parameter system (τ, s). The s parameter is related to the frequency of the sine / cosine curve in this window, indicated by ω_0. The scale and the frequency combine to produce the number of waves (n) in the window, which we now call a wavelet: $n = s\omega_0$. So, $g^*(t-\tau)$ is rewritten as:

$$g^*\left(\frac{t-\tau}{s}\right) = \frac{1}{s\sqrt{2\,\pi}}\ e^{-\frac{1}{2}\left(\frac{t-\tau}{s}\right)^2}$$

Let's now combine the 'Gaussian' component (above) with a trigonometric component, so that the window resembles a cosine curve whose amplitude decreases with distance from its centre. It is defined as:

$$g^*\left(\frac{t-\tau}{s}\right) = \phi^*(t)_{\tau,s}$$
$$= \left\{\frac{1}{s\sqrt{2\pi}}\ exp\left[-\frac{1}{2}\left(\frac{t-\tau}{s}\right)^2\right]\right\}\ exp\left[-i\omega_0\left(\frac{t-\tau}{s}\right)\right]$$

Where the function inside the curly brackets is the Gaussian and the second 'exp' term is the sine / cosine component. This is (effectively) the Morlet wavelet, and it is a complex function.

The Morlet wavelet is basically an extension of Gabor's original study (Gabor, 1946) in which he defined an 'elementary signal' which is essentially the Morlet wavelet (Morlet et al, 1982a, 1982b). The basic difference between these two bodies of work lies in Morlet's work, which overcame inherent difficulties (instabilities) in Gabor's work resulting in a substantial improvement in the 'spatial resolution as frequency gets higher and higher' (Debnath, 1998).

We call the above expression the Morlet daughter wavelets, each daughter wavelet is given by specific, user-defined, values of τ and s. The parameter $\omega_0 = \frac{n}{s}$ signifies the number of ('easily discernible') waves (n) in the wavelet.

The Morlet mother wavelet is the generic form of the above set of daughter wavelets and is obtained by setting $\tau = 0$ and $s = 1$, so that:

$$\phi^*(t)_{0,1} = \phi^*(t) = \left\{\frac{1}{\sqrt{2\pi}}\ exp\left(-\frac{t^2}{2}\right)\right\}\ exp(-i\omega_0 t)$$

The component inside the curly brackets is analogous to the 'standard normal' distribution in which the mean is zero and the standard deviation is unity.

Using the Morlet daughter wavelet expression and the convolution theorem, we can show that the convolution of the data and the (Morlet) window is a simple multiplication of these functions in frequency space. Further, the Fourier transform of the Morlet window is analytic – it is a Gaussian curve.

Recall that convolution is simply the passing of one function (the data) through another function (the window) or vice versa – which we want to undertake to improve the time / frequency resolution of the data set. We take the Fourier transform of the data and the Fourier transform of the Morlet window and multiply the two. However, as the Fourier transform of the Morlet window is analytic, we only need to select and apply values for τ and s to generate its Fourier transform.

Take the Fourier transform of the convolution, then use the convolution theorem:

$$\mathcal{F}[(x * h)(\tau, s)] = \mathcal{F}\left[\int_{-\infty}^{\infty} x(t)\, h^*\left(\frac{t-\tau}{s}\right)\, dt\right]$$

$$= \mathcal{F}[x(t)]\, \mathcal{F}\left[h^*\left(\frac{t-\tau}{s}\right)\right] = \mathcal{F}[x(t)]\, \mathcal{F}\left[\phi^*\left(\frac{t-\tau}{s}\right)\right]$$

Substitute in the expression for the Morlet wavelet:

$$\mathcal{F}\left[\phi^*\left(\frac{t-\tau}{s}\right)_{\tau,s}\right]$$

$$= \int_{-\infty}^{\infty} \left\{\frac{1}{s\sqrt{2\pi}}\, exp\left[-\frac{1}{2}\left(\frac{t-\tau}{s}\right)^2\right]\, exp\left[-i\omega_0\left(\frac{t-\tau}{s}\right)\right]\right\} exp(-i\omega t)\, dt$$

Let $T = \frac{t-\tau}{s}$, then $t = sT + \tau$ and $dt = s\, dT$. Substitute into the above equation:

$$\mathcal{F}[\phi^*(T)] = \int_{-\infty}^{\infty} \left\{\frac{1}{s\sqrt{2\pi}}\, exp\left[-\frac{1}{2}(T)^2\right]\, exp[+i\omega_0(T)]\right\} exp[-i\omega(sT + \tau)]\, s\, dT$$

Rearranging:

$$\mathcal{F}[\phi^*(T)] = \frac{1}{\sqrt{2\pi}}\, exp(-i\omega\tau) \int_{-\infty}^{\infty} exp\left[-\frac{1}{2}(T)^2 + i\omega_0(T) - i\omega sT\right]\, dT$$

$$= \frac{1}{\sqrt{2\pi}}\ exp(-i\omega\tau) \int_{-\infty}^{\infty} exp\left\{-\left[\frac{T^2}{2} + i(\omega s - \omega_0)T\right]\right\}\ dT$$

Complete the square for the expression inside the square brackets and rearrange:

$$= \frac{1}{\sqrt{2\pi}}\ exp(-i\omega\tau) \int_{-\infty}^{\infty} exp\left\{-\left[\frac{T}{\sqrt{2}} + \frac{i(\omega s - \omega_0)}{\sqrt{2}}\right]^2 - \frac{(\omega s - \omega_0)^2}{2}\right\}\ dT$$

$$= \frac{1}{\sqrt{2\pi}}\ exp(-i\omega\tau)\ exp\left[-\frac{(\omega s - \omega_0)^2}{2}\right] \int_{-\infty}^{\infty} exp\left\{-\left[\frac{T}{\sqrt{2}} + \frac{i(\omega s - \omega_0)}{\sqrt{2}}\right]^2\right\}\ dT$$

Let $X = \frac{T}{\sqrt{2}} + \frac{i(\omega s - \omega_0)}{\sqrt{2}}$, then $\sqrt{2}\,dX = dT$. Thence:

$$\mathcal{F}[\phi^*(T)] = \mathcal{F}\left[\phi^*\left(\frac{t-\tau}{s}\right)\right]$$

$$= \frac{1}{\sqrt{2\pi}}\ exp(-i\omega\tau)\ exp\left[-\frac{(\omega s - \omega_0)^2}{2}\right] \int_{-\infty}^{\infty} exp\{-[X]^2\}\ \sqrt{2}\,dX$$

$$= \frac{1}{\sqrt{2\pi}}\ exp(-i\omega\tau)\ exp\left[-\frac{(\omega s - \omega_0)^2}{2}\right] \sqrt{\pi}\sqrt{2}$$

$$= exp(-i\omega\tau)\ exp\left[-\frac{(\omega s - \omega_0)^2}{2}\right]$$

The Fourier transform of the Morlet wavelet is analytic. So, from the convolution theorem,

$$\mathcal{F}[(x * h)(\tau, s)] = \mathcal{F}[x(t)]\ \mathcal{F}\left[\phi^*\left(\frac{t-\tau}{s}\right)\right]$$

$$= \mathcal{F}[x(t)]\ \ exp(-i\omega\tau)\ exp\left[-\frac{(\omega s - \omega_0)^2}{2}\right]$$

We can generate the 'convolved' time series $y(t)$ by taking the inverse Fourier transform of this expression.

$$y(t) = \mathcal{F}^{-1}\left\{\mathcal{F}[x(t)] \;\; exp(-i\omega\tau)\, exp\left[-\frac{(\omega s - \omega_0)^2}{2}\right]\right\}$$

$$= \int_{-\infty}^{\infty}\left\{\mathcal{F}[x(t)] \;\; exp(-i\omega\tau)\, exp\left[-\frac{(\omega s - \omega_0)^2}{2}\right]\right\} exp(i\omega t)\;\; d\omega$$

$$= \int_{-\infty}^{\infty}\left\{X(\omega)\;\; exp(-i\omega\tau)\, exp\left[-\frac{(\omega s - \omega_0)^2}{2}\right]\right\} exp(i\omega t)\;\; d\omega$$

Morlet wavelet: summary

Mother: $\phi_{0,1}(t) = \phi(t) = \left\{\frac{1}{\sqrt{2\pi}}\;\; exp\left(-\frac{t^2}{2}\right)\right\}\;\; exp(-i\omega_0 t)$

Daughters: $\phi_{\tau,s}(t) = \left\{\frac{1}{s\sqrt{2\pi}}\; exp\left[-\frac{1}{2}\left(\frac{t-\tau}{s}\right)^2\right]\right\}\;\; exp\left[-i\omega_0\left(\frac{t-\tau}{s}\right)\right]$

Where:

τ is the translation (or shift in time)

$s = \frac{n}{\omega_0}$ is the scale that depends on the integer n. Strictly, n is the scaling parameter, not s. It is the number of waves or cycles in the wavelet.

ω_0 is the frequency of the wavelet (i.e. from the trigonometric component of the Morlet wavelet definition).

So, the steps for obtaining the wavelet spectrum now become:

1. Take the Fourier transform (of the original time series).
2. This only needs to be done once.
3. Select a value of τ and s.
4. Multiply the functions inside the curly brackets.
5. Repeat steps 3 and 4.

6. Plot scale (or our preference is the Fourier period, which, for the Morlet wavelet, is almost the same as the scale) and wavelet energy (analogous to spectral energy) as a function of time.

The window function that is the Morlet transform needs to obey a certain set of rules. Without going into the mathematical details, some of these rules for the window function include:

(a) It is of finite length. We ensure the window is of finite length by choosing a function such that it tapers to zero at its ends. The window is then said to have close 'compact support'. (Strictly speaking, the Morlet wavelet is of infinite length. However, because of the Gaussian part of the function, it rapidly approaches zero with increasing distance from the centre of the wavelet. Therefore, for practical purposes, the length of the Morlet wavelet can be regarded as finite).
(b) The total energy must be zero. This is achieved by ensuring that the area under the curve that forms the window is zero. The Morlet wavelet automatically satisfies this condition (i.e. the area under a sine / cosine curve is zero).
(c) The windows must form a set of orthonormal bases. This is basically a mathematical space in which the unit vectors are all mutually orthogonal (or perpendicular). Again, the Morlet wavelet satisfies this condition.
(d) To be admissible as a wavelet, the following needs to apply:

$$\int_{-\infty}^{\infty} \frac{\{\mathcal{F}[\phi(t)]\}^2}{\omega}\, d\omega \quad < \infty$$

Cohen (2019) states that, provided the wavelet tapers to zero at both ends and that the area under the wavelet curve is zero, then the conditions above will be (approximately) met.

9.1.1. Other definitions of the Morlet wavelet

As with the Fourier transform (see Chapter 2), there are multiple definitions of the Morlet wavelet transform. Some definitions of the Morlet mother wavelet are provided below (note: we've transformed these definitions to the symbols used above).

Emery and Thomson (1997):

$$\phi(t) = e^{i\omega_o t}\, e^{-\frac{1}{2}t^2}$$

Torrence and Compo (1998), Grinsted et al (2004):

$$\phi(t) = \frac{1}{\sqrt[4]{\pi}} e^{i\omega_o t}\, e^{-\frac{1}{2}t^2}$$

Kelly, et al. (2003),

$$\phi(t) = \frac{1}{\sqrt[4]{\pi}} e^{-i\omega_o t}\, e^{-\frac{1}{2}t^2}$$

Polikar (2006), Cohen (2019):

$$\phi(t) = e^{i\omega_o t}\, e^{-\frac{1}{2}t^2}$$

Ashmead (2010):

$$\phi(t) = \left(e^{-i\omega_o t} - e^{-\frac{\omega_0^2}{2}}\right) e^{-\frac{1}{2}t^2}$$

Generally, these definitions differ only be the leading constant and they have no practical impact on the overall result. Also, some authors use $e^{i\omega_o t}$ while others use $e^{-i\omega_o t}$. The only difference that this makes is in the wavelet phase spectrum, which will differ by 180°.

As with using the Fourier transform, the Morlet wavelet definition should be specified when presenting any results. Of these definitions, only that of Ashmead (2010) can be formally regarded as a Morlet wavelet. However, provided ω_o is sufficiently large, the 'correction' term $e^{-\frac{\omega_0^2}{2}} \rightarrow 0$ and can be effectively ignored. For example, if $\omega_o = 3$ then $e^{-\frac{\omega_0^2}{2}} \sim 0.01$.

The wavelet examples that are used below are based on the Torrence and Compo (1998) and Grinsted et al (2004) definition of the Morlet wavelet. In these papers, $\omega_o = 6$ and $e^{-\frac{\omega_0^2}{2}} \sim 10^{-8}$, which is sufficiently close to zero to ignore the correction term used by Ashmead (2010).

9.1.2. Admissibility

For the Morlet wavelet, the admissibility condition reduces to: $\mathcal{F}[\phi(t)]_{\omega=0} = 0$. If we go through the mathematics, we find that the Morlet wavelet (as defined in the box above) does not strictly satisfy the admissibility condition. The definition needs to be that used by Ashmead (2010), where the mother and daughter Morlet wavelets are:

Mother: $$\phi(t) = \left[e^{-i\omega_0 t} - e^{-\omega_0^2/2}\right] e^{-t^2/2}$$

Daughters: $$\phi_{\tau,s}(t) = \frac{1}{\sqrt{|s|}}\left[e^{-i\omega_0\left(\frac{t-\tau}{s}\right)} - e^{-\omega_0^2/2}\right] e^{-\frac{1}{2}\left(\frac{t-\tau}{s}\right)^2}$$

The primary difference between Ashmead's formulation and the others above, is the presence of the 'correction term', $-e^{-\omega_0^2/2}$. It enables the Morlet wavelet to formally meet the admissibility condition. Provided ω_0 $(= {}^{n}/_{s})$ is large-ish (about 6 or more), then $e^{-\omega_0^2/2}$ is small and can be effectively ignored.

Formally:

$$\mathcal{F}[\phi(t)] = \int_{-\infty}^{\infty} \left\{\left[e^{-i\omega_0 t} - e^{-\omega_0^2/2}\right] e^{-t^2/2}\right\} e^{-i\omega t}\, dt$$

$$= \int_{-\infty}^{\infty} \left[e^{-i\omega_0 t - \frac{1}{2}t^2 - i\omega t} - e^{-\frac{1}{2}\omega_0^2 - \frac{1}{2}t^2 - i\omega t}\right] dt$$

$$= \int_{-\infty}^{\infty} \left[e^{-i\omega_0 t - \frac{1}{2}t^2 - i\omega t}\right] dt - \int_{-\infty}^{\infty} \left[e^{-\frac{1}{2}\omega_0^2 - \frac{1}{2}t^2 - i\omega t}\right] dt$$

$$= \int_{-\infty}^{\infty} e^{-\left(\frac{t}{\sqrt{2}}+i\frac{\omega_0+\omega}{\sqrt{2}}\right)^2} e^{-\frac{(\omega_0+\omega)^2}{2}}\ dt\ -\ \int_{-\infty}^{\infty} e^{-\left(\frac{t}{\sqrt{2}}+i\frac{\omega}{\sqrt{2}}\right)^2} e^{-\left(\frac{\omega_o^2}{2}+\frac{\omega^2}{2}\right)}\ dt$$

$$= e^{-\frac{(\omega_0+\omega)^2}{2}} \int_{-\infty}^{\infty} e^{-\left(\frac{t}{\sqrt{2}}+i\frac{\omega_0+\omega}{\sqrt{2}}\right)^2}\ dt\ - e^{-\left(\frac{\omega_o^2}{2}+\frac{\omega^2}{2}\right)} \int_{-\infty}^{\infty} e^{-\left(\frac{t}{\sqrt{2}}+i\frac{\omega}{\sqrt{2}}\right)^2}\ dt$$

Let $X = \frac{t}{\sqrt{2}} + i\frac{\omega_0+\omega}{\sqrt{2}}$, then $dt = X\sqrt{2}$ and let $Y = \frac{t}{\sqrt{2}} + i\frac{\omega}{\sqrt{2}}$, then $dt = Y\sqrt{2}$. Substituting into the respective expressions above we get:

$$\mathcal{F}[\phi(t)] = e^{-\frac{(\omega_0+\omega)^2}{2}} \int_{-\infty}^{\infty} e^{-(X)^2}\sqrt{2}\ dX\ - e^{-\left(\frac{\omega_o^2}{2}+\frac{\omega^2}{2}\right)} \int_{-\infty}^{\infty} e^{-(Y)^2}\sqrt{2}\ dY$$

$$= \sqrt{2\pi}\ e^{-\frac{(\omega_0+\omega)^2}{2}} - \sqrt{2\pi}\ e^{-\left(\frac{\omega_o^2}{2}+\frac{\omega^2}{2}\right)}$$

$$= \sqrt{2\pi}\left\{ e^{-\frac{(\omega_0+\omega)^2}{2}} - e^{-\left(\frac{\omega_o^2}{2}+\frac{\omega^2}{2}\right)} \right\}$$

$$= \sqrt{2\pi}\left\{ e^{-\left(\frac{\omega_0^2+2\,\omega_o\,\omega+\omega^2}{2}\right)} - e^{-\left(\frac{\omega_o^2}{2}+\frac{\omega^2}{2}\right)} \right\}$$

$$= \sqrt{2\pi}\left\{ e^{-\left(\frac{\omega_0^2+\omega^2}{2}\right)} e^{-\omega_o\,\omega} - e^{-\left(\frac{\omega_o^2}{2}+\frac{\omega^2}{2}\right)} \right\}$$

$$= \sqrt{2\pi}\ e^{-\left(\frac{\omega_o^2}{2}+\frac{\omega^2}{2}\right)}\{e^{-\omega_o\,\omega} - 1\}$$

$$\mathcal{F}[\phi(t)]_{\omega=0} = \sqrt{2\pi}\ e^{-\left(\frac{\omega_o^2}{2}+\frac{0^2}{2}\right)}\{e^{-\omega_o\,0} - 1\}$$

$$= \sqrt{2\pi}\ e^{-\left(\frac{\omega_o^2}{2}\right)}\ \{1 - 1\} \quad = \quad 0$$

Hence, admissibility is satisfied.

9.1.3. Translation and scale

The Morlet wavelet is complex and depends on two parameters, translation (τ) and scale (s).

Translation simply moves the window through the time series (or equivalently, moves the time series past the window). Its effect is shown in Figure 9–1 for three values of the translation. (Note, only the real component of the Morlet wavelet is shown). Three translations are shown: $\tau = 0$, $\tau = 1$ and $\tau = 5$ (time units). The scale of all three plots is $n = 10$. The translation operation moves the wavelet such that it is centred on a new time. For the same scale, the shape of the wavelet remains identical for all translations.

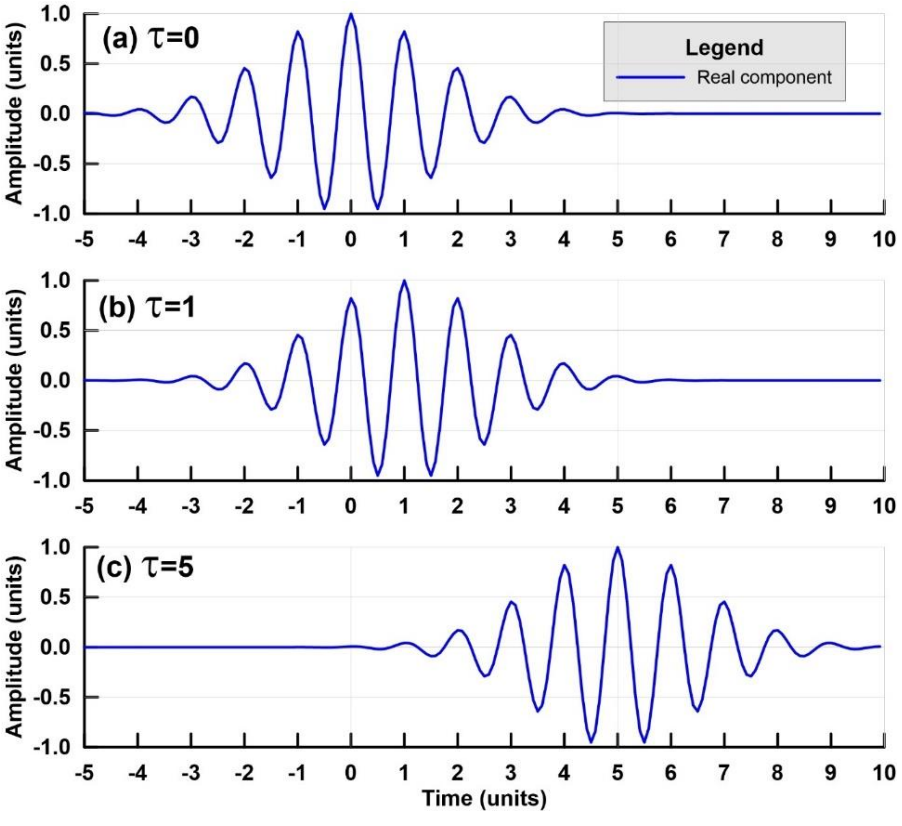


Figure 9–1. The Morlet wavelet (real component only): the effect of different translations.

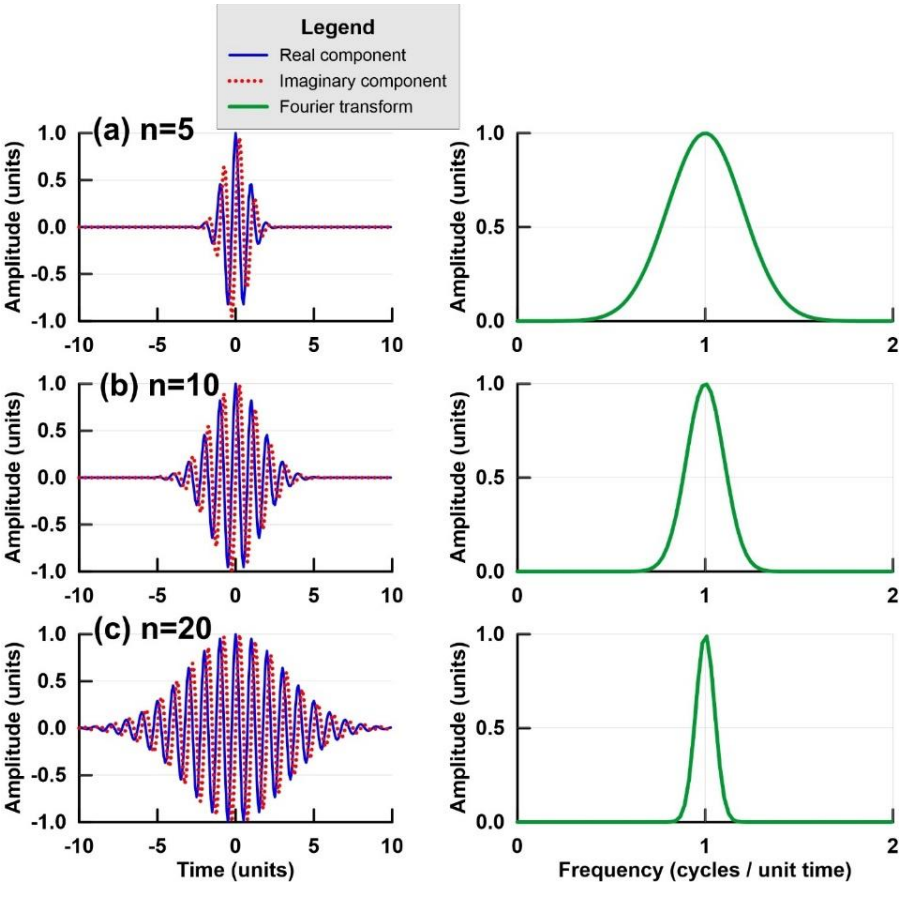


Figure 9–2. The Morlet wavelet: the effect of different scales and their corresponding Fourier transforms.

Scale is effectively the width of the window – or the number of cycles in the wavelet. It also requires us to specify the frequency of these cycles (ω_o). The effect of the scale is demonstrated in Figure 9–2 for three values corresponding to $n = 5, n = 10$ and $n = 20$ (cycles in the wavelet). (The time, amplitude and frequency units are arbitrary). This plot also shows the Fourier transforms of the corresponding wavelets. Now, we might think that having the smallest possible scale (i.e. small n) is the way to go. However, as we reduce the scale and achieve better temporal resolution, we lose frequency resolution i.e. the corresponding plots of

their Fourier transforms becomes broader (Figure 9–2). A judgement needs to be made in selecting n. Often, $n = 7$ is used.

It is generally more convenient to use the Fourier period rather than the scale. For the Morlet wavelet, the Fourier period is about 1.03 times the scale (Torrence and Compo, 1998). There is little difference (3%) in the magnitude of the scale and the Fourier period. Here, it is assumed that the scale and the Fourier period are the same.

9.2. Examples of the Morlet wavelet

The following examples all use the Morlet wavelet definitions of Torrence and Compo (1998) and Grinsted et al (2004). These papers both provide references to Matlab scripts for the Morlet wavelet that are used in these examples.

In each of the wavelet spectral plots there is a red shaded area and contours that are filled in black. The former represents regions in which the boundaries influence the wavelet spectrum (Section 8.2, Figure 8–15). Contours that lie is this red zone will be influenced by the boundary and interpretation of those contours should be treated with extreme caution. Contours that are filled in black lie above the 95% confidence limits i.e. it is unlikely that the results could have been generated by random processes.

To calculate the 95% confidence limits Torrence and Compo (1998) assume a background spectrum. This is based on a Markov process in which a white noise (i.e. random) spectrum was coupled with a red noise spectrum (i.e. in which the spectral power decreases as frequency increases – analogous to the -5/3rds power law). One hundred thousand background spectra were generated from which 95% confidence limits were determined for each scale.

A similar approach could have been used to determine the confidence limits for the Fourier (spectral) analyses (Section 4.2). However, the simpler band averaging approach was chosen.

The wavelet scale is presented in these examples as the (Fourier) period. To satisfy the admissibility condition, Torrence and Compo (1998) choose $\omega_0 = 6$. In so doing, the wavelet scale closely matches the Fourier period. Hence the axis that is labelled the 'Period' in the wavelet plots (approximately) corresponds to the wavelet scale. The difference is about 3%.

9.2.1. Example 8-9 – Southern Oscillation Index

The Southern Oscillation Index (SOI) is a parameter that is used to estimate if an El Nino or La Nina phenomenon is likely to occur. It is based on the monthly average difference in atmospheric pressure between Darwin and Tahiti. Monthly SOI data between 1935 and 2020 are used in this example.

Questions: At what time scales (periods) and at what times do peaks in the wavelet spectrum exceed the 95% confidence limit?

Data used:

- Monthly Southern Oscillation Index data between 1935 and 2020 – a total of 1,024 data points (Figure 9–3).
- The Nyquist period is 2 months (Nyquist frequency ½ cycle per month).
- The data set is about 85 years long, therefore the longest period features we can realistically resolve is about 10 years (i.e. we want about 5-10 cycles in the data set before we are confident of interpreting the low frequency end of the wavelet spectrum).

Pre-processing:

- The linear trend is removed from the data to give SOI anomalies.
- No other data smoothing or signal removal is undertaken.

Testing:

- Results from the Dickey-Fuller test for stationarity indicate a Tau value of - 15.00, while the Tau critical value is -2.86. As Tau

calculated is less than Tau critical, we reject the null hypothesis that the data are not stationary.

Smoothing:

- Confidence limits for the wavelet spectra were generated using Monte Carlo techniques (see Torrence and Compo, 1998).

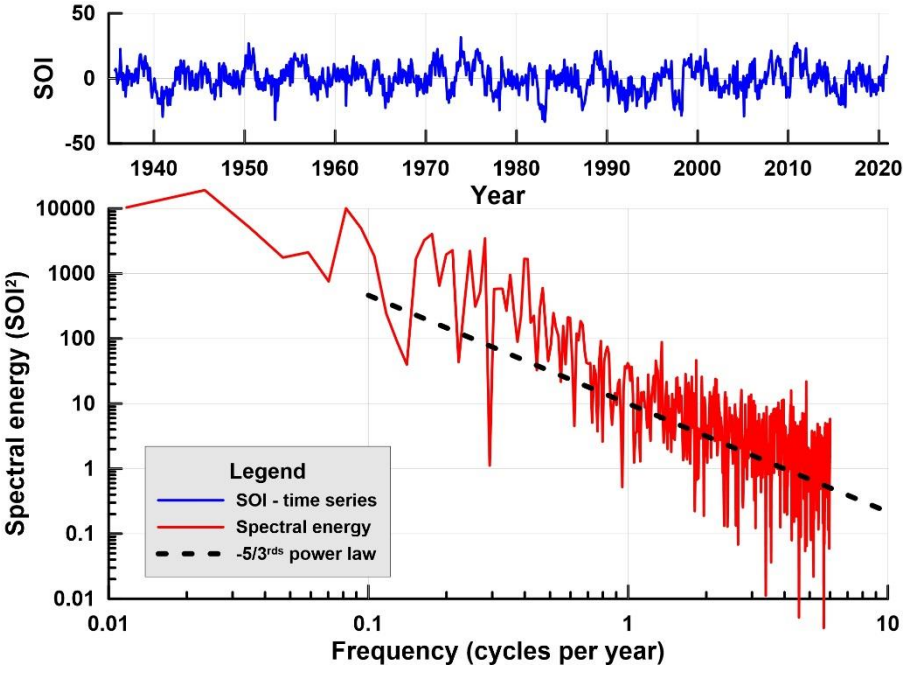


Figure 9–3. Top: Time series of the monthly SOI anomalies between 1935 and 2020. Bottom: corresponding spectral analysis of the SOI anomalies.

Description:

- The time series and Fourier spectrum of the SOI anomalies are presented in Figure 9–3.
- Spectral energy decreases (approximately) according to the -5/3^{rds} power law.

- There are no peaks in the spectrum that might be associated with signals of particular importance.
- Based on the evidence provided by the spectral analysis, we might conclude that there are no signals of importance in these data.

Using the Torrence and Compo (1998) Matlab software, the Morlet wavelet was used to estimate the wavelet energy spectrum on the monthly SOI anomalies. Results are presented in Figure 9–4 where the wavelet energy is a function of time and period (or scale). In Figure 9–4, areas shaded with red represent regions of the wavelet spectrum that are influenced by the boundaries. Therefore, contours lying in the red region should be ignored. Contours that are filled with black exceed the 95% confidence limits. Such areas indicate the period at which a particular signal is dominant. 60 scales (i.e. 60 values of '*s*') were used in the analysis, ranging from 2 months to 30 years.

Description:

- Examination of Figure 9–4 identifies three regions of significant wavelet spectral energy that are not apparent from the spectral analysis.
- At periods less than about 0.5 years there is a large amount of localised high wavelet energy activity. These features are short lived and occur throughout the entire data set.
- A second region, with periods between about 2 and 8 years, is apparent over most of the time span. With the exception of the periods 1960-1965 and 2000-2005, the wavelet spectral energy lies above the 95% confidence limit. There is some suggestion that the period is decreasing over time, although it is difficult to determine if this is a trend or part of a very long term cycle (circa >150 years). The El Nino / La Nina cycle lies in this period and the evidence (provided by the decreasing temporal trend) suggests that this cycle is becoming more frequent as time progresses.
- The third region occurs between periods of 8 and 16 years. The wavelet spectral energy lies above the 95% confidence limit from about 1970 onwards.

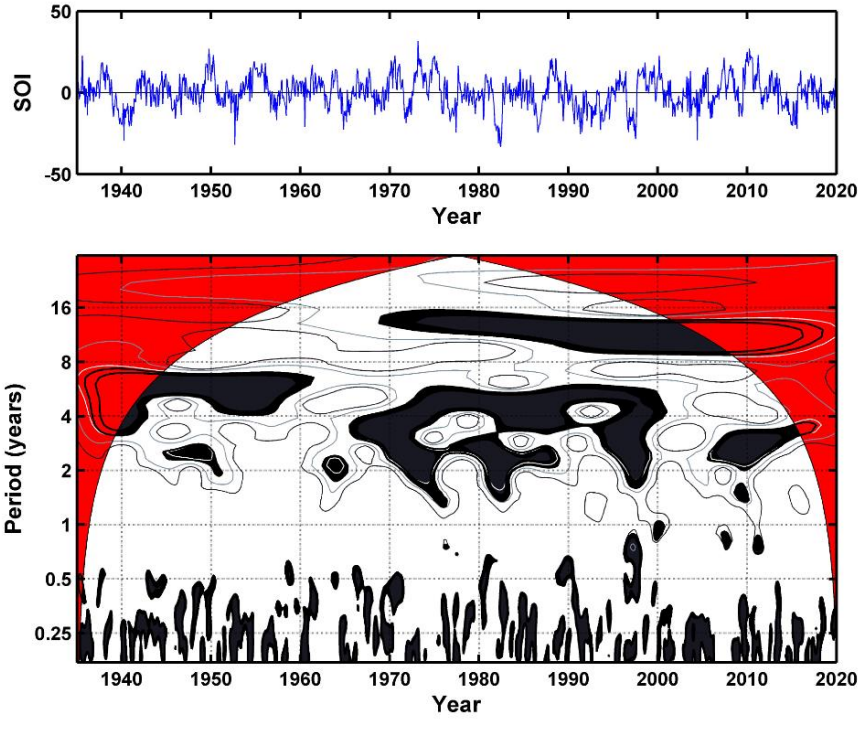


Figure 9–4. Top: Time series of the monthly SOI anomalies between 1935 and 2020. Bottom: corresponding Morlet wavelet of the SOI anomalies. Areas shaded in red are influenced by the boundaries in the data, while areas coloured black exceed the 95% confidence limit.

Returning to the question we set out to answer:

(a) The dominant, consistent signal occurs in the 4-8 year period. From about 1970 onwards, a long period feature starts to become increasingly important, imparting more energy into the system. The signal in the 4-8 year period responds to the increased long period feature by moving into a shorter period band (viz. 2-4 years). In so doing, the system (represented here by the SOI cycle) can more effectively dissipate energy. One consequence is that, as time progresses, the El Nino / La Nina cycle occurs more

frequently – periods of low and high rainfall along the Australian east coast occur more frequently.

9.2.2. Example 8-10 – High frequency internal waves

When the water column becomes stratified, internal waves may propagate along the density interface resulting from the stratification. Here we examine high frequency internal waves on 14 June 2019, which we assume can be defined by the change in the depth of the 20°C isotherm (data from Example 8-8). (The depth of the 20°C isotherm is used as a surrogate for the more appropriate parameter – the depth at which the Brunt-Väisälä frequency is a maximum).

Questions: At what scales (periods) and at what times do peaks in the wavelet spectrum exceed the 95% confidence limit?

Data used:

- 5-minute water temperature data from 14 June 2019.
- A total of 256 data points.
- The Nyquist period is 10 minutes (Nyquist frequency is 1/10 cycles per minute or 6 cycles per hour).
- The data set is 21.25 hours long, therefore the longest period features we can realistically resolve is about 2-4 hours (i.e. we want about 5-10 cycles in the data set before we are confident of interpreting the low frequency end of the wavelet spectrum).

Pre-processing:

- The linear trend is removed from the data to give anomalies of the depth of the 20°C isotherm.
- No other data smoothing or signal removal is undertaken.

Testing:

- The Dickey-Fuller test for stationarity resulted in a Tau value of -5.57. The Tau critical value is -2.87. As Tau calculated is less than Tau critical, we reject the null hypothesis that the data are not

stationary. We find this result interesting as the time series plot (Figure 9–5) appears to indicate a non-stationary time series.

Smoothing:

- No smoothing of the wavelet energy spectrum was undertaken.
- Confidence limits for the wavelet spectra were generated using Monte Carlo techniques (see Torrence and Compo, 1998).

Description:

- A time series plot of the depth of the 20°C isotherm anomalies is given in Figure 9–5, together with the corresponding Fourier spectrum. For clarity of the plot, no spectral confidence limits are presented.
- The time series plot clearly shows periods where the depth of the anomalies changes rapidly (likely to be associated with internal waves) and periods where there is very little change in the depth.
- The spectrum is essentially featureless – there are no spectral peaks of note. The spectrum approximately follows the -5/3^{rds} power law.
- However, there are two small elevations in spectral energy near frequencies of 3-4 cycles per hour. Although these elevations lie below the 95% confidence limits (not shown here), their presence may suggest a process of importance that is transitory.

The analysis is repeated using the Morlet wavelet. Again, the wavelet package of Torrence and Compo (1998) is used. Results are presented in Figure 9–6. Areas shaded with red on Figure 9–6 are influenced by the boundary, while contours shaded in grey lie above the 95% confidence limit – they have high wavelet spectral energy.

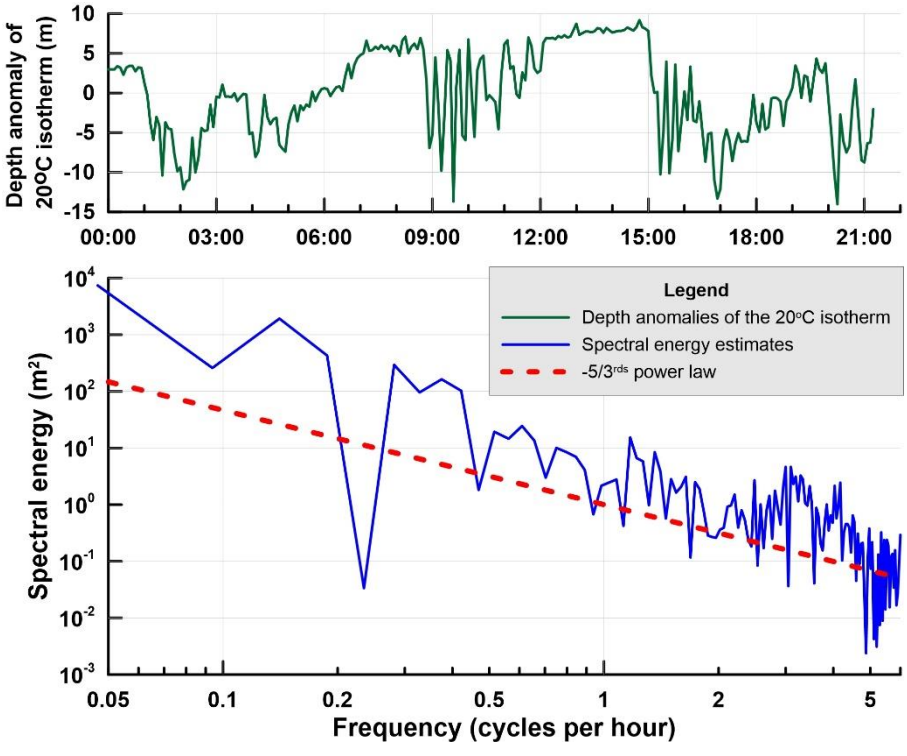


Figure 9–5. Top: Time series plot of the depth of the 20°C isotherm anomalies on 14 June 2019. Bottom: Spectral analysis of the depth of the 20°C isotherm anomalies.

Returning to the question we set out to answer:

(a) Three regions of high wavelet spectral energy are observed: near 09;00-11:00, 15:00-17:00 and 20:00-21:00. The period is between about 0.167 hours and 0.45 hours (10 and 27 minutes, respectively). These regions visually align with the high frequency activity identified in the time series plot. Near 11:00 on the wavelet spectral plot, a small, but significant region of internal wave activity is also identified. The corresponding period is about 0.6-0.7 hours (36 to 42 minutes). From the time series plot, it can be clearly seen that the period of the internal wave has increased,

compared with the internal wave packet observed between 09:00 and 11:00.

The wavelet analysis clearly identifies when the high frequency internal waves occur.

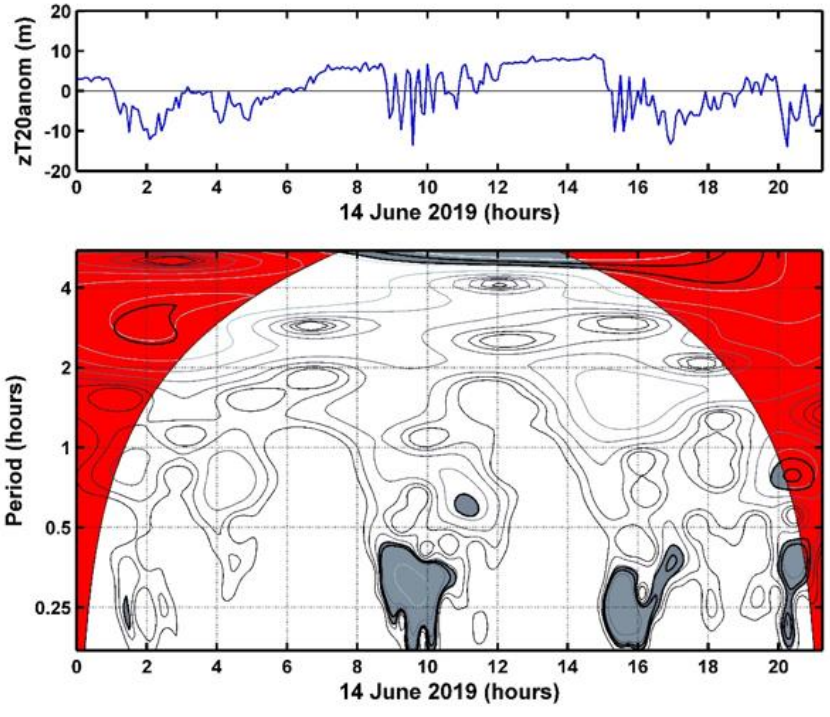


Figure 9–6. Morlet wavelet results. Top: time series plot of the depth of the 20°C isotherm anomalies on 14 June 2019. Bottom: Wavelet spectrum of the depth of the 20°C isotherm anomalies. Areas shaded in red are influenced by the boundaries in the data, while areas coloured black exceed the 95% confidence limit.

9.2.3. Example 8-11 – Internal waves – extended time period

The previous example examined high frequency internal waves over a short time period (circa 24 hours). In this example, high frequency internal waves are examined over an extended period of time, about 340 hours.

Question: Can wavelet analysis clearly detect high frequency internal waves (with periods less than about 1 hour) when a long time span is used?

Data used:

- 5-minute data of the depth of the 20°C isotherm.
- The total time spans about 340 hours (4,096 records).
- Water temperature data were collected from a thermistor string comprising 14 thermistors spaced at approximately 4m.
- Linear interpolation was used to estimate the depth of the 20°C isotherm – a surrogate for internal waves.

Pre-processing:

- A linear trend was removed from the data prior to analysis.

Testing:

- The Dickey-Fuller test indicated that the anomalies formed a stationary data set. Tau calculated -11.97 and Tau critical -2.86. As Tau calculated is less than Tau critical, we reject the null hypothesis that the data are not stationary.

Smoothing:

- No smoothing of the wavelet energy spectrum was undertaken.
- Confidence limits for the wavelet spectra were generated using Monte Carlo techniques (see Torrence and Compo, 1998).

Description of results:

- The Morlet wavelet was used to estimate the wavelet energy spectrum using the approach (and accompanying Matlab

software) outlined in Torrence and Compo (1998). Results are presented in Figure 9–7. Areas shaded in red are affected by the boundaries and interpretation of the results within should be treated with extreme caution (i.e. we should ignore this region of the wavelet spectrum). Areas shaded in black lie above the 95% confidence limit.

- In general, the wavelet spectrum is significant for periods exceeding 16 hours across all times. There are several regions where the wavelet spectrum lies above the 95% confidence limit for periods between 4 and 16 hours.
- At higher frequencies (i.e. low period), there are several short-lived features where the energy in the wavelet spectrum lies above the 95% confidence limit.

Returning to the question we set out to answer:

(a) At the high frequency (i.e. low period) end of the wavelet spectrum, there are several short-lived features where the wavelet energy lies above the 95% confidence limit. These occur near times of 60, 150, 200, 275 and 290 hours. The period of such features is generally between about 0.25 and 0.5 hours and are likely to be associated with high frequency high internal wave activity. Such activity can be identified in the original time series plot (although the time scale of Figure 9–7 makes it a little difficult to identify).

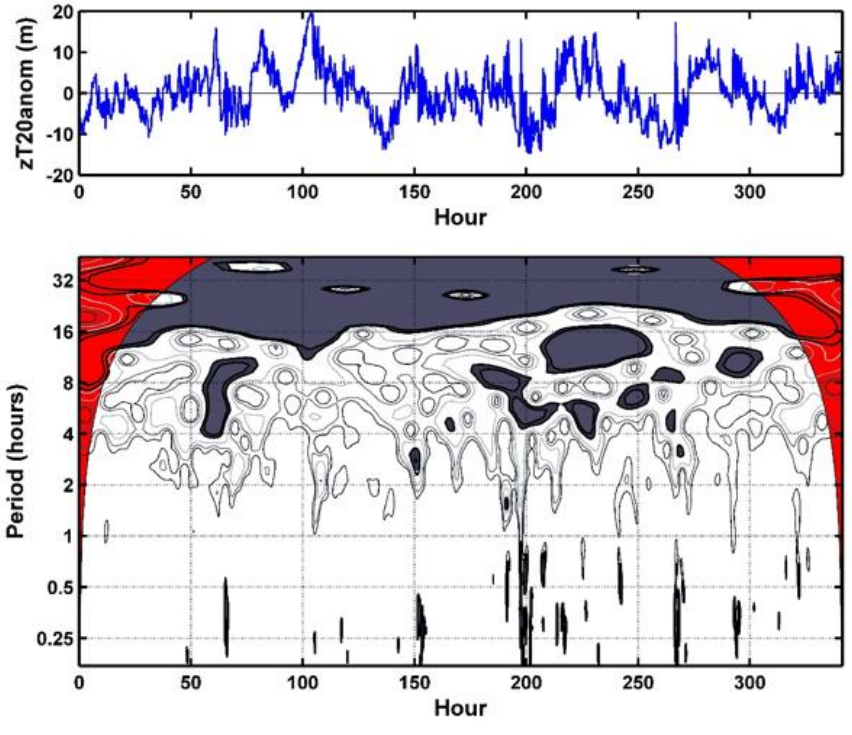


Figure 9–7. Morlet wavelet results. Top: time series plot of the depth of the 20°C isotherm anomalies over 340 hours. Bottom: Wavelet spectrum of the depth of the 20°C isotherm anomalies over 340 hours. Areas shaded in red are affected by the boundaries. Areas shaded in black lie above the 95% confidence limit.

As the question centred around high frequency internal waves, we could smooth the time series data the remove the low frequency features and reanalyse the data, focusing on the higher frequencies. A cosine-Lanczos filter was generated to remove the features with periods exceeding about 2 hours. Results using the Morlet wavelet are presented in Figure 9–8. Removing the low frequency features result in all of the spectral energy lying in the periods of interest. It then becomes virtually impossible to identify at which times the high frequency internal waves are important.

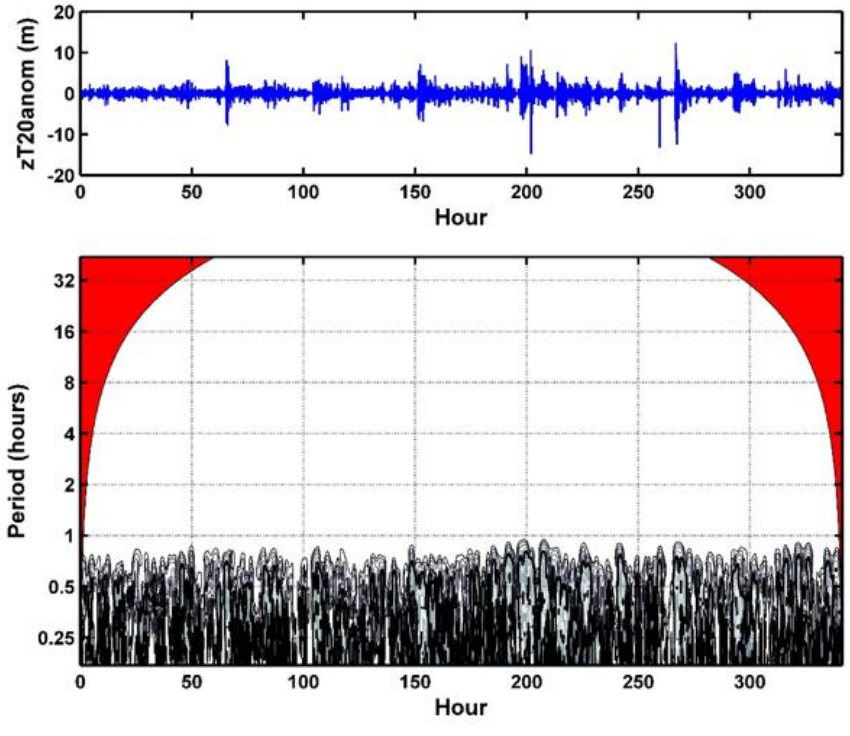


Figure 9–8. Morlet wavelet results. Top: Time series plot of the depth of the 20°C isotherm anomalies over 340 hours, filtered using a cosine-Lanczos filter. Bottom: Wavelet spectrum of the depth of the 20°C isotherm anomalies over 340 hours. Areas shaded in red are affected by the boundaries. Areas shaded in black lie above the 95% confidence limit.

So, if high frequency features are of interest, the advice is to use the unsmoothed data to generate an initial wavelet spectrum. This is, perhaps, a little counter intuitive. This should identify when high frequency features occur e.g. as observed in Figure 9–7. Then select the times of interest and undertake a wavelet analysis using the shorter data sets e.g. as observed in Figure 9–6.

9.3. The Morlet wavelet cross spectrum

In much the same way that the Fourier cross spectrum can be used to examine the relationship between two data sets, a wavelet cross spectrum can be generated. As might be expected, the wavelet cross spectrum has many parallels to cross spectral analysis (see Section Chapter 5). The major difference arises because the wavelet cross spectrum is a two parameter system, while the cross spectrum is a single parameter system.

The Fourier cross spectrum of two time series generated two auto spectra (or one cross amplitude spectrum), a coherence squared spectrum and a phase difference spectrum. Similarly, the wavelet cross spectrum results in two auto wavelet spectra (or one wavelet cross spectrum), a wavelet coherence squared spectrum and a wavelet phase difference spectrum.

In Section 5.1 the cross spectrum was defined in terms of the co and quad spectra which was also shown to be the product of the first Fourier transform with the complex conjugate of the second Fourier transform.

$$\mathcal{F}_{xy(p)} = CO_p - i\,QU_p = \mathcal{F}_{x(p)}\mathcal{F}^*_{y(p)}$$

Let $\phi_{x(\tau,s)} = (\phi_x)$ and $\phi_{y(\tau,s)} = (\phi_y)$ be the Morlet wavelet transforms of the two time series $x(t)$ and $y(t)$. Then, by analogy with the Fourier cross spectrum, the wavelet cross spectrum is given by:

$$\phi_{xy} = \phi_x\,\phi_y^*$$

The wavelet cross spectrum, coherence spectrum and phase difference spectrum presented here follows that of Grinsted et al (2004). (Throughout this section, the term 'wavelet' explicitly assumes that the Morlet wavelet is being used). In turn, the Grinsted et al paper builds on the wavelet approach and definition used in Torrence and Compo (1998), where the Morlet mother wavelet is defined as:

$$\phi(t) = \frac{1}{\sqrt[4]{\pi}}\, e^{-\frac{1}{2}t^2}\, e^{i\omega_o t}$$

The Morlet daughter wavelets are then:

$$\phi_{\tau,s}(t) = \frac{1}{s\sqrt[4]{\pi}} \; exp\left[-\frac{1}{2}\left(\frac{t-\tau}{s}\right)^2\right] \; exp\left[-i\omega_0\left(\frac{t-\tau}{s}\right)\right]$$

Grinsted et al (2004) choose $\omega_o = 6$ as it provides a 'good balance between time and frequency localization'.

For convenience, we break each wavelet into its real and imaginary part (noting that we have a real and imaginary value for each of the τ, s parameters):

$$\phi_x = a + ib \qquad and \qquad \phi_y = c + id$$

$$\phi_{xy} = \phi_x \, \phi_y^* = (ac + bd) - i(ad - bc)$$

Apart from the wavelet cross spectrum, two important functions are the coherence and the phase difference i.e. the 'correlation' between the two data sets and whether one data set lags (or leads) the second data set. The wavelet coherence and phase difference spectra are given by:

Wavelet coherence squared spectrum:

$$Coh^2 = \frac{\left|\phi_{xy}\right|^2}{|\phi_x|^2 \;\; \left|\phi_y\right|^2} = \frac{|(ac + bd) - i(ad - bc)|^2}{|a + ib|^2 \;\; |c + id|^2}$$

Wavelet phase difference spectrum:

$$\psi = tan^{-1}\left[\frac{Im(\phi_{xy})}{Re(\phi_{xy})}\right] = tan^{-1}\left[\frac{-(ad - bc)}{(ac + bd)}\right]$$

As we saw with the Fourier coherence squared, the above expression for the wavelet coherence squared spectrum gives the trivial result of unity. So, to obtain non trivial results, we need to do some averaging. In Fourier space this was easy – there was only one parameter, p, and we averaged the Fourier spectrum in p-space (or frequency). However, wavelet analysis results in two parameters τ and s and the question is raised over which parameter do we average? The short answer is that we average over both parameters.

Note: Grinsted et al (2004) average over 'time and scale'. This is broadly equivalent to averaging over τ (translation or time) and s (scale or frequency).

Therefore, the (averaged) wavelet coherence squared spectrum is:

$$Coh^2 = \frac{\left|\mathcal{S}\left[s^{-1}\,\phi_{xy}\right]\right|^2}{\mathcal{S}[s^{-1}\,|\phi_x|^2]\;\;\mathcal{S}\left[s^{-1}\,|\phi_y|^2\right]}$$

Where $\mathcal{S}$ (i.e. the scripted '$\mathcal{S}$') is a smoothing operator given by:

$$\mathcal{S}[\phi] \;=\; \mathcal{S}_{scale}\{\mathcal{S}_{time}[\phi(s)]\}$$

And s (i.e. small 's') is the wavelet scale.

$\mathcal{S}_{scale}$ indicates smoothing along the wavelet scale axis, and

$\mathcal{S}_{time}$ indicates smoothing along the time axis.

And the (average) wavelet phase difference spectrum is:

$$\psi = \; tan^{-1}\left[\frac{-(ad-bc)}{(ac+bd)}\right]$$

The results are easier to interpret if the scale is converted to period. For the Morlet wavelet, the scale and the Fourier period are essentially the same, differing only by 3%.

9.4. Examples of the Morlet cross wavelet

9.4.1. Example 8-12 – Across shelf currents at two depths

The first example of the wavelet cross spectrum examines the across shelf component of the currents at two depths in the water column. The across shelf component is approximated by the east-west component of the currents (i.e. the coastline runs approximately north-south).

Question: How does the coherence of the across shelf component of the currents at two depths change over time?

Data used:

- The east-west (approximately across shelf) currents at two depths in the water column.
- Data were collected from a location approximately 3 km offshore in waters 65 m deep.
- The data time step was 5 minutes and a total of 1,024 data points were used (about 3.5 days long), Figure 9–9.

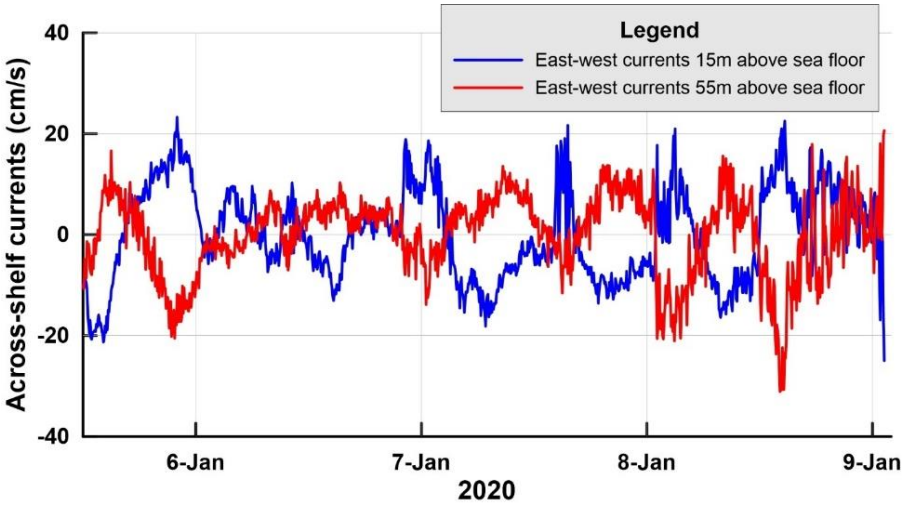


Figure 9–9. Time series plot of the east-west components of the currents 15 m and 55 m above the sea floor.

As the location is relatively close to shore, we might expect the across shelf currents at the different depths to move in opposite directions to (approximately) maintain mass balance in the system. Further, internal waves are known to be present in this region and internal wave theory indicates that currents above and below the internal wave will move in opposite directions. Therefore, we might expect that the coherence between these two data sets will be 'out of phase' i.e. onshore at one depth and offshore at the other depth and vice versa. An examination of the time series data (Figure 9–9) indicates exactly this feature.

Pre-processing:

- A linear trend was removed from each data set.
- No further pre-processing of the data was undertaken prior to the analysis. (In this particular example, the removal of the linear trend had little effect on the results).

Testing:

- The Dickey-Fuller test was undertaken on both data sets. The critical value was -2.86, the Tau-statistic for the upper layer currents was -5.87, while that for the lower currents was -7.95.
- As both Tau statistics lie below the critical value, we reject the null hypothesis that the time series are not stationary.

Smoothing:

- Confidence limits for the wavelet spectra were generated using Monte Carlo techniques (see Torrence and Compo, 1998 and Grinsted et al, 2004).

Description of results:

- Both wavelet auto spectra show essentially the same results.
- From time 35 hours onwards, near the Fourier period of 15 hours, there is a band of high wavelet energy.
- At Fourier periods between about 0.33 and 1 hour, high wavelet energy is observed near times 50, 60 and 80 hours.
- Similar features are observed in the cross wavelet spectrum. The phase difference arrows generally point to the left, indicating that the same processes are out of phase. The same processes are observed in both data sets, although they occur in opposite directions (i.e. a positive value in one data set corresponds to a negative value in the other data set, and vice versa).

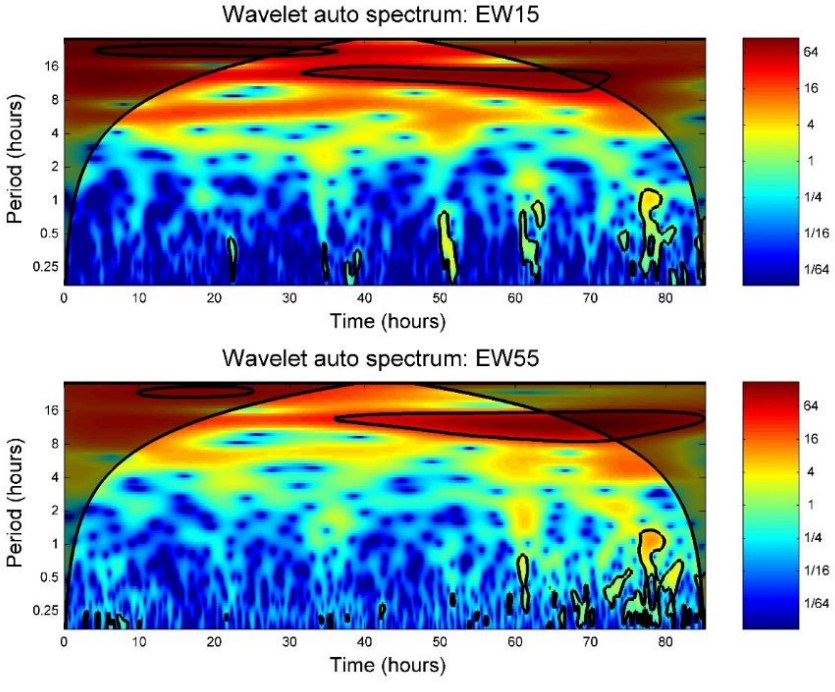


Figure 9–10. Wavelet auto spectra for each data: top plot EW55 (55 m above the sea floor), bottom plot EW15 (15 m above the sea floor). The shaded black areas indicate influences of the boundaries. The heavy black lines indicate the 95% confidence limit.

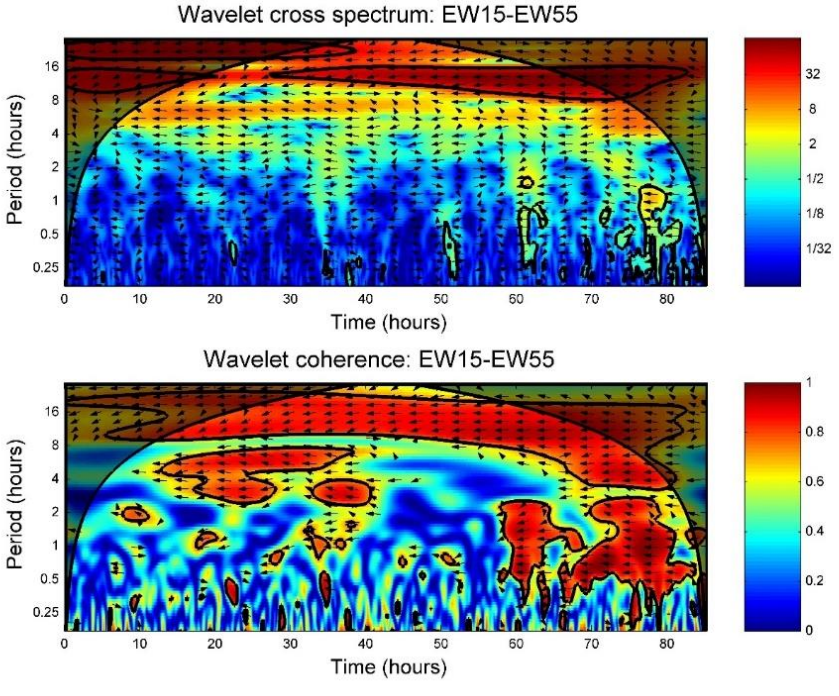


Figure 9–11. Top: Wavelet cross spectrum between EW55 and EW15. Bottom: wavelet coherence squared (indicated by the coloured contours) and phase difference (indicated by the arrows) spectra. The shaded black areas indicate influences of the boundaries. The heavy black lines indicate the 95% confidence limit.

Returning to the question asked:

(a) Where the coherence squared exceeds the 95% confidence limit, the two plots are generally out of phase. Positive values in one layer correspond to negative values in the other layer, and vice versa. This is consistent at all scales and over all times. This 'out of phase behaviour' can also be clearly observed in the time series plot, Figure 9–9.

9.4.2. Example 8-13 – Temperature and along shore currents

This example examines the along shore component of the currents (about 15 m below the water surface) with the water temperature at the same depth. The coastline runs approximately north-south hence the currents are running approximately parallel to the shoreline. This is a long term data set, over several years, and our interest lies primarily in the low frequency end of the spectrum.

Question: Does the temperature respond to changes in the along shore currents and how does this change over time?

Data used:

- The north-south current anomalies about 15 m below the sea surface and the water temperatures at the same location in the water column.
- Three years of data were selected; between 1 January 2019 and 31 October 2021.
- Data were collected from a location approximately 3 km offshore in waters 65 m deep.

Pre-processing:

- The original 5 minute data were averaged into hourly bins using a 7 point Gaussian filter.
- The hourly data were subject to the 120i913 filter described in Thompson (1983) to remove tidal and inertial signals.
- A linear trend and the annual signals were removed from each data set.
- The result is a set of daily temperature and along shore current data.
- A total of 1,024 data points (days) were used in the analysis (slightly less than 3 years. The anomalies are plotted in Figure 9–12.

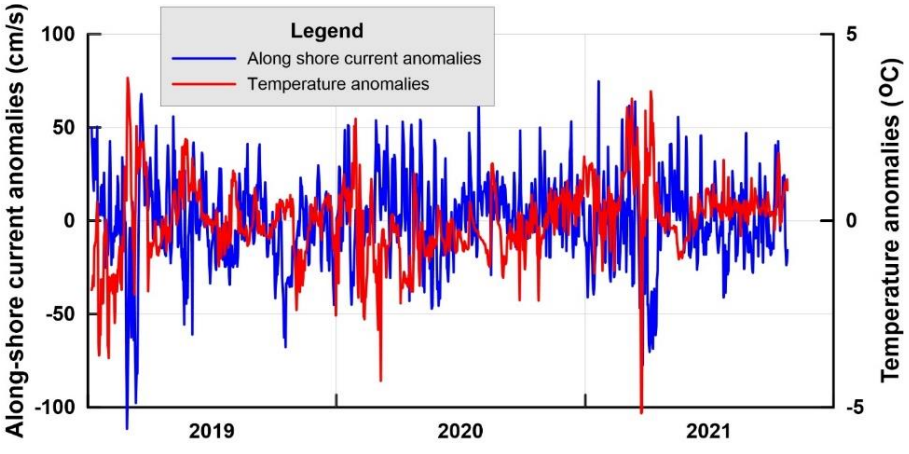


Figure 9–12. Time series plot of the anomalies of the north-south current component and the water temperature approximately 15 m below the sea surface.

Testing:

- The Dickey-Fuller test was undertaken on both data sets. The critical value was -2.86, the Tau-statistic for the current anomalies was -7.54, while that for the temperature anomalies was -8.01.
- As these both lie below the critical value, we cannot reject the null hypothesis that the time series are non-stationary. Therefore, we conclude that both data sets are stationary.

Smoothing:

- Confidence limits for the wavelet spectra were generated using Monte Carlo techniques (see Torrence and Compo, 1998 and Grinsted et al, 2004).

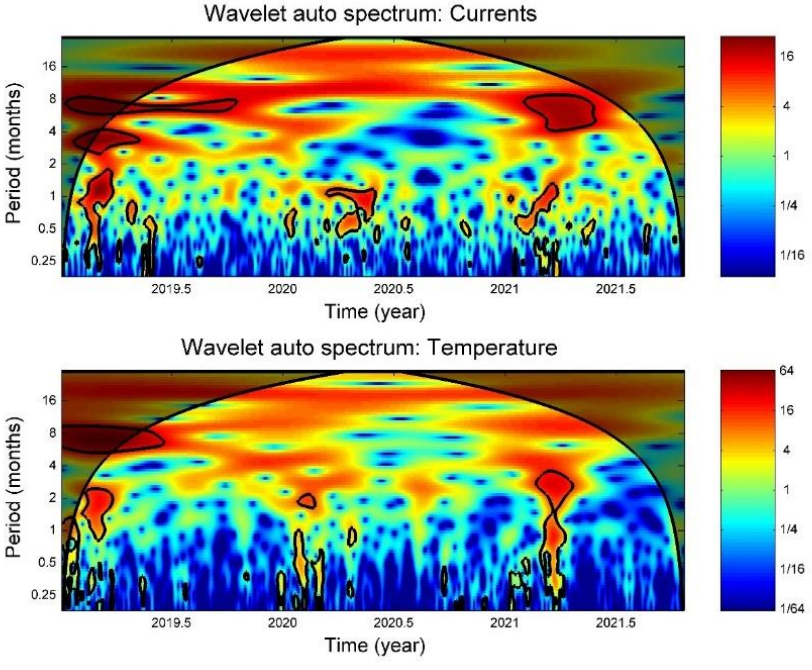


Figure 9–13. Wavelet auto spectra for each data set: top plot current anomalies (15 m below the sea surface), bottom plot temperature anomalies (15 m below the sea surface). The shaded black areas indicate influences of the boundaries. The heavy black lines indicate the 95% confidence limit.

Description of results:

- Both wavelet auto spectra show essentially the same results. High wavelet spectral energy is observed near early 2019, mid-2020 and early 2021. The Fourier period is less than 3 months.
- Similar features are observed near the same times in the cross wavelet spectrum. The phase difference is variable, with the arrows often point to the left, indicating that processes occur in opposite directions (i.e. a positive temperature anomaly corresponds to a negative value in the current anomalies, and vice versa).

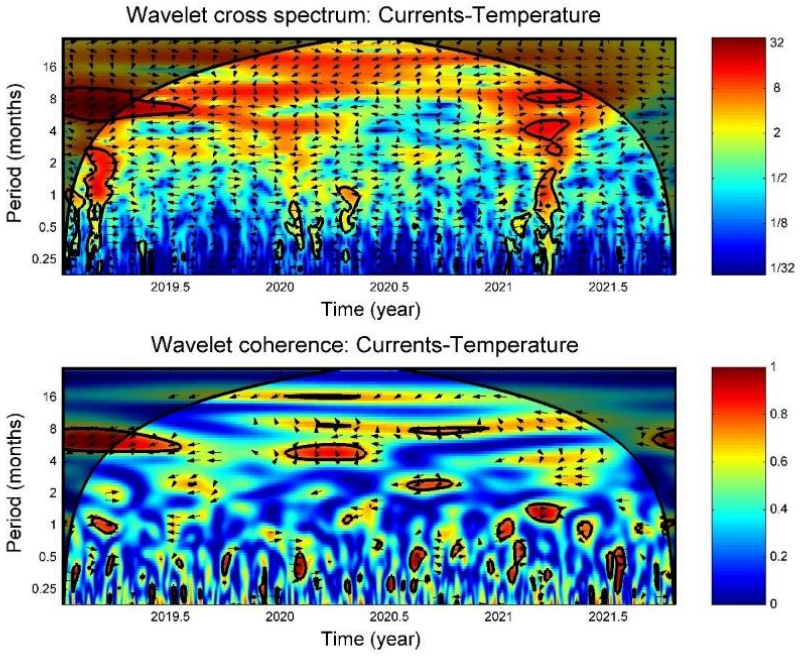


Figure 9–14. Top: Wavelet cross spectrum between the current and temperature anomalies. Bottom: wavelet coherence squared (indicated by the coloured contours) and phase difference (indicated by the arrows) spectra. The shaded black areas indicate influences of the boundaries. The heavy black lines indicate the 95% confidence limit.

Returning to the question asked.

(a) Coherence squared exceeds the 95% confidence limit in a band near the annual period. This is not consistent over time. There are several times when shorter period features also have high coherence squared values.

(b) The phase difference is broadly out of phase, although this is not consistent over time. This suggests that high temperature anomalies are more likely to occur when the current anomalies are negative i.e. flows towards the south (or from the tropics) bring with them warmer than normal waters.

9.4.3. Example 8-14 – Southern Oscillation Index and Tasman Sea surface temperature

The final example of the cross wavelet spectrum examines the Southern Oscillation Index (SOI) and the Tasman Sea surface temperatures over the 85 year period between 1935 and 2020. The SOI is a metric used to monitor the long period El Nino / La Nina cycle and is based on the monthly average difference in atmospheric pressure between Darwin and Tahiti. The Tasman Sea lies between Australia and New Zealand. Data were obtained from the Australian Bureau of Meteorology website. With such a long term data set, emphasis is placed on inter annular features.

Question: Does the coherence squared between the SOI and the Tasman Sea surface temperature exceed the 95% confidence limit for periods exceeding about 12 months? If so, are such exceedances consistent over the time period examined?

Data used:

- Monthly SOI values and monthly Tasman Sea surface temperatures. Time series plots of the SOI and Tasman Sea surface temperature anomalies are presented in Figure 9–15.

Pre-processing:

- No processing was undertaken on the SOI data.
- Anomalies were obtained for the Tasman Sea surface temperatures by removing the monthly averages generated between 1961 and 1990.

Testing:

- The Dickey-Fuller test was undertaken on both data sets. The critical value was -2.86, the Tau-statistic for the SOI was -14.99, while that for the Tasman Sea surface temperature anomalies was -8.19.
- As these both lie below the critical value, we reject the null hypothesis that the time series are not stationary.

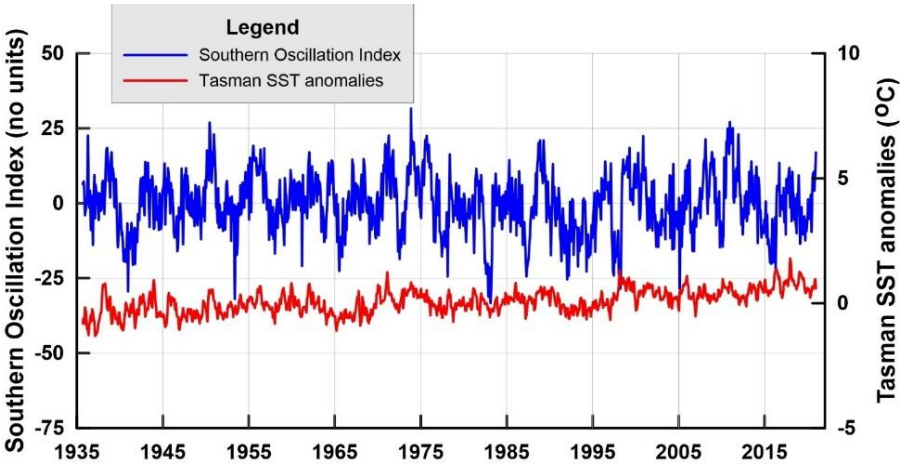


Figure 9–15. Time series plot of the Southern Oscillation Index and the Tasman Sea surface temperature anomalies between 1935 and 2020.

Smoothing:

- Confidence limits for the wavelet spectra were generated using Monte Carlo techniques (see Torrence and Compo, 1998 and Grinsted et al, 2004).

Description of results:

- The wavelet auto spectrum for the SOI shows several periods of high energy: between 1940 and 1960 in the 4-8 year period, between 1965 and 2000 in the 2-6 year period and between 2005 and 2015 in the 2-4 year period. This suggests that there is a long term movement towards shorter periods i.e. the El Nino / La Nina cycle is becoming more frequent as time progresses.
- The wavelet auto spectrum for the Tasman Sea surface temperature anomalies shows some isolated regions throughout the data set. The periods of these high energy features are generally less than 1 year. There are no long period features in the data set.
- The wavelet coherence squared lies above the 95% confidence limit between 1940 and 1990 for periods between 2-8 years.

Beyond 1990, regions of high coherence squared become patchy and their period is substantially shorter – about 1-2 years. The two data sets are in-phase.

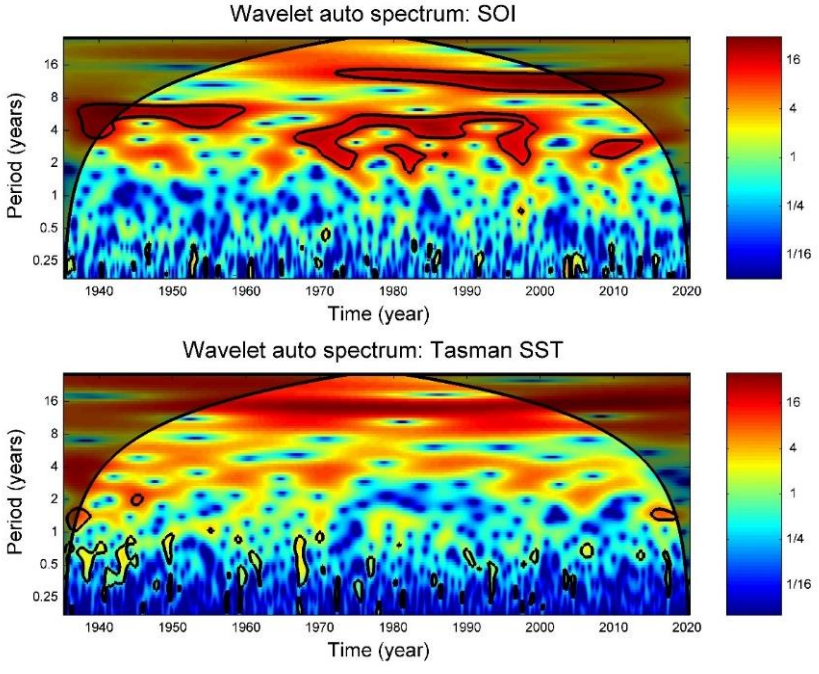


Figure 9–16. Wavelet auto spectra for each data: top plot Southern Oscillation Index (SOI), bottom plot Tasman Sea surface temperature (Tasman SST). The shaded black areas indicate influences of the boundaries. The heavy black lines indicate the 95% confidence limit.

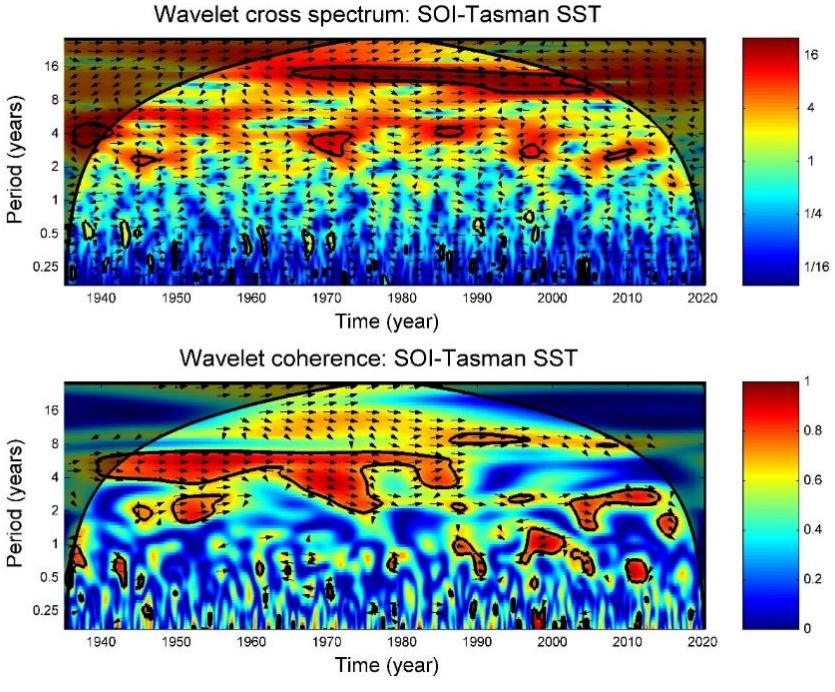


Figure 9–17. Top: Wavelet cross spectrum between Southern Oscillation Index (SOI) and Tasman Sea surface temperature (Tasman SST). Bottom: wavelet coherence squared (indicated by the coloured contours) and phase difference (indicated by the arrows) spectra. The shaded black areas indicate influences of the boundaries. The heavy black lines indicate the 95% confidence limit.

Returning to the questions asked:

(a) The coherence squared between the SOI and the Tasman Sea surface temperature anomalies exceed the 95% confidence limit for periods between 1-8 years across the whole time domain examined.

(b) Between 1940 and 1990 the coherence squared between the SOI and the Tasman Sea surface temperature anomalies exceed the 95% confidence limit for periods between 2-8 years. Between 1990 and 2020 coherence squared values exceeded the 95%

confidence limits became patchy and the period reduced to 1-2 years.

9.5. Summary

A time series provides excellent resolution in time but no resolution in frequency. We know precisely when an event occurs but we have no information on how frequently it occurs. In contrast, a spectrum gives excellent resolution in frequency but no resolution in time. We don't know when an event of particular frequency occurs.

The STFT is a compromise in resolution between time and frequency. It gives neither the temporal resolution of a time series nor the frequency of the Fourier transform – but a bit of both.

The wavelet transform is a more efficient resolution in time and scale (proportional to the inverse of frequency). This process is summarised in Figure 9–18.

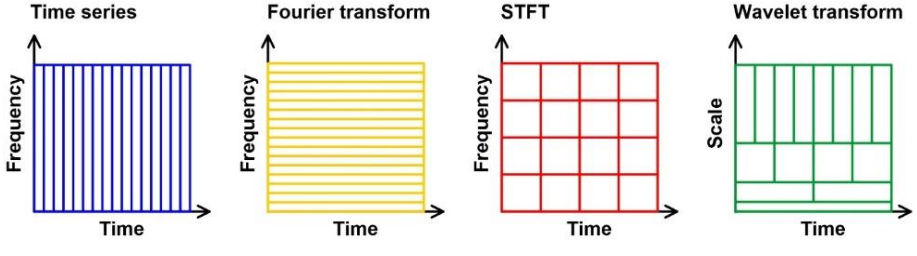


Figure 9–18. The trade-off in resolution between frequency and time. The time series has good resolution in time but no resolution in frequency. The spectrum (via the Fourier transform) has good frequency resolution but no time resolution. The STFT contains some resolution in time and some resolution in frequency. The wavelet transform provides a good balance between temporal and frequency resolution.

We can achieve this by passing the time series through a window, a process known as convolution. Convolution is a computationally expensive operation, but we can take advantage of the fact that convolution in time is equivalent to multiplication in frequency (and vice

versa). If we choose our window function judicially (i.e. such that it has an analytic Fourier transform and it obeys a number of rules), we can substantially speed up the convolution process.

If the window is a combination of a Gaussian function and a sine/cosine function, then the result is the Morlet wavelet. It's Fourier transform is analytic (it become a Gaussian function), hence we can speed up the convolution process by converting the time series to frequency space.

With spectral analysis we had a single parameter system (frequency) and we can view this as a two-dimensional plot (spectral energy as a function of frequency). The Morlet wavelet analysis results in a two parameter system (time and frequency, or translation and scale) and we view this as a three-dimensional plot (spectral energy as a function of time and scale or time and period).

Cross wavelet analysis can be undertaken in a manner that is almost identical to cross spectral analysis. The primary difference is that, for spectral analysis, averaging is done in the frequency domain. For cross wavelet analysis, averaging is done in both scale and time.

Chapter 10. Tips and Tricks

10.1. General

10.1.1. Different Fourier transform packages

There is an ambiguity in the definition of the Fourier transform (see Section 2.2, Chapter 2). We need to understand the exact formulation used by each package, particularly so if we're comparing the results from different packages.

10.1.2. Something for nothing?

At first glance it might appear that we're getting something for nothing. Our original data set shows the size of a signal as a function of time (or space). This thing called Fourier analysis transforms this into spectral amplitude AND phase as a function of frequency. Further, wavelet analysis transforms this into amplitude AND phase (scale) AND time (translation) as functions of frequency.

So, where do the additional pieces of information come from? In Fourier (or spectral) analysis the additional piece of information comes about because we have specified that the data can be represented as the sum of sine and cosine functions. It is this assumption, that the data can be represented in a very specific way, that leads us to the additional information. Similarly, wavelet analysis passes a specific waveform (or window) through the data. In our case, we've assumed that this waveform is a 'complex exponential multiplied by a Gaussian window' – the Morlet wavelet / transform. Again, it is this assumption that enables wavelet analysis to represent the data as two independent variables (amplitude and phase) as functions of two dependent variables (time and frequency).

10.1.3. Start simple

Starting a new topic is often confusing and we might not be sure of the results or how to interpret them. This is particularly so as the spectral

analysis becomes more complex (e.g. cross spectral analysis or rotary spectral analysis). We'd recommend starting simple. Generate a synthetic data set with one or two signals at specific frequencies. If you've done everything correctly, the spectral analysis should show peaks at frequencies corresponding to these one or two signals.

A test for cross (or rotary) analysis is to use the same data for the two sets of spectral analyses. This should result in a coherence squared value of unity across the entire spectrum and a phase difference of zero. (Noting that there is a $k\pi$ ambiguity in the phase).

If we're dealing with rotary spectra using a pair of vectors, one check is to set both vectors to be the same. In such cases, the inner rotary coherence squared and phase should be unity and zero, respectively (noting the $k\pi$ ambiguity in the phase). The outer rotary coherence squared and phase will not be unity or zero, respectively. We can check on the outer coherence squared and phase spectra by setting all of the Fourier components to, for example, the real component of time series one. In this case, the outer coherence squared and phase should be unity and zero, respectively (again, noting the $k\pi$ ambiguity in the phase).

10.2. Stationarity

In theory, a data set needs to be stationary (or at least weakly stationary) to perform a spectral analysis. Essentially, a data set is approximately stationary if the mean and variance of subsets of the data do not change. The statistical tests that are used to check for homogeneity of the means and variances might include ANOVA and Bartlett's test. Other tests for stationarity include the Dickey-Fuller and KPSS tests.

However, these (and other equivalent) tests require the data to be independent. It is quite likely that real data sets have serial dependence, either in space or time. Further, these tests are all based on hypotheses. When we have very large number of data points (as we often do with spectral analysis) the null hypothesis is almost always rejected. The Dickey-Fuller and KPSS tests have contrary null hypotheses, so one test (the Dickey-Fuller test) will almost always give the result that the time

series is stationary (i.e. we reject the null hypothesis that the data set is not stationary), while the KPSS test will almost always give the result that the time series is not stationary (i.e. we reject the null hypothesis that the data set is stationary). Similarly, the ANOVA and homogeneity of variance tests results will almost always fail (i.e. their respective null hypotheses will be rejected).

So, we're left with a dilemma. Do we accept the stationarity assumption, in which case it is unlikely that we would do any spectral analysis at all? Or, do we accept that the data set is unlikely to be strictly stationary and proceed anyway?

It would seem a pity to forgo a good analysis technique, so we'd recommend that the tests for stationarity be done and the results noted, remove the trend from the data set, then proceed with the spectral analysis. This then raises the question of whether stationarity is actually important for spectral analysis.

10.3. Resolution versus practical interpretation

There is a limit to what can be resolved in a data set. At the low frequency end, this is governed by the length of the data set. At the high frequency end, this resolution is limited by the Nyquist frequency (or Nyquist period). This differs from the practical interpretation of signals in the data set. Fundamentally, spectral analysis relies on the repetition of a signal. The more a signal repeats itself, the more confident we are that the signal is important. As a general rule, we like to see 5-10 cycles before we are confident in the interpretation of that signal.

Hence, we impart greater restrictions (than that of signal resolution alone) to be placed on the data before we are confident in correctly interpreting the results. At the low frequency end, we generally like to have about 5-10 cycles before we are confident in our interpretation of the signal. For example, if we have 12 months of data, then the lowest frequency signal that we'd be comfortable in interpreting is about 1-2 months. This ensures that we have 6-12 cycles in the data.

At the high frequency end, we'd ideally like about 5 times the Nyquist frequency before we're confident in interpreting the signal. In Section 3.7 we showed that, although sampling once every 6 months enabled us to resolve the annual signal, the more frequently we sample, the better we can represent a signal. For example, if we have data sampled every month, then the Nyquist period is 2 months. Our rule of thumb suggests that the highest frequency signal that we'd be comfortable in practically interpreting is about one year.

These are broad rules of thumb which may differ for different data sets.

10.4. Excel FFT routine

Chatfield (1996) uses the following definition for the Fourier coefficients.

$$a_p + ib_p = \frac{2}{N}\sum_{t=1}^{N} x_t \left[e^{i\frac{2\pi pt}{N}}\right] \qquad for \quad p = 1, 2, 3, \ldots . \left(\frac{N}{2}\right) - 1$$

The Excel FFT routine appears to use a slightly different formulation. Viz:

$$a_p - ib_p = \frac{2}{N}\sum_{t=1}^{N} x_t \left[e^{-i\frac{2\pi pt}{N}}\right] \qquad for \quad p = 1, 2, 3, \ldots . N$$

Comparing the results from Excel and Chatfield (1996) we note that the sign of the imaginary parts of the expression are reversed. This then has flow-on effects to the phase lag of the auto-spectrum and to several results from the cross-spectrum (and rotary spectrum.

Care needs to be taken when comparing results that have been generated using different Fourier transform programs or packages.

10.5. Spectral frequencies

In many (perhaps most) textbooks, Δt is tacitly assumed to be unity. Usually this is not a problem, because Δt might be 1 minute or 1 hour or 1 year.

If we sample, say, once every hour, and we set $\Delta t = 1$, then the frequency term in our output becomes 'cycles per hour' and that is fine. However, if we sample at, say, once every 39 minutes and we use $\Delta t = 1$, then the unit of our output frequency is 'cycles per 39 minutes'. This is not a particularly useful frequency and it is not easy to interpret the results.

We immediately say, 'why would anyone want to sample once every 39 minutes'? This is a reasonable question and under most circumstances we'd probably prefer to sample once every 30 minutes or once an hour. However, there are a couple of ways in which this could occur.

Firstly, we could have accidently set up our instrument to sample at 39 minutes, rather than some other more logical value. Yes, incorrectly setting up an instrument does happen, hopefully not too often.

Secondly, we always want to optimise the time available to us. Suppose we have 4 weeks to undertake an experiment which requires data from a deployed instrument. If we sampled hourly, we'd have 672 data points (over 4 weeks). If we planned on using the FFT, we require 2^n data points. Now 672 lies between 512 (2^9) and 1,024 (2^{10}), so our proposed hourly sampling puts us in a quandary. However, if we sampled at 39 minutes, we'd have 1,033 data points (close enough to 1,024).

If we use $\Delta t = \frac{60}{39} = 1.54$ then the units of our frequency are 'cycles per hour'.

10.6. Sum of spectral estimates

The sum of the spectral estimates equals the variance of the original time series. Assume we're using the Excel 'Fourier analysis' package.

Remove the trend from the data and perform the FFT. Results are returned as a complex number. Convert the complex number into its real and imaginary components.

Complex numbers in Excel are strings or texts. Excel can undertake mathematical operations using complex numbers (e.g. IMSUM, IMPRODUCT, IMSIN), but the result is still a string/text (even if the

result is a real number). Sometimes it is more convenient to generate (and use) the real and imaginary components by specifying IMREAL(complex number) and IMAGINARY(complex number). These results are real numbers to which Excel operations can be applied. Just remember that one of these numbers is actually the imaginary component.

Calculate the amplitude squared at each frequency.

$$(Amplitude)^2 = (Real)^2 + (Imaginary)^2$$

Sum all of the $(Amplitude)^2$ components. Divide this sum by the square of the number of data points in the FFT (N^2). The result is the variance of the original data set.

$$Variance = \frac{(Amplitude)^2}{N^2}$$

This is a way of checking the validity of the spectral analysis.

NOTE. If the trend is NOT removed, then the energy at the 'zeroth' frequency is not included in this sum.

10.7. Confidence limits

Often, we get carried away with the importance of confidence limits. A spectral peak lying below the 95% confidence limit does not necessarily mean that the processes resulting in this peak is unimportant.

The choice of 95% is arbitrary, a commonly accepted percentage. It is a statistical artefact that is not necessarily related to the (physical) process itself. A corollary is that there is a 5% chance that a statistically significant peaks actually results by random chance.

It may simply be that the energy at that frequency is insufficient to manifest itself at the 95% confidence level. If the frequency of this process varies over time (i.e. it is a non-stationary process) then it may not satisfy the 95% confidence limit criterion.

Care needs to be taken with the interpretation of the results, as it should be for interpreting the results from any data processing tool. The 95%

confidence limits should be viewed as an indication of importance and not used as a blanket rule.

10.8. A single vector

Checks can be made on the results from a single vector (or two scalars). If the two scalars are set to be equal (or if they are very close in space), then the phase difference spectrum should be zero (or small) and the coherence squared spectrum should be unity (or close to unity). Large departures from these values may indicate a problem with the software or with the processing.

10.9. A pair of vectors

Rotary spectra involving a pair of vectors can be a bit difficult to understand, let alone interpret. This is particularly so for the outer rotary components (i.e. comparing a clockwise rotating component of one vector with the anti-clockwise rotating component of another vector). One trick that is worth considering is to use the same data set for all four vector components. In so doing, all phases should equal zero and all coherence squared values should equal unity.

A second trick is to use the same vector for both data sets. This should result in the inner rotary phase spectra being zero and the inner coherence squared spectra having values of unity.

Thirdly, two vectors that are closely separated in space should give almost identical results. For example, two sets of currents from adjacent cells from the same Acoustic Doppler Current Profiler (ADCP) will be fairly close in space. This should result in the inner rotary phase spectra being 'close to zero' and the inner coherence squared spectra having values 'close to unity'.

These are simple checks that can be done to validate the rotary spectra program that is used.

Chapter 11. References

Below are the references that have been used in this manuscript. They are ones that we find most useful. However, it is recognised that different people think about topics in different ways and other texts and papers may resonate better with some people. The more you read, the better chance you'll have at understanding a topic.

Ashmead, J. (2010). Morlet wavelets in quantum mechanics. DOI: 10.12743/quanta.v1i1.5.

Bendat, J.S. and Piersol, A.G. (2000). Random data. Analysis and measurement procedures. Wiley-Interscience Publication, New York, NY, Third edition, 594p.

Bloomfield, P. (2000). Fourier analysis of time series. An introduction. Wiley-Interscience Publication, New York, NY, Second edition, 261p.

Boland, F.M. and Hamon, B.V. (1970). The East Australian Current, 1965-1968. Deep-Sea Research, Vol.17, No.4, pp.777-794.

Boland, F.M. and Church, J.A. (1981). The East Australian Current 1978. Deep-Sea Research, Vol.28A, pp.937-957Cheng, N. S. (1997). Simplified Settling Velocity Formula for Sediment Particles. Journal of Hydraulic Engineering Vol.123, No.2, pp.149-152.

BoM (2019). http://www.bom.gov.au/climate/change/index.shtml.

Carrier, G.F., Krook, M. and Pearson, C.E. (1966). Functions of a complex variable: theory and technique. McGraw-Hill Book Company, New York, 438p.

Church, J.A., Freeland, H.J. and Smith, R.L. (1986a). Coastal trapped waves on the east Australian continental shelf. Part I: Propagation of modes. Journal of Physical Oceanography, Vol.16, No.11, pp.1929-1943.

Church, J.A., Freeland, H.J. and Smith, R.L. (1986b). Coastal trapped waves on the east Australian continental shelf. Part II: Model verification. Journal of Physical Oceanography, Vol.16, No.11, pp.1945-1957.

Church, J.A. and Freeland, H.J. (1987). The energy source for the coastal-trapped waves in the Australian Coastal Experiment region. Journal of Physical Oceanography, Vol.17, pp.289-300.

Churchill, R.V. (1963). Fourier series and boundary value problems. McGraw-Hill Book Company, New York, NY, Second edition, 248p.

Chatfield, C. (1996). The analysis of time series. An introduction. Chapman and Hall, Texts in Statistical Sciences, London, UK, Fifth edition, 283p.

Cohen, M.X. (2019). https://www.youtube.com/watch?v =7ahrcB5HL0k.

Cooley, J.W. and Tukey, J.W. (1965). An algorithm for the machine calculation of complex Fourier series. Mathematics of Computation. Vol.19 (No.90), pp.297–301.

Cresswell, G.R. and Boland, F.M. (1981). Internal waves detected with a continental shelf current-meter array. Australian Journal of Marine and Freshwater Research, Vol.32, pp.1-8.

Danielson, G.C. and Lanczos, C. (1942). Some improvements in practical Fourier analysis and their application to X-ray scattering from liquids. Journal of the Franklin Institute, Vol. 233, pp. 365-380.

Debnath, L. (1998). Wavelet transforms and their applications. Proceedings of the Indian National Science Academy, Vol.64, A, No.6, pp. 685-713.

Emery, W.J. and Thomson, R.E. (1997). Data analysis methods in physical oceanography. Pergamon Press, Elsevier Science Ltd, Oxford, UK, 634p.

Fang, X.H., Boland, F.M. and Cresswell, G.R. (1984). Further observations of high-frequency current variations on the continental shelf

near Sydney, New South Wales. Australian Journal of Marine and Freshwater Research, Vol.35, pp.611-618.

Foreman, M.G.G. (1996). Manual for tidal heights analysis and prediction. Pacific Marine Science Report, 77-10. Institute of Ocean Sciences, Sidney, BC, Canada, 66p.

Foreman, M.G.G. and Henry, R.F. (1989). The harmonic analysis of tidal model time series. Advances in Water Resources, Vol. 12, pp.109-120.

Freeland, H.J., Boland, F.M., Church, J.A., Clarke, A.J., Forbes, A.M.G., Huyer, A., Smith, R.L., Thompson, R.O.R.Y., and White, N.J. (1986). The Australian coastal experiment: A search for coastal-trapped waves. Journal of Physical Oceanography, Vol.16, pp.1230-1249.

Fugal, D.L. (2009). Conceptual wavelets in digital signal processing: an in-depth, practical approach for the non-mathematician. Space and signals technical publishing, San Diego, California, USA, 302p.

Gabor, D. (1946). Theory of communication. Journal of the Institution of Electrical Engineers, Vol. 93, pp. 429-441.

Gibbs, M. T. and Middleton, J.H. (1997). Barotropic and baroclinic tides on the Sydney continental shelf. Continental Shelf Research, Vol.17, Issue 9, pp.1005-1027.

Godin, G. (1972). The analysis of tides. Liverpool University Press, 264p.

Gonella, J. (1972). A rotary-component method for analysing meteorological and oceanographic vector time series. Deep-Sea Research, Vol. 19, pp. 833-846.

Good, I.J. (1958). The Interaction Algorithm and Practical Fourier Analysis. Journal of the Royal Statistical Society. Series B (Methodological), Vol. 20, No. 2, pp. 361-372.

Griffin, D.A. and Middleton, J.H. (1990). Local and remote wind forcing of New South Wales inner shelf currents and sea level. Journal of Physical Oceanography, Vol.21, pp. 304-322.

Grinsted, A., Moore, J. and Jevrejeva, S. (2004). Application of the cross wavelet transform and wavelet coherence to geophysical time series. Nonlinear Processes in Geophysics (2004) Vol. 11, pp. 561–566.

Hamon, B.V. (1965). The East Australian Current, 1960-1964. Deep-Sea Research, Vol.12, pp.899-921.

Hamon, B.V. and Kerr, J.D. (1968). Time and space scales of variation in the East Australian Current, from merchant ship data. Australian Journal of Marine and Freshwater Research, Vol.19, pp.101-106.

Hamon, B.V. (1982). Notes on filtering. Unpublished manuscript, CSIRO Division of Oceanography, Australia, 32p.

Hassanieh, H, Indyk, P. Katabi, D. and Price, E. (2012). Simple and practical algorithm for sparse Fourier transform. Society for Industrial and Applied Mathematics, Proceedings of the 23rd annual ACM-SIAM symposium on discrete algorithms, doi.org/10.1137/1.9781611973099.93, pp.1183-1194.

Hayashi, Y. (1979). Space-time spectral analysis of rotary vector series. Journal of the Atmospheric Sciences, Vol. 36, No. 5, pp. 757-766.

Heideman, M.T., Johnson, D.H. and Bussus, C.S. (1984). Gauss and the history of the Fast Fourier Transform. Institute of Electrical and Electronic Engineers, Acoustics, Speech and Signal Processing Magazine. Volume 1, Issue 4, pp.14–21.

Kanji, G.K. (2001). 100 Statistical Tests. Sage Publications Ltd, London, UK, 215p.

Kelly, B.B., Hughes, P.A., Aller, H.D. and Aller, M.F. (2003). The cross wavelet transform and analysis of quasi-periodic behavior in the Pearson-Redhead VLBI survey sources. The Astrophysical Journal, Vol. 591, pp. 695–713.

Kwiatkowski, D., Phillips, P. C. B., Schmidt, P. and Shin, Y. (1992). Testing the null hypothesis of stationarity against the alternative of a unit root. Journal of Econometrics, Vol.54, Nos. 1-3, pp. 159–178. doi:10.1016/0304-4076(92)90104-Y.

Lighthill, M.J. (1978). Introduction to Fourier analysis and generalised functions. Cambridge University Press, Cambridge, UK, 79p.

MacDonald, H. S., Baird, M. E., and Middleton, J. H. (2009). Effect of wind on continental shelf carbon fluxes off southeast Australia: a numerical model. Journal of Geophysical Research, Vol.114, C05016.

McClean-Padman, J., and L. Padman, (1991). Summer upwelling on the Sydney inner continental shelf: The relative roles of local wind forcing and mesoscale eddy encroachment. Continental Shelf Research, Vol.11, pp.321–345, doi:10.1016/0278-4343(91)90025-2.

Middleton, J.H. (1982). Outer rotary cross spectra, coherences, and phases. Deep-Sea Research, Vol.29, Npo.10A, pp.1267-1269.

Mooers, C.N.K. (1973). A technique for the cross spectrum analysis of pairs of complex-valued time series, with emphasis on properties of polarized components and rotational invariants. Deep-Sea Research, Vol.20, pp.1129-1141.

Morlet, J., Arens; G., Fourgeau, E.; and Giard, D. (1982a). Wave propagation and sampling theory; Part I, Complex signal and scattering in multilayered media. Geophysics (1982) 47 (2): 203–221. https://doi.org/10.1190/1.1441328.

Morlet, J., Arens; G., Fourgeau, E.; and Giard, D. (1982b). Wave propagation and sampling theory; Part II, Sampling theory and complex waves. Geophysics (1982) 47 (2): 222–236. https://doi.org/10.1190/1.1441329.

Morrow, R. Jones, I.S.F., Smith, R.L. and Stabeno, P.J. (1990). Bass Strait forcing of Coastal Trapped Waves: ACE revisited. Journal of Physical Oceanography, Vol.20, pp. 1528-1538.

Parker, B.B. (2007). Tidal analysis and prediction. National Oceanic and Atmospheric Administration, U.S. Department of Commerce, NOAA Special Publication NOS CO-OPS 3, 388p.

Polikar, R. (2006). The wavelet tutorial. http://users.rowan.edu /~polikar/WAVELETS/ WTtutorial.html. 79p.

Press, W.H., Teukolsky, S.A., Vetterling, W.T. and Flannery, B.P. (1992). Numerical Recipes in FORTRAN – The Art of Scientific Computing. Cambridge University Press, Cambridge, UK, Second Edition, 963p.

Ramirez, R.W. (1985). The FFT – Fundamentals and concepts. Prentice-Hall Inc., Englewood Cliffs, NJ, USA, 178p.

Roughan, M. and Middleton, J.H. (2002). A comparison of observed upwelling mechanisms off the East Coast of Australia. Continental Shelf, Vol.22, pp.2551-2572.

Roughan, M. and Middleton, J.H. (2004). On the East Australian Current: variability, encroachment and upwelling. Journal of Geophysical Research, Oceans, Vol.109, Issue C7, doi:10.1029/2003JC001833, 16p.

Runge, C. (1903). Uber die zerlegung empirisch gegebener periodischer funktionen in sinuswellen. Z. Math. Phys., Vol. 48, pp. 443-456. [Translation in Heideman, M.T., Johnson, D.H. and Burrus, S. Gauss and the history of the Fast Fourier Transform, IEEE ASSP Magazine October 1984].

Sahai, H. and Ageel M.I. (2000). The analysis of Variance: Fixed, random and mixed models. Birkhauser Publications, Boston, USA, 742p.

Schaeffer, A., Gramoulle, A., Roughan, M. and Mantovanelli, A. (2017). Characterizing frontal eddies along the East Australian Current from HF radar observations. Journal of Geophysical Research, Oceans, Vol.122, Issue 5, pp.3964-3980.

Short, A.D. and Trenaman, N.L. (1992). Wave climate of the Sydney region, an energetic and highly variable ocean wave regime. Australian Journal of Marine and Freshwater Research, Vol.43, No.4, pp.765 – 791.

Spiegel, M.R. (1971). Advanced mathematics for engineers and scientists. Schaum's outline series, McGraw-Hill book company, New York, NY, 261p.

Tate, P.M., Jones, I.S.F. and Hamon, B.V. (1989). Time and space scales of surface temperatures in the Tasman Sea, from satellite data. Deep Sea Research Vol.36, No.3 pp.419 430.

Tate, P.M., Holden, C.J. and Tate, D.J. (2019). Influence of Plume Advection and Particle Settling on Wastewater Dispersion and Distribution. Marine Pollution Bulletin, Vol.145, pp. 678-690.

Thomson, W. (1860). On the reduction of observations of underground temperature; with application to Professor Forbes' Edinburgh observations, and the continued Carlton Hill Series. Transactions of the Royal Society, Vol. 22, No. 2, pp. 405-427.

Thompson, R. O. R. Y. (1983). Low-pass filters to suppress inertial and tidal frequencies. Journal of Physical Oceanography, Vol. 13, pp. 1077-1083.

Tilburg, C.E., Hurlburt, H.E., O'Brien, J.J. and Shriver, J.F. (2001). The dynamics of the East Australian Current system: the Tasman Front, the East Auckland Current, and the East Cape Current. Journal of Physical Oceanography, Vol.31, pp.2917-2943.

Torrence, C. and Compo, G.P. (1998). A practical guide to wavelet analysis. Bulletin of the American Meteorological Society, Vol. 79, No. 1, pp. 61-79.

Turner, I.L., Harley, M.D., Short, A.D., Simmons, J.A., Bracs, M.A., Phillips, M.S. and Splinter, K.D. (2016). A multi-decade dataset of monthly beach profile surveys and inshore wave forcing at Narrabeen, Australia. Nature, Scientific Data, Vol.3:160024; DOI: 10.1038/sdata.2016.24.

Walters, R.A. and Heston, C. (1982). Removing tidal-period variations from time-series data using low-pass digital filters. Journal of Physical Oceanography, Vol.12, pp. 112-115.

Warner, R.M. (1998). Spectral analysis of time-series data. The Guildford Press, New York, NY, 225p.

Wood, J.E., Roughan, M. and Tate, P.M. (2012). Finding a proxy for wind stress over the coastal ocean. Marine and Freshwater Research, Vol. 63, pp.528-544.

Zurbenko, I.G. (1991). Spectral analysis of non-stationary time series. International Statistical Review, Vol.59, No.2, pp.163-173.

Chapter 12. Appendix A – The Fast Fourier Transform

In Chapter 2 we introduced the Fourier transform. Here, we'll use a bit of theory to show how the discrete Fourier transform and the fast Fourier transform are obtained.

12.1. Discrete Fourier transform

Real data sets are rarely continuous. Mostly they consist of discrete values at specific points in time (or space). In such cases, the integrals are replaced by summations. However, because real data sets are finite in length, the limits become finite.

In much the same way that 'n' data points can be represented exactly by a polynomial of order 'n-1', 'N' data points can be represented by 'N' sine / cosine curves. The finite Fourier transform (or discrete Fourier transform, DFT) of the time series x_t or $x(t)$ can be defined as:

$$\mathcal{F}(p) = \frac{1}{N}\sum_{t=0}^{N-1} x(t)\ e^{-i\,2\pi\frac{pt}{N\Delta t}} \qquad for \quad p = 0, 1, 2 \ldots . N-1$$

$$= \frac{1}{N}\sum_{t=0}^{N-1} x(t)\ \left[cos\left(2\pi\frac{pt}{N\Delta t}\right) + i\ sin\left(2\pi\frac{pt}{N\Delta t}\right)\right] \qquad for \quad p = 0, 1, 2 \ldots . N-1$$

Usually, Δt is assumed to be unity and is often neglected in Fourier theory. This is fine when a convenient time step is used (e.g. 1 second, 1 hour, 1 day etc), but if sampling is, say, once every 90 minutes then it is best to set $\Delta t = 1.5\ t$ and the results will have units of 'cycles per hour'. If we left $\Delta t = 1$ then our results would have units of 'cycles per 90 minutes' – probably not the most convenient of units.

The inverse Fourier transform is then:

$$x(t) = \frac{N}{2\,\pi}\sum_{p=0}^{N-1}\mathcal{F}(p)\; e^{-i\,2\pi\frac{pt}{N\Delta t}} \qquad for \quad t = 0, 1, 2 \ldots . N-1$$

$$= \frac{N}{2\,\pi}\sum_{p=0}^{N-1}\mathcal{F}(p)\;\left[cos\left(2\pi\frac{pt}{N\Delta t}\right) + i\;\; sin\left(2\pi\frac{pt}{N\Delta t}\right)\right] \qquad for \;\; t$$

$$= 0, 1, 2 \ldots . N-1$$

Using the Fourier transform, if $p = 0$ then:

$$\mathcal{F}(0) = \frac{1}{N}\sum_{t=0}^{N-1} x(t)\;\left[cos(0) + i\;\; sin(0)\right] = \frac{1}{N}\sum_{t=0}^{N-1} x(t) = \langle x(t)\rangle$$

The complex term vanishes and we're left with the 'zeroth' terms simply being the average of the observations.

If $p = \frac{N}{2}$ then:

$$\mathcal{F}\left(\frac{N}{2}\right) = \frac{1}{N}\sum_{t=0}^{N-1} x(t)\;\left[cos\left(2\pi\frac{\frac{N}{2}t}{N\Delta t}\right) + i\;\; sin\left(2\pi\frac{\frac{N}{2}t}{N\Delta t}\right)\right]$$

$$= \frac{1}{N}\sum_{t=0}^{N-1} x(t)\;\left[cos\left(\pi\frac{t}{\Delta t}\right) + i\;\; sin\left(\pi\frac{t}{\Delta t}\right)\right]$$

$$= \frac{1}{N}\sum_{t=0}^{N-1} x(t)\;\left[cos\left(\pi\frac{t}{\Delta t}\right)\right]$$

$$= \frac{1}{N}\sum_{t=0}^{N-1} x(t)\;\left[-1\right]^t$$

We now look at what happens either side of $p = \frac{N}{2}$. We'll start with $p = \frac{N}{2} - 1$ and $p = \frac{N}{2} + 1$.

$p = \frac{N}{2} - 1$	$p = \frac{N}{2} + 1$
$\mathcal{F}\left(\frac{N}{2} - 1\right) = \frac{1}{N}\sum_{t=0}^{N-1} x(t)\ e^{-i\left[2\pi\frac{\left(\frac{N}{2}-1\right)t}{N\Delta t}\right]}$	$\mathcal{F}\left(\frac{N}{2} + 1\right) = \frac{1}{N}\sum_{t=0}^{N-1} x(t)\ e^{-i\left[2\pi\frac{\left(\frac{N}{2}+1\right)t}{N\Delta t}\right]}$
$= \frac{1}{N}\sum_{t=0}^{N-1} x(t)\ e^{-i\left[\pi\, t - \frac{2\,\pi\, t}{N\Delta t}\right]}$	$= \frac{1}{N}\sum_{t=0}^{N-1} x(t)\ e^{-i\left[\pi\, t + \frac{2\,\pi\, t}{N\Delta t}\right]}$
$= \frac{1}{N}\sum_{t=0}^{N-1} x(t)\ \left\{-cos\left[\pi\, t - \frac{2\,\pi\, t}{N\Delta t}\right] - i\sin\left[\pi\, t - \frac{2\,\pi\, t}{N\Delta t}\right]\right\}$	$= \frac{1}{N}\sum_{t=0}^{N-1} x(t)\ \left\{-cos\left[\pi\, t + \frac{2\,\pi\, t}{N\Delta t}\right] - i\sin\left[\pi\, t + \frac{2\,\pi\, t}{N\Delta t}\right]\right\}$
$= \frac{1}{N}\sum_{t=0}^{N-1} x(t)\ \left\{-\cos(\pi\, t)\, cos\left(-\frac{2\,\pi\, t}{N\Delta t}\right) + \sin(\pi\, t)\, sin\left(-\frac{2\,\pi\, t}{N\Delta t}\right) - i\ \sin(\pi\, t)\, cos\left(-\frac{2\,\pi\, t}{N\Delta t}\right) - i\ \cos(\pi\, t)\, sin\left(-\frac{2\,\pi\, t}{N\Delta t}\right)\right\}$	$= \frac{1}{N}\sum_{t=0}^{N-1} x(t)\ \left\{-\cos(\pi\, t)\, cos\left(\frac{2\,\pi\, t}{N\Delta t}\right) + \sin(\pi\, t)\, sin\left(\frac{2\,\pi\, t}{N\Delta t}\right) - i\ \sin(\pi\, t)\, cos\left(\frac{2\,\pi\, t}{N\Delta t}\right) - i\ \cos(\pi\, t)\, sin\left(\frac{2\,\pi\, t}{N\Delta t}\right)\right\}$
$= \frac{1}{N}\sum_{t=0}^{N-1} x(t)\ \left\{-\cos(\pi\, t)\, cos\left(\frac{2\,\pi\, t}{N\Delta t}\right) + i\ \cos(\pi\, t)\, sin\left(\frac{2\,\pi\, t}{N\Delta t}\right)\right\}$	$= \frac{1}{N}\sum_{t=0}^{N-1} x(t)\ \left\{-\cos(\pi\, t)\, cos\left(\frac{2\,\pi\, t}{N\Delta t}\right) - i\ \cos(\pi\, t)\, sin\left(\frac{2\,\pi\, t}{N\Delta t}\right)\right\}$

We find that the real parts are identical and the imaginary parts differ only by the sign. Generalising, we find that:

$$Real\left(t=\frac{N}{2}-q\right)=\ Real\left(t=\frac{N}{2}+q\right)$$

$$Imaginary\left(t=\frac{N}{2}-q\right)=\ -Imaginary\left(t=\frac{N}{2}+q\right)$$

i.e. they are complex conjugates.

This means that we can rewrite our Fourier transform as:

$$\mathcal{F}(p)=\langle x(t)\rangle+\frac{2}{N}\sum_{t=0}^{N-1}x(t)\ \left[cos\left(2\pi\frac{pt}{N\Delta t}\right)+i\ \ sin\left(2\pi\frac{pt}{N\Delta t}\right)\right]$$
$$+\frac{1}{N}\sum_{t=0}^{N-1}x(t)\ (-1)^t\qquad for\quad p=\ 1,2\ldots\ldots\frac{N}{2}-1$$

The frequency is given by:

$$\omega_p\ =\quad 2\,\pi\frac{p}{N\,\Delta t}\qquad for\quad p=1,2,3\ldots\ldots\left(\frac{N}{2}-1\right)$$

12.2. Fast Fourier transform

The fast Fourier transform (FFT) is simply a way to manipulate the DFT to substantially speed up the execution of the algorithm. But, like most efficiency gains, there is a penalty. The 'penalty' is that we need to factorise 'N'. The more we can factorise 'N', the faster becomes the execution. The most efficient factorisation comes when $N=2^r$, where r is an integer.

Popularisation of the FFT came about after the publication by Cooley and Tukey (1965). However, it appears that the FFT dates back to Gauss in the early 1800s. Although never formally published, an algorithm similar to that of Cooley and Tukey appeared in Gauss's collected work (circa 1866 – 11 years after his death).

Calculating the DFT by hand was a long and tedious process and subject to human errors. As a mathematical tool, its use was limited. After the second world war, computers (the machines not the people) became more widely available and used. This opened up new analytical approaches including the Fourier transform.

In 1965 computers were much slower and far less accurate than they are today. Control Data Corporation caused major consternation at IBM when they announced their CDC 6600 computer could undertake 3 x 10^6 FLOPS (floating point operations per second). Today's fastest computers are up to about 10^{18} FLOPS. Undertaking a discrete Fourier transform required much computing power and time, even at 3 x 10^6 FLOPS.

Cooley and Tukey (1965) described a way of undertaking the Fourier transform that used far fewer number of mathematical operations. While the DFT algorithm requires of the order N^2 mathematical operations, their 'radix 2' FFT algorithm requires of the order $N\,log_2 N$ operations. Further, because fewer mathematical operations were required, truncation errors were less and the results were more accurate. While this doesn't look like much of a saving in the number of operations, the differences become substantial when N is large.

Today, the speed and accuracy of modern computers makes the use of the FFT less important (for our particular work in ocean sciences where data sizes are generally less than 10,000). Even on a PC, the time taken to run a 10,000 point DFT is virtually instantaneous. Nevertheless, many inbuild routines in data analysis packages such as Excel, Matlab and even our Fortran programs use the 'radix 2' FFT algorithm. Much computer time can be saved if many Fourier transforms are required.

The big downside to using the 'radix 2' FFT algorithm is that the number of input data points needs have be integer power of two. We can always optimise our experiment to ensure we have close to 2^r data points, we can truncate or extrapolate our data sets accordingly or we can split the data set into multiple sizes of 2^r or even subsample at a different value of Δt. For example, if we have 6,144 data points at hourly intervals, we can subsample to generate 4,096 (2^{12}) data points at 90 minute intervals. Exactly what we do will depend on the question that we want to answer.

We now show how the 'radix 2' FFT algorithm may be generated. (This is just one of several ways in which this can be done). In essence, the FFT takes advantage of the fact that the sine and cosine functions are symmetric and circular, viz:

$$cos(\theta) = cos(-\theta) \qquad and \qquad sin(\theta) = -sin(-\theta)$$

$$cos\left(\frac{2\,\pi\,k\,t}{N}\right) = cos\left[\frac{2\,\pi\,(N+k)\,t}{N}\right] \quad and \quad sin\left(\frac{2\,\pi\,k\,t}{N}\right) = sin\left[\frac{2\,\pi\,(N+k)\,t}{N}\right]$$

In Section 12.1 we showed that the Fourier transform can be represented as:

$$\mathcal{F}(p) = \frac{1}{N}\sum_{t=0}^{N-1} x(t)\; e^{-i\,2\pi\frac{pt}{N\Delta t}} \qquad for \quad p = 0, 1, 2 \ldots . N-1$$

Without loss of generality, we remove the $1/N$ term (it plays no part in generating the FFT algorithm and is simply a constant for a given value of N). The most efficient form of the FFT occurs when the number of data points is a power of 2 i.e. $N = 2^r$, and we'll assume that is the case here. We're also explicitly assuming that $\Delta t = 1$. Hence, for a given value of p,

$$\mathcal{F}(p) \sim \sum_{t=0}^{N-1} x(t)\; e^{-i\,2\pi\frac{pt}{N}}$$

We now separate this sum into t even and t odd. (This will start to get a bit messy, but we will persist for a while). We do this by explicitly letting the even part be $t = 2m$ and the odd part be $t = 2m + 1$. Therefore,

$$\mathcal{F}(p) = \sum_{m=0}^{N/2-1} x(2m)\left[e^{-i\frac{2\pi p(2m)}{N}}\right] + \sum_{m=0}^{N/2-1} x(2m+1)\left[e^{-i\frac{2\pi p(2m+1)}{N}}\right]$$

Half of the observations will be at time even and half will be at time odd. Therefore, each of the two summations is only over half of the total length of the data. Rearranging, we get:

$$\mathcal{F}(p) = \sum_{m=0}^{N/2-1} x(2m)\left[e^{-i\frac{2\pi p(m)}{N/2}}\right] + \sum_{m=0}^{N/2-1} x(2m+1)\left[e^{-i\frac{2\pi p(m)}{N/2} - i\frac{\pi p(1)}{N/2}}\right]$$

$$= \sum_{m=0}^{N/2-1} x(2m)\left[e^{-i\frac{2\pi p(m)}{N/2}}\right] + \sum_{m=0}^{N/2-1} x(2m+1)\left[e^{-i\frac{2\pi p(m)}{N/2}}\ e^{-i\frac{\pi p}{N/2}}\right]$$

$$= \sum_{m=0}^{N/2-1} x(2m)\left[e^{-i\frac{2\pi p(m)}{N/2}}\right] + e^{-i\frac{\pi p}{N/2}} \sum_{m=0}^{N/2-1} x(2m+1)\left[e^{-i\frac{2\pi p(m)}{N/2}}\right]$$

$$= \sum_{m=0}^{N/2-1} x(2m)\left[e^{-i\frac{2\pi p(m)}{N/2}}\right] + W_N^p \sum_{m=0}^{N/2-1} x(2m+1)\left[e^{-i\frac{2\pi p(m)}{N/2}}\right]$$

Noting that we can take the $e^{-i\frac{\pi}{N/2}} = W_N$ term outside the summation because it does not depend on m. Examination of this expression shows that the first summation is simply a DFT of the even time elements of $x(t)$ but over $N/2$ time steps, while the second summation is a DFT of the odd time elements of $x(t)$ over $N/2$ time steps.

That is:

$$\mathcal{F}(p) = \mathcal{F}_{even}(p) + W_N^p\, \mathcal{F}_{odd}(p)$$

Where:

$$\mathcal{F}_{even}(p) = \sum_{m=0}^{N/2-1} x(2m)\left[e^{-i\frac{2\pi p(m)}{N/2}}\right]$$

$$\mathcal{F}_{odd}(p) = \sum_{m=0}^{N/2-1} x(2m+1)\left[e^{-i\frac{2\pi p(m)}{N/2}}\right]$$

One FFT of length N undertakes of the order N^2 computations. We have two DFTs but each of length $N/2$, so each of them will take $(N/2)^2$

computations or $N^2/2$ in total. So, in doing a single split, we've (approximately) halved the number of computations we need to undertake.

We chose N to be a power of two, so when we split $\mathcal{F}(p)$ into even and odd parts and halve N, the resulting 'half time series' are also even (and a power of two). If we halved the number of computations by doing one split, then, if we do another split, we should reduce the number of computations needed by another factor of two. This is exactly what we do.

Split the m's into odd and even parts, by letting the even part be $m = 2n$ and the odd part be $m = 2n + 1$. We then get:

$$\mathcal{F}_{even}(p) = \sum_{n=0}^{N/4-1} x(4n)\left[e^{-i\frac{2\pi p(2n)}{N/2}}\right] + e^{-i\frac{4\pi p}{N}} \sum_{n=0}^{N/4-1} x(4n+2)\left[e^{-i\frac{2\pi p(2n)}{N/2}}\right]$$

$$= \sum_{n=0}^{N/4-1} x(4n)\left[e^{-i\frac{2\pi p(n)}{N/4}}\right] + e^{-i\frac{4\pi p}{N}} \sum_{n=0}^{N/4-1} x(4n+2)\left[e^{-i\frac{2\pi p(n)}{N/4}}\right]$$

$$= \sum_{n=0}^{N/4-1} x(4n)\left[e^{-i\frac{2\pi p(n)}{N/4}}\right] + e^{-i\frac{2\pi p}{N/2}} \sum_{n=0}^{N/4-1} x(4n+2)\left[e^{-i\frac{2\pi p(n)}{N/4}}\right]$$

$$= \sum_{n=0}^{N/4-1} x(4n)\left[e^{-i\frac{2\pi p(n)}{N/4}}\right] + W_N^{2p} \sum_{n=0}^{N/4-1} x(4n+2)\left[e^{-i\frac{2\pi p(n)}{N/4}}\right]$$

$$= \mathcal{F}_{even-even}(p) + W_N^{2p}\,\mathcal{F}_{even-odd}(p)$$

$$= W_N^0\,\mathcal{F}_{even-even}(p) + W_N^{2p}\,\mathcal{F}_{even-odd}(p)$$

$$W_N^p\,\mathcal{F}_{odd}(p) = W_N^p \sum_{n=0}^{N/4-1} x(4n+1)\left[e^{-i\frac{2\pi pn}{N/4}}\right] + W_N^p \sum_{n=0}^{N/4-1} x(4n+3)\left[e^{-i\frac{2\pi p(2n+1)}{N/2}}\right]$$

$$= W_N^p \sum_{n=0}^{N/4-1} x(4n+1)\left[e^{-i\frac{2\pi p(n)}{N/4}}\right] + W_N^p\; e^{-i\frac{4\pi p}{N}} \sum_{n=0}^{N/4-1} x(4n+3)\left[e^{-i\frac{2\pi p(n)}{N/4}}\right]$$

$$= W_N^p \sum_{n=0}^{N/4-1} x(4n+1)\left[e^{-i\frac{2\pi p(n)}{N/4}}\right] + W_N^p\; W_N^{2p} \sum_{n=0}^{N/4-1} x(4n+3)\left[e^{-i\frac{2\pi p(n)}{N/4}}\right]$$

$$= W_N^p \sum_{n=0}^{N/4-1} x(4n+1)\left[e^{-i\frac{2\pi p(n)}{N/4}}\right] + W_N^{3p} \sum_{n=0}^{N/4-1} x(4n+3)\left[e^{-i\frac{2\pi p(n)}{N/4}}\right]$$

$$= W_N^p\,\mathcal{F}_{odd-even}(p) + W_N^{3p}\,\mathcal{F}_{odd-odd}(p)$$

$$\mathcal{F}(p) = \mathcal{F}_{even}(p) + W_N^p\,\mathcal{F}_{odd}(p)$$

$$\mathcal{F}(p) = W_N^0\,\mathcal{F}_{even-even}(p) + W_N^{2p}\,\mathcal{F}_{even-odd}(p) + W_N^{1p}\,\mathcal{F}_{odd-even}(p) + W_N^{3p}\,\mathcal{F}_{odd-odd}(p)$$

$$\mathcal{F}(p) = W_N^0\,\mathcal{F}_{ee}(p) + W_N^{2p}\,\mathcal{F}_{eo}(p) + W_N^{1p}\,\mathcal{F}_{oe}(p) + W_N^{3p}\,\mathcal{F}_{oo}(p)$$

Where we've replaced $even$ with e and odd with o. In matrix form, for $N = 4$, this looks like:

$$\mathcal{F}(p) = \begin{bmatrix} W_4^0 & W_4^{2p} & W_4^{1p} & W_4^{3p} \end{bmatrix} \begin{bmatrix} \mathcal{F}_{ee}(p) \\ \mathcal{F}_{eo}(p) \\ \mathcal{F}_{oe}(p) \\ \mathcal{F}_{oo}(p) \end{bmatrix} \qquad for \quad p = 0, 1, 2, 3$$

We continue to split the results in the same way until further splits are no longer possible. The original length of our time series was $N = 2^r$, so we can make r number of splits. When we've come to the end of our splitting, we find that the $\mathcal{F}_e(p)$'s and the $\mathcal{F}_o(p)$'s are just the elements of the time series i.e.

$$\mathcal{F}_{eeee.....ee}(p) = x(0)$$

$$\mathcal{F}_{eeee.....eo}(p) = x(^N/_2)$$

And so on. So, our matrix for $N = 4$ becomes:

$$\mathcal{F}(p) = \begin{bmatrix} W_4^0 & W_4^{2p} & W_4^{1p} & W_4^{3p} \end{bmatrix} \begin{bmatrix} x(0) \\ x(2) \\ x(1) \\ x(3) \end{bmatrix} \qquad for \quad p = 0, 1, 2, 3$$

$$[\mathcal{F}] = \begin{bmatrix} W_4^0 & W_4^0 & W_4^0 & W_4^0 \\ W_4^0 & W_4^2 & W_4^1 & W_4^3 \\ W_4^0 & W_4^4 & W_4^2 & W_4^6 \\ W_4^0 & W_4^6 & W_4^3 & W_4^9 \end{bmatrix} \begin{bmatrix} x(0) \\ x(2) \\ x(1) \\ x(3) \end{bmatrix}$$

$$[\mathcal{F}] = \begin{bmatrix} \mathcal{F}(0) \\ \mathcal{F}(1) \\ \mathcal{F}(2) \\ \mathcal{F}(3) \end{bmatrix} = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & -1 & -i & i \\ 1 & 1 & -1 & -1 \\ 1 & -1 & i & -i \end{bmatrix} \begin{bmatrix} x(0) \\ x(2) \\ x(1) \\ x(3) \end{bmatrix}$$

Now this needs to be multiplied by $^2/_N$ except for $\mathcal{F}(0)$ and $\mathcal{F}(^N/_2)$ which are each multiplied by $^1/_N$. Hence,

$$\begin{bmatrix} \mathcal{F}(0) \\ \mathcal{F}(1) \\ \mathcal{F}(2) \\ \mathcal{F}(3) \end{bmatrix} = \frac{1}{4} \begin{bmatrix} x(0) + x(2) + x(1) + x(3) \\ 2\,\{x(0) - x(2) - i\,x(1) + i\,x(3)\} \\ x(0) + x(2) - x(1) - x(3) \\ 2\,\{x(0) - x(2) + i\,x(1) - i\,x(3)\} \end{bmatrix}$$

There is a quick way to determine the numerical coefficients of the exponents of W_N and the order of the $x(t)$'s in the matrix, which involves 'bit reversal' (or STIB). We'll demonstrate this using $N = 4$.

	Time	BITS	STIB	Time associated with STIB	Value of $x(t)$ at STIB time
$x(0)$	0	0 0	0 0	0	$x(0)$
$x(1)$	1	0 1	1 0	2	$x(2)$
$x(2)$	2	1 0	0 1	1	$x(1)$
$x(3)$	3	1 1	1 1	3	$x(3)$

Using the bit reversal, we can generalise the matrix that is given by:

$$[\mathcal{F}] = \begin{bmatrix} W_4^0 & W_4^0 & W_4^0 & W_4^0 \\ W_4^0 & W_4^2 & W_4^1 & W_4^3 \\ W_4^0 & W_4^4 & W_4^2 & W_4^6 \\ W_4^0 & W_4^6 & W_4^3 & W_4^9 \end{bmatrix} \begin{bmatrix} x(0) \\ x(2) \\ x(1) \\ x(3) \end{bmatrix}$$

Specify the following variables. Let the rows and columns in the matrix be given by $r\ (= 0, 1, 2, \ldots N-1)$ and $c\ (= 0, 1, 2, \ldots N-1)$, respectively. Let r_{BR} indicate the rows in bit reversal (STIB) order and Let c_{BR} indicate the columns in bit reversal order. Then the matrix above can be written as:

$$[\mathcal{F}(r)] = \left[W_N^{(r \times c_{BR})}\right][x(r_{BR})]$$

The computational saving comes about because many elements of $W_N^{(r \times c_{BR})}$ are the same – they only need to be evaluated once. In fact, many of the elements of $W_N^{(r \times c_{BR})}$ only have one of four values: ± 1 or $\pm i$.

12.3. Examples

12.3.1. Example 1: N=4

Let's put all this together in a simple example. Assume we have the following data set, comprising four data anomalies (i.e. data with the trend removed) uniformly spaced in time. Viz:

Time	**t=0**	**t=1**	**t=2**	**t=3**
Data x(t)	-0.120	0.155	0.050	-0.085

We start with our matrix, which we need to populate.

$$[\mathcal{F}(r)] = \left[W_N^{(r \times c_{BR})}\right][x(r_{BR})]$$

The bit reversal order is: (0, 2, 1, 3).

c	0	1	2	3	
c_{BR}	0	2	1	3	
r					
0	W_4^0	W_4^0	W_4^0	W_4^0	$x(0)$
1	W_4^0	W_4^2	W_4^1	W_4^3	$x(2)$
2	W_4^0	W_4^4	W_4^2	W_4^6	$x(1)$
3	W_4^0	W_4^6	W_4^3	W_4^9	$x(3)$

c	0	1	2	3
c_{BR}	0	2	1	3

$$r \begin{matrix} 0 \\ 1 \\ 2 \\ 3 \end{matrix} \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & W_4^2 & W_4^1 & -W_4^1 \\ 1 & -W_4^2 & W_4^2 & W_4^2 \\ 1 & W_4^2 & -W_4^1 & W_4^1 \end{bmatrix} \begin{bmatrix} x(0) \\ x(2) \\ x(1) \\ x(3) \end{bmatrix}$$

$$\begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & -1 & -i & i \\ 1 & 1 & -1 & 1 \\ 1 & -1 & i & -i \end{bmatrix} \begin{bmatrix} -0.120 \\ 0.050 \\ 0.155 \\ -0.085 \end{bmatrix}$$

$$\begin{bmatrix} \mathcal{F}(0) \\ \mathcal{F}(1) \\ \mathcal{F}(2) \\ \mathcal{F}(3) \end{bmatrix} = \begin{matrix} 1/4 \\ 1/2 \\ 1/2 \\ 1/4 \end{matrix} \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & -1 & -i & i \\ 1 & 1 & -1 & 1 \\ 1 & -1 & i & -i \end{bmatrix} \begin{bmatrix} -0.120 \\ 0.050 \\ 0.155 \\ -0.085 \end{bmatrix}$$

$$\begin{bmatrix} \mathcal{F}(0) \\ \mathcal{F}(1) \\ \mathcal{F}(2) \\ \mathcal{F}(3) \end{bmatrix} = \begin{bmatrix} 0.000 & i\,0.000 \\ -0.085 & -i\,0.120 \\ -0.035 & i\,0.000 \\ -0.085 & i\,0.120 \end{bmatrix}$$

We can check the results using the discrete Fourier transform, which is given by (with $\Delta t = 1$):

$$\mathcal{F}(p) = \langle x(t) \rangle + \frac{2}{N}\sum_{t=0}^{N-1} x(t)\ \left[cos\left(2\pi\frac{pt}{N}\right) + i\ sin\left(2\pi\frac{pt}{N}\right)\right] + \frac{1}{N}\sum_{t=0}^{N-1} x(t)\ (-1)^t \qquad for \quad p = 0, 1, 2 \ldots \ldots \frac{N}{2} + 1$$

Now:

$$\mathcal{F}(0) = \langle x(t) \rangle = \frac{(-0.120 + 0.155 + 0.050 - 0.085)}{4} = 0$$

$$\mathcal{F}(2) = \frac{1}{N}\sum_{t=0}^{N-1} x(t)\ (-1)^t$$

$$= \frac{1}{4}[-0.120(-1)^0 + 0.155(-1)^1 + 0.050(-1)^2 - 0.085(-1)^3] = \frac{1}{4}[-0.120 - 0.155 + 0.050 + 0.085]$$

$$= -0.035$$

And,

$$\mathcal{F}(p) = \frac{2}{N}\sum_{t=0}^{N-1} x(t)\ \left[cos\left(2\pi\frac{pt}{N}\right) + i\ sin\left(2\pi\frac{pt}{N}\right)\right] \qquad for \quad p = 1, 3$$

$$\mathcal{F}(1) = \frac{2}{N}\sum_{t=0}^{N-1} x(t)\ \left[cos\left(2\pi\frac{1t}{N}\right) + i\ sin\left(2\pi\frac{1t}{N}\right)\right]$$

$$= \frac{2}{4}\Big\{x(0)\ [cos(0) + i\ sin(0)]\ +\ x(1)\left[\cos\left(\frac{2\pi}{4}\right) + i\ sin\left(\frac{2\pi}{4}\right)\right] + x(2)\left[\cos\left(\frac{4\pi}{4}\right) + i\ sin\left(\frac{4\pi}{4}\right)\right] + x(3)\left[\cos\left(\frac{6\pi}{4}\right) + i\ sin\left(\frac{6\pi}{4}\right)\right]\Big\}$$

$$= \frac{2}{4}\{x(0) + i\,x(1) - x(2) - i\,x(3)\}$$

$$= \frac{1}{2}\{[x(0) - x(2)] + i\,[x(1) - x(3)]\}$$

$$= \frac{1}{2}\{[-0.120 - 0.050] + i\,[0.155 + 0.085]\}$$

$$\mathcal{F}(1) = -0.085 - i\,0.120$$

$$\mathcal{F}(3) = \mathcal{F}(1)^* = -0.085 + i\,0.120$$

It might not appear that the computational savings are all that great, and they're not when N is small. The big computational advantage comes when N is large. In fact, there is little perceivable difference even when $N = 10{,}000$ (towards the upper end of the length of time series of many oceanographic applications). However, applications such as image processing use very large values of N, and the differences in computational speed using the DFT and the FFT can be large.

The Fourier transform was also run using the Excel FFT algorithm. All three sets of results are given below.

	DFT		**FFT**		**Excel**	
	Real	**Imaginary**	**Real**	**Imaginary**	**Real**	**Imaginary**
F(0)	0.000	0.000	0.000	0.000	0.000	0.000
F(1)	-0.085	-0.120	-0.085	-0.120	-0.170	-0.240
F(2)	-0.035	0.000	-0.035	0.000	-0.140	0.000
F(3)	-0.085	0.120	-0.085	0.120	-0.170	0.240

Note that the Excel FFT algorithm does not multiply the results by $2/N$, for $\mathcal{F}(1)$ and $\mathcal{F}(3)$ (which is ½ in this case) or by $1/N$, for $\mathcal{F}(0)$ and $\mathcal{F}(2)$ hence the Excel values are double the magnitude of those calculated here. This is not a problem *per se*, as the difference is only a constant. However, we need to be careful if we're comparing results that have been generated using different FFT (or DFT) packages.

12.3.2. Example 2: N=8

Let's look at another example, this time with 8 data points. Again, these are anomalies, so the linear trend has been removed.

Time	t=0	t=1	t=2	t=3
Data x(t)	-0.768	0.098	0.597	0.564

Time	t=4	t=5	t=6	t=7
Data x(t)	0.162	-0.282	-0.414	0.043

We start with our matrix, which we need to populate.

$$[\mathcal{F}(r)] = \left[W_N^{(r \times c_{BR})}\right][x(r_{BR})]$$

The bit reversal order is: (0, 4, 2, 6, 1, 5, 3, 7).

c	0	1	2	3	4	5	6	7	
c_{BR}	0	4	2	6	1	5	3	7	
r									
0	W_8^0	W_8^0	W_8^0	W_8^0	W_8^0	W_8^0	W_8^0	W_8^0	$x(0)$
1	W_8^0	W_8^4	W_8^2	W_8^6	W_8^1	W_8^5	W_8^3	W_8^7	$x(4)$
2	W_8^0	W_8^8	W_8^4	W_8^{12}	W_8^2	W_8^{10}	W_8^6	W_8^{14}	$x(2)$
3	W_8^0	W_8^{12}	W_8^6	W_8^{18}	W_8^3	W_8^{15}	W_8^9	W_8^{21}	$x(6)$
4	W_8^0	W_8^{16}	W_8^8	W_8^{24}	W_8^4	W_8^{20}	W_8^{12}	W_8^{28}	$x(1)$
5	W_8^0	W_8^{20}	W_8^{10}	W_8^{30}	W_8^5	W_8^{25}	W_8^{15}	W_8^{35}	$x(5)$
6	W_8^0	W_8^{24}	W_8^{12}	W_8^{36}	W_8^6	W_8^{30}	W_8^{18}	W_8^{42}	$x(3)$
7	W_8^0	W_8^{28}	W_8^{14}	W_8^{42}	W_8^7	W_8^{35}	W_8^{21}	W_8^{49}	$x(7)$

We now evaluate each of the $W_N^{(r \times c_{BR})}$'s noting that there are many symmetries and many cases where $W_N^{(r \times c_{BR})}$ takes on one of the four values: ± 1 or $\pm i$. The symmetry means that, in this case, we only need to calculate W_8^1 through to W_8^7.

c	0	1	2	3	4	5	6	7	
c_{BR}	0	4	2	6	1	5	3	7	
r									
0	1	1	1	1	1	1	1	1	-0.768
1	1	-1	$-i$	i	W_8^1	$-W_8^1$	W_8^3	$-W_8^3$	0.162
2	1	1	-1	-1	$-i$	$-i$	i	i	0.597
3	1	-1	i	$-i$	W_8^3	$-W_8^3$	W_8^1	$-W_8^1$	-0.414
4	1	1	1	1	-1	-1	-1	-1	0.098
5	1	-1	$-i$	i	$-W_8^1$	W_8^1	$-W_8^3$	W_8^3	-0.282
6	1	1	-1	-1	i	i	$-i$	$-i$	0.564
7	1	-1	i	$-i$	$-W_8^3$	W_8^3	$-W_8^1$	W_8^1	0.043

$$\begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & -1 & -i & i & 0.71 - i\,0.71 & -0.71 + i\,0.71 & -0.71 - i\,0.71 & 0.71 + i\,0.71 \\ 1 & 1 & -1 & -1 & -i & -i & i & i \\ 1 & -1 & i & -i & -0.71 - i\,0.71 & 0.71 + i\,0.71 & 0.71 - i\,0.71 & -0.71 + i\,0.71 \\ 1 & 1 & 1 & 1 & -1 & -1 & -1 & -1 \\ 1 & -1 & -i & i & -0.71 + i\,0.71 & 0.71 - i\,0.71 & 0.71 + i\,0.71 & -0.71 - i\,0.71 \\ 1 & 1 & -1 & -1 & i & i & -i & -i \\ 1 & -1 & i & -i & 0.71 + i\,0.71 & -0.71 - i\,0.71 & -0.71 + i\,0.71 & 0.71 - i\,0.71 \end{bmatrix} \begin{bmatrix} -0.768 \\ 0.162 \\ 0.597 \\ -0.414 \\ 0.098 \\ -0.282 \\ 0.564 \\ 0.043 \end{bmatrix}$$

Matrix multiplication gives (remembering to multiple by $^{2}/_{N}$ except for $\mathcal{F}(0)$ and $\mathcal{F}(^{N}/_{2}) = \mathcal{F}(4)$ each of which are multiplied by $^{1}/_{N}$):

$$\begin{bmatrix} \mathcal{F}(0) \\ \mathcal{F}(1) \\ \mathcal{F}(2) \\ \mathcal{F}(3) \\ \mathcal{F}(4) \\ \mathcal{F}(5) \\ \mathcal{F}(6) \\ \mathcal{F}(7) \end{bmatrix} = \begin{bmatrix} 0.0000 & \text{i } 0.0000 \\ -0.2575 & \text{-i } 0.4123 \\ -0.1975 & \text{i } 0.1975 \\ -0.2075 & \text{i } 0.0935 \\ -0.1059 & \text{i } 0.0000 \\ -0.2075 & \text{-i } 0.0935 \\ -0.1975 & \text{-i } 0.1975 \\ -0.2575 & \text{i } 0.4123 \end{bmatrix}$$

Again, we ran this same data through the Excel FFT algorithm. The results were:

	FFT		Excel		Excel (adjusted)	
	Real	**Imaginary**	**Real**	**Imaginary**	**Real**	**Imaginary**
$\mathcal{F}(0)$	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
$\mathcal{F}(1)$	-0.2575	-0.4123	-1.0300	-1.6491	-0.2575	-0.4123
$\mathcal{F}(2)$	-0.1975	0.1975	-0.7900	0.7900	-0.1975	0.1975
$\mathcal{F}(3)$	-0.2075	0.0935	-0.8299	0.3740	-0.2075	0.0935
$\mathcal{F}(4)$	-0.1059	0.0000	-0.8469	0.0000	-0.1059	0.0000
$\mathcal{F}(5)$	-0.2075	-0.0935	-0.8299	-0.3740	-0.2075	-0.0935
$\mathcal{F}(6)$	-0.1975	-0.1975	-0.7900	-0.7900	-0.1975	-0.1975
$\mathcal{F}(7)$	-0.2575	0.4123	-1.0300	1.6491	-0.2575	0.4123

Note again, that the Excel FFT algorithm does not multiply the results by $2/N$, or $1/N$ accordingly, hence the unadjusted Excel values differ from those calculated here.

12.4. The sparse Fourier transform

As a closing remark, there is a way to undertake a DFT that is faster than the FFT – the sparse Fourier transform (SFT) e.g. Hassanieh et al (2012). It is designed for handling very large data sets, so it's application to our work in ocean sciences may not be critical, and it is not discussed here in any detail. Nevertheless, it is a relatively new approach that speeds up a Fourier analysis and deserves some mention.

In many situations (particularly in the medical and image enhancement fields) the spectrum is dominated by a few relatively large peaks. Much of the energy in the spectrum is small and unimportant, and can be effectively ignored.

Suppose we have a data set (e.g. time series) of length N. Suppose also, that most of the energy is contained in k ($\ll N$) spectral bands. The FFT transforms the entire data set from which we can identify the k's that are important. This requires $O(N\,log_2 N)$ operations. The basic idea behind the SFT is to recover the spectral energy at the k's without having to analyse the entire data set. This requires $O(k\,log_2 N)$ operations.

Chapter 13. Appendix B – Principal Component Analysis

In Example 5-9 (cross-spectral analysis using a vector of current components) we used one aspect of Principal Component Analysis (PCA) to determine the orientation of orthogonal axes that maximise (and minimise) the variance. This is further explored below.

More generally, PCA can be used to reduce the dimensionality of a problem. Some oceanographic instruments (e.g. thermistor strings, Acoustic Doppler Current Profilers or ADCPs) will measure their respective variables at multiple depths in the water column. This may mean that we have 10 or 20 or more sets of data. Including all of the cross products and combinations, this may mean thousands of different analyses. PCA is a way in which the number of data sets can be reduced to just a few (i.e. the dimension of the problem is reduced), resulting in far fewer analyses. There are no formal statistical tests required before undertaking a PCA.

PCA can also be used in a single time series to examine the different patterns (or 'modes') of variability. We can then see how these modes of variability change over time. This aspect of PCA is outside the scope of this book.

PCA has two major downsides. The first is that we lose some variability in our system. The second is that it is often difficult to relate the resulting 'modes of variability' to physical processes.

13.1. Axis Orientation

PCA can be used to rotate the axes of a vector such that the main axis aligns along the dominant direction. This is the axis about which the variance is smallest – the first principal axis. (This may sound a little counter-intuitive, but it is simply the definition that is used). Orthogonal to this axis is the second principal axis – the direction along which the variance in the data is maximised.

If we're analysing current meter data from a near-shore location, the direction of the first principal axis should align (approximately) with the shoreline (or local bathymetry). Coastal currents (in the absence of rivers or similar discharges) will generally flow in a direction parallel with the coast.

The direction of the first principal axis (θ_1) is given by:

$$\theta_1 = 0.5 \times tan^{-1}\left[\frac{2 \times covariance(x,y)}{variance(x) - variance(y)}\right]$$

There is an ambiguity in this direction because any axis extends through zero in the '+/-' direction. The direction of the second principal axis is given by:

$$\theta_2 = \theta_1 \pm 90^o$$

13.2. Dimension Reduction

To illustrate how to reduce the dimensions using PCA it might be best to show the steps via a simple example. In this example we have two time series x_1 and x_2 but this works exactly the same as if we had 5 or 10 or any number of time series. (If we had 5 time series we need to find the roots of a quintic polynomial – so a bit difficult by hand, so by restricting it to two time series we only have to solve a quadratic equation). In many oceanographic applications, our data will originate from a set of time series in the vertical (e.g. from a thermistor string), but PCA can also be applied to a laterally spaced array of variables (e.g. temperature or currents) from which horizontal modes of variability can be obtained.

Further, the data do not have to be of the same type. In Example 13.4.1 we use temperature data sets, but we could equally use, say, one temperature, one salinity, three nitrogen species, two phosphorus species and so on. Example 13.4.2 uses a set of 14 different water quality variables.

Suppose we have the following two time series (each comprising four time steps).

Time	1	2	3	4
x1	3	5	6	2
x2	2	5	4	2

As we've done with most of our examples, the first step is to remove the linear trend to generate a set of anomalies. Some texts suggest removing the mean to produce anomalies, but we prefer removing a trend (in this case linear), or perhaps a seasonal curve.

Time	1	2	3	4
x1	-1.30	0.90	2.10	-1.70
x2	-1.40	1.70	0.80	-1.10

At each time step, we set up a matrix by multiplying the data by its transpose. In this case we have four time steps, so we'll generate four matrices. We have two variables, so the size of each matrix is 2 x 2.

Time 1

	-1.30	-1.40
-1.30	1.69	1.82
-1.40	1.82	1.96

Time 2

	0.90	1.70
0.90	0.81	1.53
1.70	1.53	2.89

Time 3

	2.10	0.80
2.10	4.41	1.68
0.80	1.68	0.64

Time 4

	-1.70	-1.10
-1.70	2.89	1.87
-1.10	1.87	1.21

Average the four (in this case) matrices, which we'll call $\underline{\boldsymbol{A}}$:

2.45	1.73
1.73	1.68

Calculate the eigenvalues and eigenvectors of matrix $\underline{\boldsymbol{A}}$. The eigenvalues, $\boldsymbol{\lambda}$, are given by:

$$\underline{A} - \lambda \underline{I} = \begin{vmatrix} A_{11} - \lambda & A_{12} \\ A_{21} & A_{22} - \lambda \end{vmatrix} = 0$$

$$(A_{11} - \lambda)(A_{22} - \lambda) - A_{12}A_{21} = 0$$

$$\lambda^2 - (A_{11} + A_{22})\,\lambda + (A_{11}A_{22} - A_{12}A_{21}) = 0$$

$$\lambda = \frac{-[-(A_{11} + A_{22})] \pm \sqrt{[-(A_{11} + A_{22})]^2 - 4 \times 1 \times (A_{11}A_{22} - A_{12}A_{21})}}{2 \times 1}$$

The roots of this equation are the eigenvalues. There are two data sets, so we'll get two eigenvalues that we'll call λ_1 and λ_2. In this example: $\lambda_1 = 3.8303$ and $\lambda_2 = 0.2945$. These are associated with two 'modes' (or patterns) of variability.

Normalise the eigenvalues:

$$\lambda_1 = \frac{3.8303}{3.8303 + 0.2945} = 0.93 \quad and \quad \lambda_2 = \frac{0.2945}{3.8303 + 0.2945} = 0.07$$

The normalised eigenvalues represent the proportion that each mode contributes to the variability in the data. Here, Mode 1 contributes 93% to the overall variability of the data, while Mode 2 contributes just 7%. If we're willing to accept that we will only have 93% of the variability of the data, then we can replace the two data sets, x_1 and x_2 with a single data set that is generated from Mode 1.

If we had ten data sets, then we generate ten modes. However, we might find that 95% of the variability of the original data is contained in Mode

1 plus Mode 2. So, if we're willing to sacrifice 5% of the variability, we can represent the ten original data sets with two new characteristic data sets, one generated from Mode 1 and the second generated from Mode 2. The amount of analyses required is substantially reduced – the dimension of the problem is reduced from ten to two.

Unfortunately, these modes are just patterns in the data sets and, in general, they do not correspond to physical processes. This makes the interpretation of the results a little tricky. However, we can apply our spectral analysis to the different modes and use that information to determine the physical processes contained in each mode.

Calculate the eigenvectors, one for each eigenvalue: $\left(\underline{A} - \lambda_i \underline{I}\right) \underline{u_i} = 0$

For the first eigenvalue ($\lambda_1 = 3.8303$), we get:

$$\begin{pmatrix} -1.3805 & 1.7250 \\ 1.7250 & -2.1555 \end{pmatrix} \begin{pmatrix} u_1 \\ u_2 \end{pmatrix} = \underline{0} \quad \Rightarrow \begin{pmatrix} u_1 \\ u_2 \end{pmatrix} = \begin{pmatrix} 1 \\ 0.8003 \end{pmatrix}$$

For the second eigenvalue ($\lambda_1 = 0.2945$), we get

$$\begin{pmatrix} -2.1555 & 1.7250 \\ 1.7250 & 1.3805 \end{pmatrix} \begin{pmatrix} v_1 \\ v_2 \end{pmatrix} = \underline{0} \quad \Rightarrow \begin{pmatrix} v_1 \\ v_2 \end{pmatrix} = \begin{pmatrix} 1 \\ -1.2496 \end{pmatrix}$$

Normalise the eigenvectors to get:

$$\begin{pmatrix} u_1 \\ u_2 \end{pmatrix} = \begin{pmatrix} 0.7808 \\ 0.6248 \end{pmatrix} \quad and \quad \begin{pmatrix} v_1 \\ v_2 \end{pmatrix} = \begin{pmatrix} 0.6248 \\ -0.7808 \end{pmatrix}$$

Calculate the modes at each time.

$$Time_i\text{: } MODE\ 1 = \begin{pmatrix} x_{1i} & x_{2i} \end{pmatrix} \begin{pmatrix} u_1 \\ u_2 \end{pmatrix} \quad and\ MODE\ 2 = \begin{pmatrix} x_{1i} & x_{2i} \end{pmatrix} \begin{pmatrix} v_1 \\ v_2 \end{pmatrix}$$

Time	x1	x2	E-vector 'u'	MODE 1	
1	-1.3	-1.4	0.7808	-1.8898	
2	0.9	1.7	0.6248	1.7649	
3	2.1	0.8		2.1395	
4	-1.7	-1.1		-2.0146	

Time	x1	x2	E-vector 'v'	MODE 2		MODEs 1+2
1	-1.3	-1.4	0.6248	0.2808		-1.6090
2	0.9	1.7	-0.7808	-0.7649		0.9999
3	2.1	0.8		0.6875		2.8270
4	-1.7	-1.1		-0.2034		-2.2180

We now have a new time series (MODE 1) which is representative of both the original time series (x_1 and x_2) and explains about 93% of the variability in the original time series. The MODE 2 time series can be thought of as a 'correction' to MODE 1. If we sum all of the modes (in this case just MODE 1 plus MODE 2) we get the best representation by a single time series of the original (two) time series. The two modes along with the original time series are shown in Figure 13–1. MODE 1 gives a good approximation, simultaneously, to BOTH of the original time series (of anomalies) x_1 and x_2.

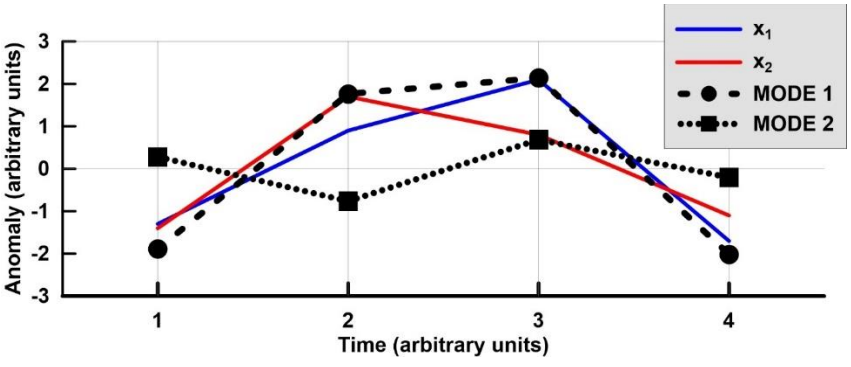


Figure 13–1. PCA modes and original data.

13.2.1. Checking

The eigenvalues and eigenvectors can be used to reconstruct the original time series, which provides a good check on the analyses. Eigenvalues are given by:

$$\lambda_1 = \underline{u}\,\underline{A}\,\underline{u^T} = (0.7808 \quad 0.6248)\begin{pmatrix}2.450 & 1.725\\ 1.725 & 1.675\end{pmatrix}\begin{pmatrix}0.7808\\ 0.6248\end{pmatrix} = 3.8305$$

$$\lambda_2 = \underline{v}\,\underline{A}\,\underline{v} = (0.6248 \quad -0.7808)\begin{pmatrix}2.450 & 1.725\\ 1.725 & 1.675\end{pmatrix}\begin{pmatrix}0.6248\\ -0.7808\end{pmatrix} = 0.2945$$

At time 'i': $x_{1i} = u_1\, MODE_{1i} + u_2\, MODE_{2i}$ and $x_{2i} = v_1\, MODE_{1i} + v_2\, MODE_{2i}$

Where:

Time	$MODE\ 1$	$MODE\ 2$	Eigenvectors	
			u	v
1	-1.8898	0.2808	0.7808	0.6248
2	1.7649	-0.7649	0.6248	-0.7808
3	2.1395	0.6875		
4	-2.0146	-0.2034		

Hence, we get (the cells shaded yellow are the original time series):

Time	u_1 x $MODE_{1i}$	u_2 x $MODE_{2i}$	v_1 x $MODE_{1i}$	v_2 x $MODE_{2i}$	x_1	x_2
1	-1.4754	0.1754	-1.1808	-0.2192	-1.3	-1.4
2	1.3780	-0.4780	1.1028	0.5972	0.9	1.7
3	1.6704	0.4296	1.3368	-0.5368	2.1	0.8
4	-1.5729	-0.1271	-1.2588	0.1588	-1.7	-1.1

A few points to note.

- We have a set of symmetric matrices at each time step. This means that the average of these matrices will also be symmetric.
- All symmetric matrices have real eigenvalues.
- The eigenvectors are all orthogonal i.e. perpendicular to each other.
- The variance of the data in each mode equals the (non-normalised) eigenvalue for that mode.
- The sum of the (non-normalised) eigenvalues (i.e. the trace of the matrix) equals the sum of the variances of each of the time series.

Because of the orthogonality property, PCA is sometimes referred to as Empirical Orthogonal Function (EOF) analysis.

13.3. Example 1: Axis rotation

In this example, we simply want to use PCA to rotate the axes of current meter data so that the east-west and north-south axes become the along shore and across shore axes.

Question: What is the optimal direction to rotate the axes to (approximately) align them with the local bathymetry?

Data used: This example uses the data from the Example in Section 5.2.8 (and 5.2.9). The time series of the east-west and north-south current components is reproduced in Figure 13–2.

Pre-processing: No pre-processing of the data is required.

Processing: A scatter plot of the original data is presented in Figure 13–3 (left hand plot). The latter plot suggests that the dominant direction is oriented slightly clockwise from the north-south axis. This alignment is approximately the same as that of the coastline (or local bathymetry, Figure 13–3, right hand plot). Using the information from Section 13.1, the magnitude of this orientation (the direction) can be quantified.

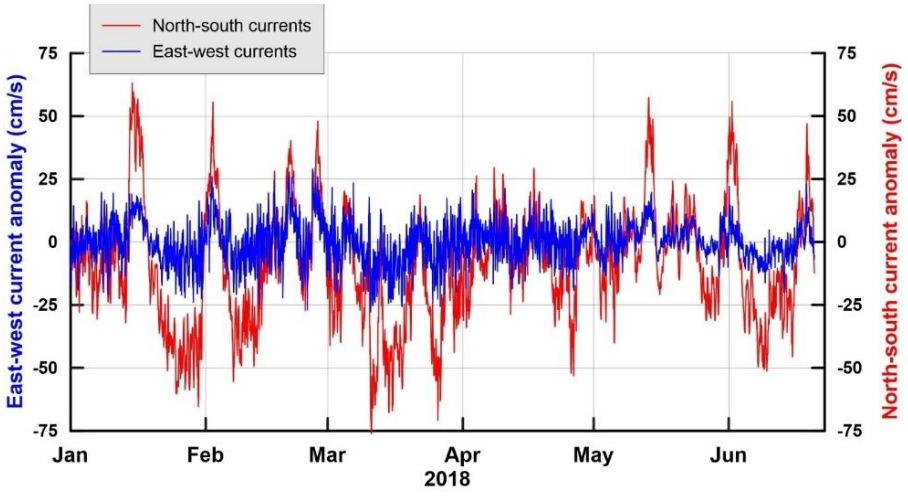


Figure 13–2. Current components – hourly data over a six month period.

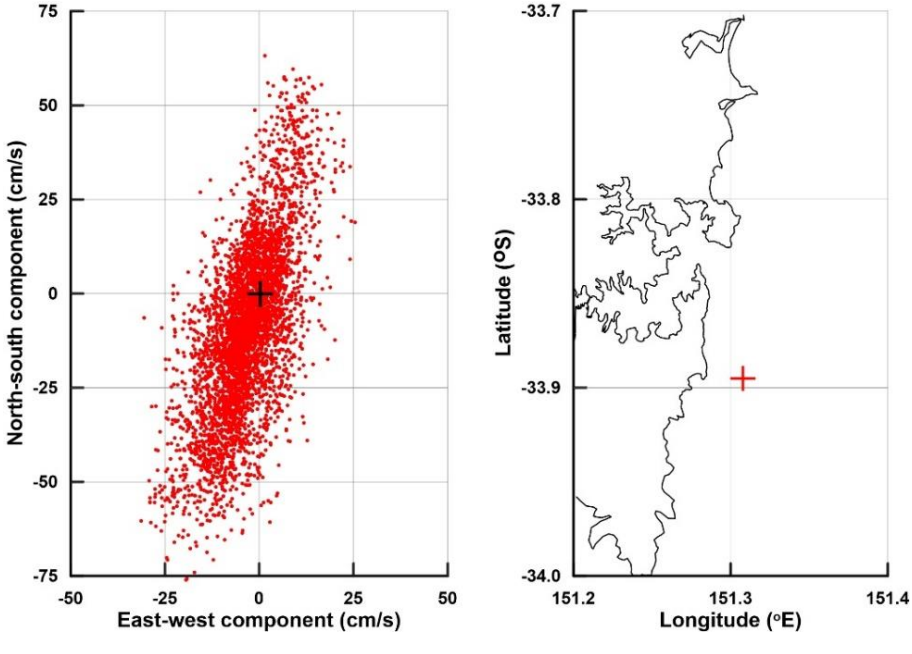


Figure 13–3. Scatter plot of current components (left) and location of the current meter (right).

The direction of the first principal axis (θ_1) is given by:

$$\theta_1 = 0.5 \times tan^{-1}\left[\frac{2 \times covariance(EW, NS)}{variance(EW) - variance(NS)}\right]$$

Where EW and NS refer to the east-west and north-south components of the currents, respectively.

For this example, the covariance is 121.3 $(cm/s)^2$ while the variances of the north-south and east-west components are 525.5 $(cm/s)^2$ and 67.9 $(cm/s)^2$ respectively. The direction of the first principal component is:

$$\theta_1 = 0.5 \times tan^{-1}\left[\frac{2 \times 121.3}{67.9 - 525.5}\right] = -14.0^o$$

A positive direction is defined as anti-clockwise from the 'north' direction.

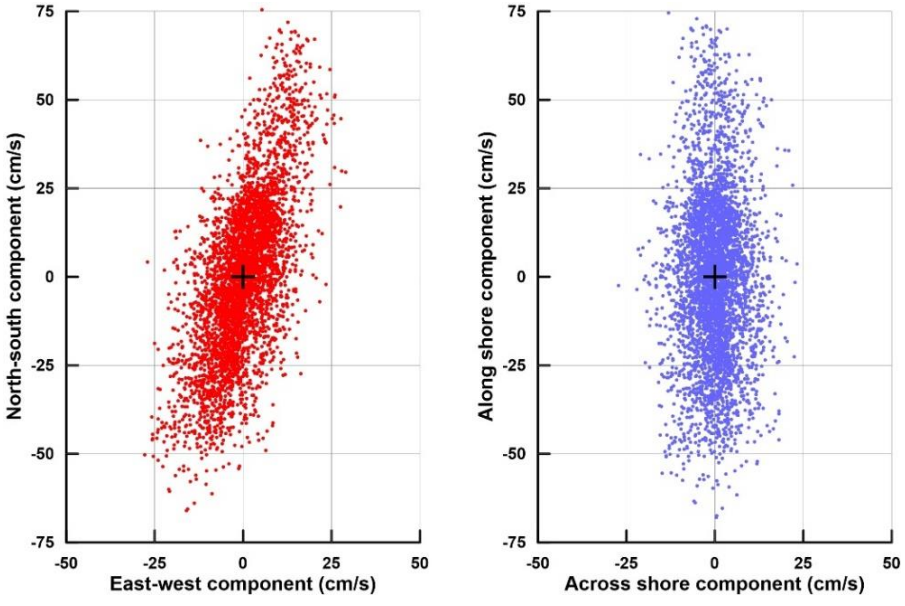


Figure 13–4. Scatter plot of east-west and north-south current components (left) and the along-shore and across-shore current components (right).

Description: The result of the axis rotation (rotated clockwise from north by 14º) is shown in Figure 13–4. The ambiguity associated with the orientation of the axes when undertaking our cross-spectral analysis estimates is now removed. One axis is now the axis along which the variance is maximised and its orthogonal axis is the axis along which the variance is minimised.

13.4. Examples

13.4.1. Example 2: Dimension Reduction – temperature

In this example, we take a set of temperature data and subject it to a PCA to reduce the amount of information required to for further analysis.

Question: How many PCA modes are required to account for 90% of the variability in the data set?

Data used: The data set comprises temperature data from 14 levels in the water column, each temperature sensor vertically separated by about 4 m (Figure 13–5). (For clarity, only 5 of the 14 sensors have been plotted). Data are sampled every 5 minutes and there is a total of 6,016 records (about three weeks).

Pre-processing: Anomalies were produced for each of the 14 temperature data sets. This was achieved by determining a cubic polynomial for each data set and removing it from the original data. (A visual inspection of the data suggested the cubic polynomial, although other curves could also be used. We wanted a consistent function across all 14 data sets). The resultant anomalies are shown in Figure 13–6 (again for clarity, only 5 sets of anomalies, corresponding to the data in Figure 13–5, are shown).

Description: PCA was carried out on the 14 sets of temperature anomalies, 5 of which are shown in Figure 13–6. Normalised eigenvalues and eigenvectors result.

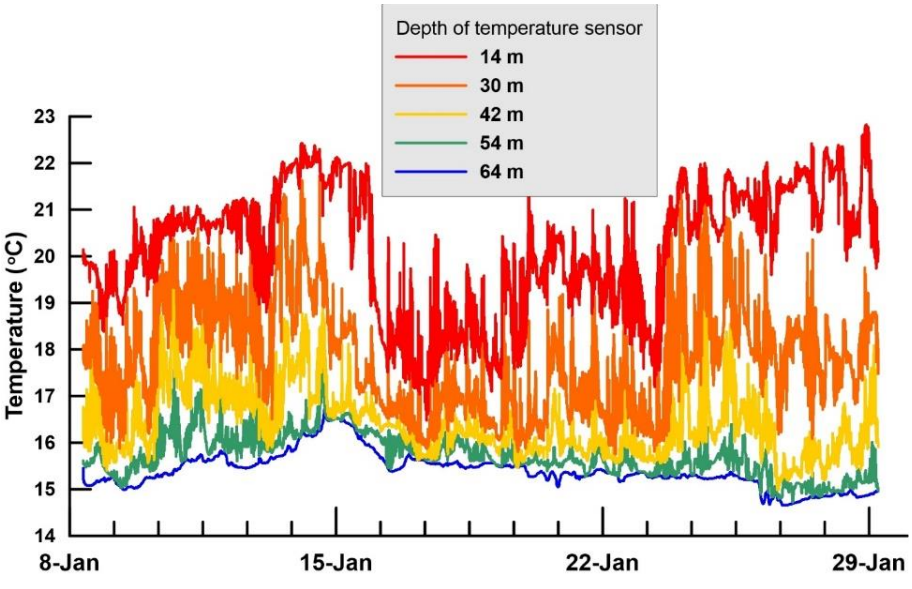


Figure 13–5. Time series – temperature data.

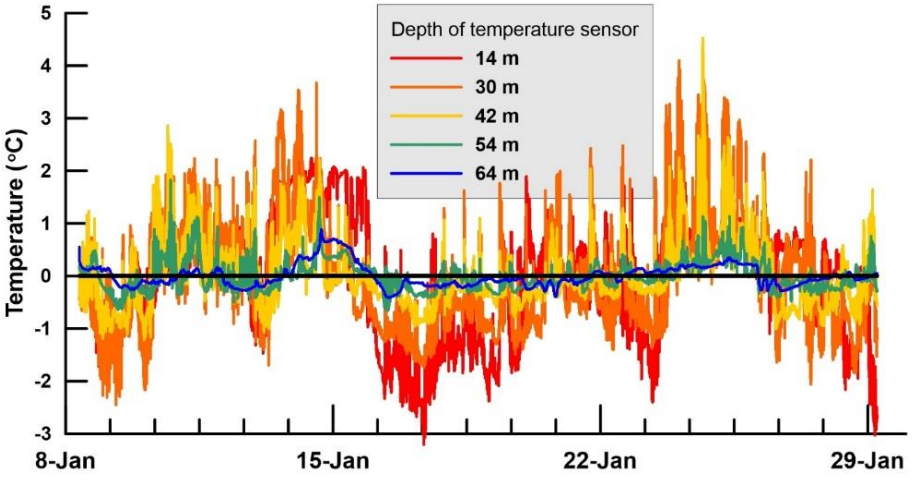


Figure 13–6. Time series – temperature anomalies.

If we plot the normalised eigenvalues as a function of the (eigen) mode we generate a 'scree' plot (Figure 13–7). The normalised eigenvalues are

simply the proportional contribution of each eigenvalue to the total variance. The greatest contribution occurs in MODE 1, the next greatest contribution occurs in MODE 2 and so on. From this plot it is clearly seen that about 80% of the variance of the whole system (i.e. the sum of the variances in each of the 14 temperature data sets) occurs in MODE 1. About 10% of the overall variance occurs in MODE 2, while the remaining modes contribute very little to the overall variance.

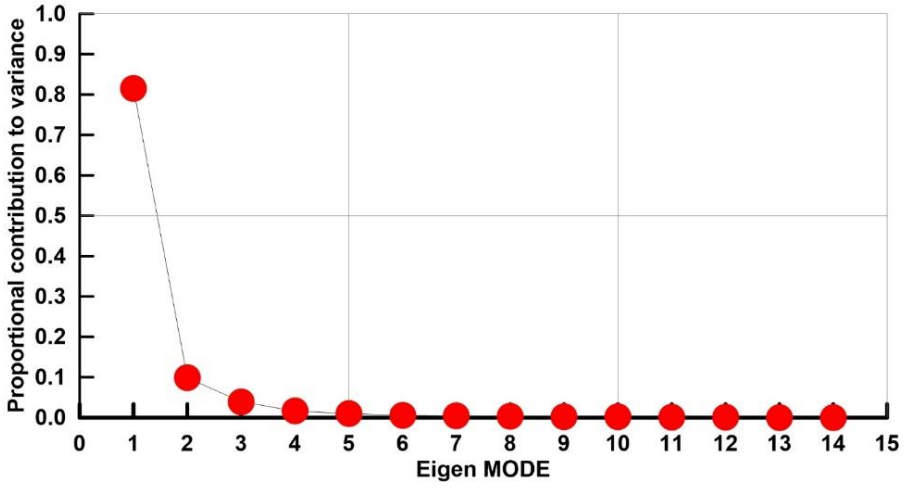


Figure 13–7. Scree plot.

Therefore, more than 90% of the variability of the whole system occurs in the first two MODES. If we are willing to accept that we will only retain 90% of the overall variability in our system, then, instead of having to analyse 14 data sets (and look at all of their interactions), we only have to analyse 2 new data sets. For example, the two new data sets could be analysed using cross-spectral analysis (Chapter 5). It is important to stress that these are new data sets, they do necessarily correlate with any of the original time series. In a sense, each MODE takes a bit of information (the most important bits first) from all 14 data sets and pushes as much as possible (within the PCA 'rules') into a single, new data set. However, if much variability is present in one of the data sets, then the MODE 1 data set may resemble that data set. In general, this does not occur.

Time series for the first three eigenvectors are shown in Figure 13–8. Most of the variance lies in MODE 1 and decreasing amounts of the variance occurs in each of the subsequent modes. Note that the temperatures and temperature anomalies are applied to a specific depth. However, the modes do not correspond to the depths – they are an overall representative of the variance at all depths.

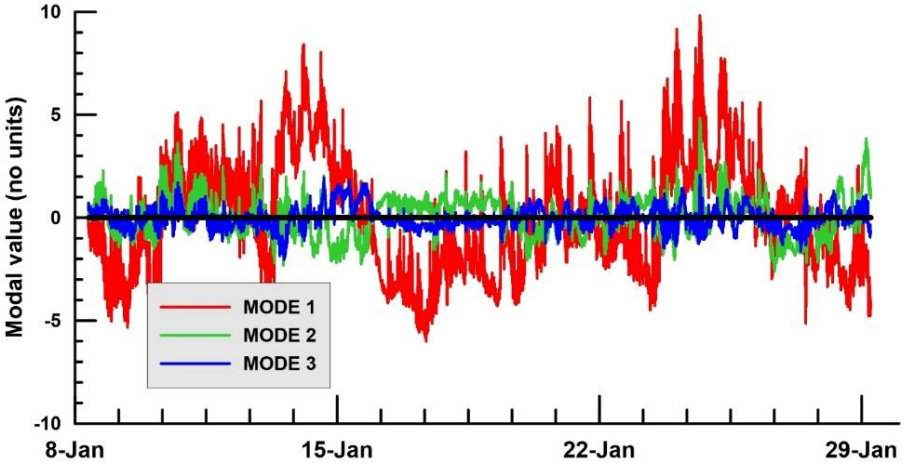


Figure 13–8. Time series – PCA MODES.

The MODE 1 data set looks broadly similar to that of the temperature anomalies from those data sets obtained from temperature sensors near the sea surface.

The variance of each of the temperature anomalies can be plotted as a function of water depth (Figure 13–9, blue circles). The sum of these 14 variances is 9.36 $^{o}C^{2}$. Most of the variance occurs in the temperature data sets closest to the water surface. There are only small differences in the variances of the top 4 or 5 temperature data sets, each close to 1.5 $^{o}C^{2}$. Superimposed on the same plot are the variances from each of the 14 eigen modes. Almost 80% of the variance is contained in MODE 1 and about 10% of the variance is contained in MODE 2. The sum of the remaining variances totals less than 10%. However, the sum of the variances from the 14 MODEs totals 9.36 $^{o}C^{2}$, the same as that for the

temperature anomalies. The sum of the variances is the same, irrespective if we're using temperature anomalies or eigen modes. However, it is the distribution of this variance that differs between the two types of data.

To answer our question, 90% of the variability of the system of temperature data can be achieved using the first two PCA modes.

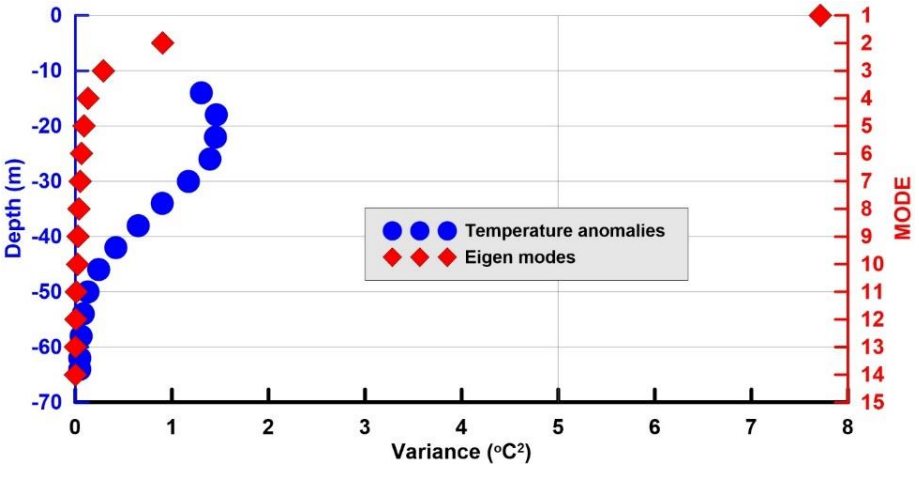


Figure 13–9. Variances – time series temperature anomalies and PCA eigen MODES.

13.4.2. Example 3: Dimension Reduction – multiple variables

The second dimension reduction example uses 14 different water quality variables from a single location in the coastal ocean. (It is coincidence that there happens to be the same number of variables as in the previous example).

Question: How many PCA modes are required to account for 90% of the variability in the data set?

Data used: The data set comprises the following water quality variables: temperature, salinity, dissolved oxygen, pH, turbidity, total nitrogen, ammonia, nitrate, nitrite, total phosphorus, orthophosphate, chlorophyll-a, faecal coliforms and enterococci. Data were collected on 14 occasions (again, this number is just a coincidence) over 12 months, from the surface waters at a single location in the coastal ocean. There are too many different variables in this example to clearly plot the original data. The two bacteria (faecal coliforms and enterococci) we log-transformed prior to any analyses.

Pre-processing: Anomalies were produced for each of the 14 variables. This was achieved by determining a fourth order polynomial for each data set and removing it from the original data. (A visual inspection of the data suggested the fourth order polynomial, although other curves could also be used. Again, for consistency, the same order polynomial type was used across all 14 variables). The resultant anomalies are shown in Figure 13–10. No units were used on the vertical axis in Figure 13–10: the units for each variable are given in the legend associated with the plots. When superimposed, the plots appear messy and give little indication of any consistent pattern.

Description: PCA was carried out on the 14 sets of variables using their anomalies. Normalised eigenvalues and eigenvectors result. The resulting scree plot is shown in Figure 13–11. In the previous example, the scree plot was dominated by the first mode. In Figure 13–11 we note that high variance is spread over many modes: MODE 1 contributes about 40% of the total variance, MODE 2 about 30%, MODE 3 about 16%, MODE 4

about 7% and MODE 5 about 5%. To answer our question, we need to use the first 4 modes to include 90% of the total variance. Given the variability of the anomalies (Figure 13–10), this result is, perhaps, not unexpected.

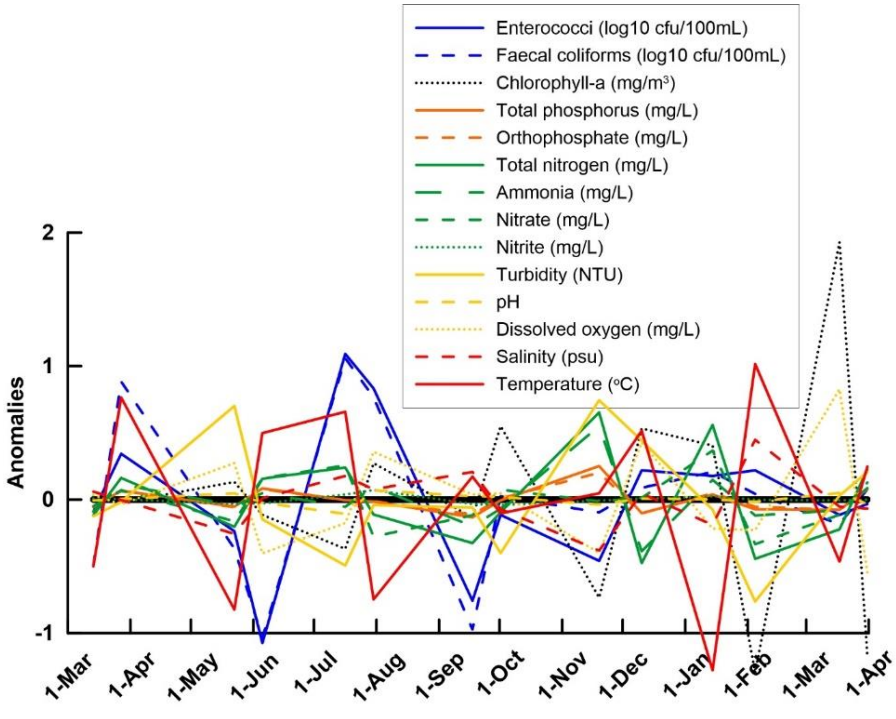


Figure 13–10. Time series data – anomalies.

If we need to include at least 90% of the overall variability in our system, then we need to include the results from the first 4 modes. We will then need to analyse 4 new data sets – better than having to analyse 14 data sets, but still a considerable about of work. Including the first 5 modes will result in >98% of the variance being included in the new data sets.

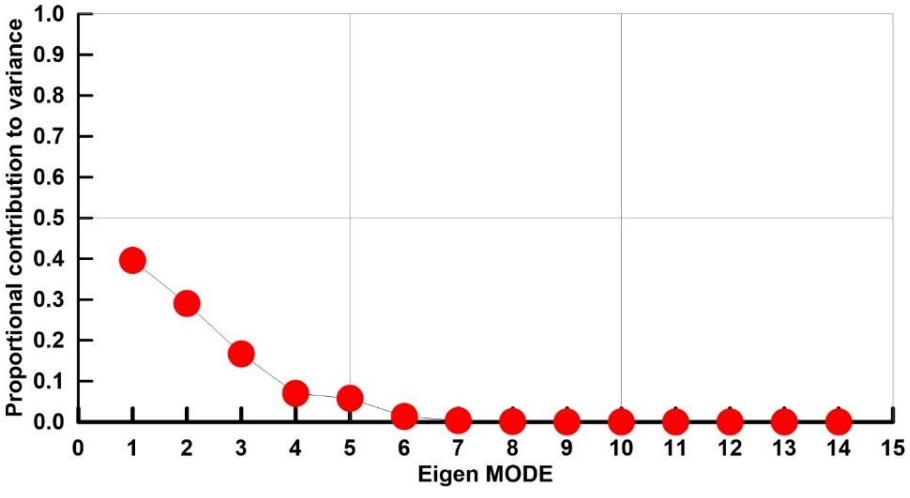


Figure 13–11. Scree plot.

Time series for the first five eigenvectors are shown in Figure 13–12. Unlike the previous example, MODE 1 does not clearly dominate the other modes. The first 4 or 5 modes all contain substantial amount of the variance. This may mean using PCA on this data set does not result in substantial efficiency gain.

The variance of each of the temperature anomalies is plotted as a function of water depth (Figure 13–13, blue circles). The sum of these 14 variances is 2.39 units^2. Most of the variance occurs in variables 1 (temperature), 12 (chlorophyll-a), 13 (faecal coliforms) and 14 (enterococci). While PCA is the most efficient arrangement of the variances, in this case the benefit of PCA is much less obvious than in the previous example. The first 5 PCA modes contain substantial amounts of the variance.

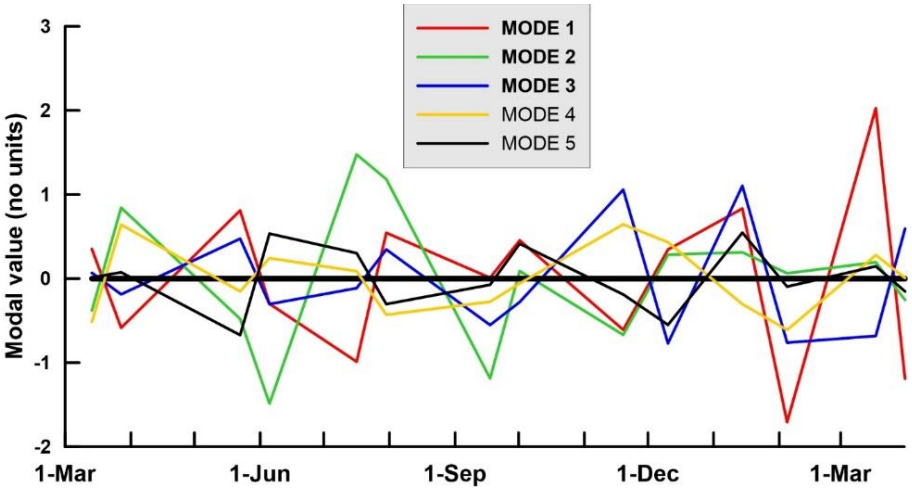


Figure 13–12. Time series – PCA MODES.

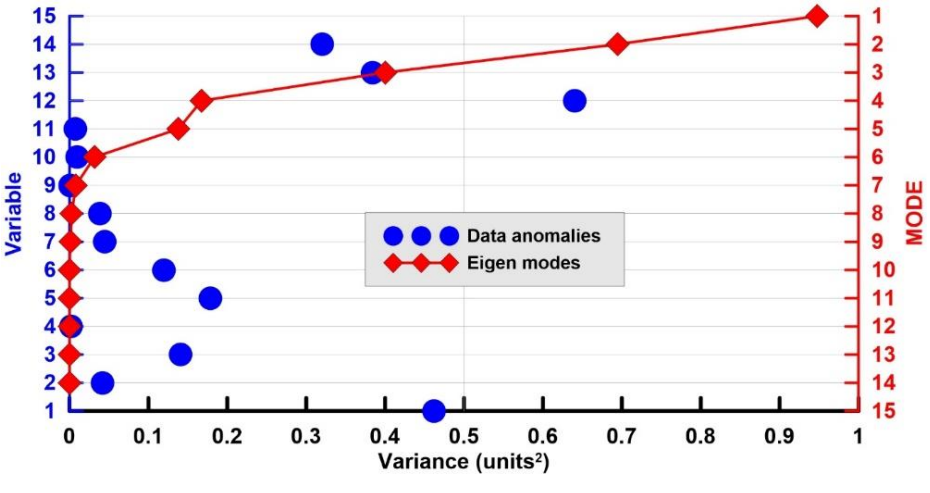


Figure 13–13. Variances – time series data anomalies and PCA eigen MODES.