1

What Is Fourier Analysis?

Fourier analysis can be defined as mathematical analysis based on the representation of a complicated waveform, *e.g.* a time-series, as the linear combination of a specific set of sinusoids. This started with Fourier series: "Fourier theory" is the general term used to describe the branch of mathematics which generalises and extends Fourier series beyond its original application.

Fourier series were initially developed in the early nineteenth century by, and named for, J-B J Fourier (1768–1830). The problem Fourier was trying to solve had nothing to do with periodic features *per se* but was a heat transfer problem in a metal plate. In brief, and skipping over quite a lot of important mathematical concepts that we will return to, Fourier could solve the same problem for a plate bounded by a sinusoid. From there, he reasoned that if he could exactly represent the boundary of the plate in question as a linear combination of sinusoids of different periods then the solution to the plate problem would be the linear combination of all the individual sinusoidal solutions. This was the first time that anyone had formalised the expansion of functions as trigonometrical series of sinusoids (*i.e.* sines and cosines).

Fourier's ideas did not gain immediate acceptance owing to their unorthodoxy and unfamiliarity and have been further developed and formalised since he first proposed them. However, Fourier's original insight that it is possible to use periodic functions to represent other functions carries forwards into general Fourier theory.

In essence, the Fourier transform provides us with a different way of viewing the world around us by transforming data from a sequential domain, such as time (and interval), to a wave domain, such as frequency (and phase). Sometimes, and particularly for periodic phenomena, the wave domain is easier to work with and more informative. All equal-interval sequential data that arise in real-life are (discrete) Fourier transformable and, if there are periodic features in those data, the Fourier transform will give us frequency and phase information. There are other, more advanced and specialised techniques for doing this, but many are based on or related to the Fourier transform, so an intuitive understanding of the Fourier transform and the necessary mathematics gives a solid basis for understanding those more advanced techniques.

Use of the Fourier transform extends beyond its relatively straightforward application in the identification of periodic features in data, *e.g.* time-series. It has applications in technologies that are used everyday and on which we depend: in essence pretty much everything which involves transmission or recording of data, including audio and video files in the entertainment and communications industries, involves use of the Fourier transform, or something closely related, in some form. In addition, there are countless scientific and mathematical applications beyond time-series analysis.

I first encountered the Fourier transform and general Fourier theory in my undergraduate mathematics courses. Based on my experience since then, including much tutoring and advising of non-specialists, my opinion is that the Fourier transform is best understood intuitively, at least by non-specialists. The mathematics can be, initially at least, more than a little daunting, and it is that initial dauntingness that I am addressing in this book.

1.1 What Does This Book Set Out to Do?

This book, and the European Geosciences Union short-courses it evolved from, were prompted by questions along the following general lines:

- What is the FFT?
- What does the FFT do?
- Why do I get a column of N strange-looking numbers out when I put a column of N 'ordinary' numbers in?
- Why do I get complex numbers out when I put real numbers in?
- OK, I know what the Fourier transform does but how do I interpret what the FFT gives me?

So, this book tries to answer such questions and give the reader an understanding of the underpinning properties of the discrete Fourier transform (DFT), as generally implemented via a Fast Fourier Transform (FFT) algorithm, in the context of time-series analysis for periodic features. With regard to the two descriptors, *i.e.* DFT ('discrete') and FFT ('fast'), although not strictly identical in meaning the terms DFT and FFT are to a great extent interchangeable in practice. In the early chapters where I am developing the theory, I will refer to the DFT as the theory applies to the discrete nature of the data and the transform. In the latter chapters where I am referring to the use of the specific tool(s) coded in software, I will refer to the FFT.

This book is a primer. It is an introduction to Fourier and associated analysis and theory which underpins many of the concepts and techniques used in time-series analysis such as least-squares spectral analysis (LSSA), wavelets, singular-spectrum analysis (SSA) and empirical mode decomposition (EMD). As a primer, it is not a complete discussion of the subject and is not intended to be. Also, it is not intended to be read in isolation either from other (introductory) books on time-series analysis in the geosciences (and other sciences) or, indeed, other books on Fourier analysis and theory. Although it is explicitly aimed at geoscientists, and the examples and exercises are oriented towards the geosciences, it should be suitable for other scientists and researchers who want a very introductory text to the field.

As a book on Fourier analysis, it is a mathematics book but is one which focuses on the identification of periodic features in real data – real time-series – and estimating their effect-size and significance. These are aspects of Fourier theory which are important in time-series analysis but which are often passed over by the majority of books on Fourier theory, which emphasise more abstract aspects less relevant to time-series analysis and specialise in, for example, digital signal processing, audio and video processing and allied fields. Some of the content of this book will be familiar to people who have specialised in fields such as signal analysis but that same material can be far from familiar to non-specialists, including – broadly, in this context – early career geoscientists. Also, although effect-size and, by implication, statistical significance are touched on in some other books, these aspects are generally not the focus of those books and it is rare to find a systematic consideration consistent with statistical correlation and linear regression.

In keeping with these aims, while the mathematical reasoning, logic and flow of ideas and concepts are presented, proofs are not presented: readers with a need or desire for proofs are referred to books listed in the Bibliography (Further Reading) and others recommended to them. This is not to deprecate mathematical proof: proof is vital to the development of mathematics and mathematical proofs are often elegant and illuminating of themselves for those who view the world mathematically. However, that does not include everyone, and there will be some whose overriding and pressing need is to resolve immediate questions in their branch of science. That said, I encourage readers to refer to proofs when their time permits as they will quite likely find something that I cannot fully present or develop here, or something that gives them insights to problems that have been perplexing them.

1.2 Choice of Software

I decided to use R, the Free and Open-Source Software (FOSS) for statistical computing for this book. This is because R is open-source and freely available for Microsoft, Apple and the open-source Linux and BSD operating systems, making it a good choice because it means that I can work examples using one piece of software knowing that it is generally available to readers. Also, there are many specialist add-on library packages for R which contain algorithms/code for more advanced time-series analysis, *e.g.* there are packages for the Lomb–Scargle periodogram, wavelets, empirical mode decomposition, singular-spectrum analysis, amongst others.

We will use the included datasets in the `datasets` package, usually included with the R base package, and the add-on library packages `astrodatR`, `astsa`, `EMD`, `RSEIS`, `TideHarmonics`, `TSA` and `tseries`. We will, in addition, use the `lomb` package for its implementation of the Lomb–Scargle periodogram. When loading library packages for datasets, look at the analytical techniques included, and look at those in other time-series packages, too: one of those techniques might be suited for a dataset you are investigating. Also, the Bibliography (Online Resources) includes some URLs of websites that host datasets and some of those websites are the original sources of the datasets in the R library packages. In addition, it is possible to export datasets from R, *e.g.* in tab-delimited or comma-separated text files, for loading into other software.

I have deliberately kept my use of R code/commands in the book as basic as possible. This is for two reasons. First, to make worked examples as accessible as possible to readers who are more familiar with other software. Second, I avoid using R-unique ways of doing things so as to keep the commands as generic as possible for readers who wish to translate commands to other software. Although syntax varies between software packages, comparable software packages will have commands for performing the FFT, correlating and linearly regressing vectors of numbers, adding rows and columns to arrays and matrices, ordering arrays and matrices, plotting basic line and point graphs, and so on. For example, the basic forward and inverse FFT commands in R are, for data-vector `z`, `fft(z, inverse = FALSE)` and `fft(z, inverse = TRUE)` respectively: in Matlab and Octave (open-source, Matlab-like) the corresponding commands are `fft(z)` and `ifft(z)`, and in Scilab (open-source) the corresponding commands are `fft(z, −1)` and `fft(z, 1)`.

With regard to R and the worked examples, I am assuming that you have a working command of basic R syntax, *e.g.* downloading, installing and loading library files, loading data-frames, referencing rows and columns in data-frames and matrices, basic plotting and saving data-objects and workspaces. However, don't worry: there are plenty of books and online resources for R, not least the documentation available from the R-project website and mirrors. Commands for the more specialised operations such as cross- and auto-correlations and FFT are explained as they are introduced. Also, R has comprehensive help files, including help files on datasets. If you are an expert R-user, you will identify where you can make use of R-unique commands.

1.3 Structure of the Book

The book opens with a consideration and extension of statistical correlation and linear regression in Chapter 2. Correlation and regression should be familiar from mathematics and statistics courses in undergraduate science programmes, and so this gives an accessible starting-point for the rest of the book, but also a return point for statistical effect-size and significance in the final chapter. Also, covariance and correlation are closely related to the Fourier transform, although the majority of the theory is beyond the scope of this book, and to LSSA periodogram approaches in the final chapter. Chapters 3 and 4 cover most of the core Fourier and discrete Fourier theory, which I have aimed to present in an intuitively accessible, if not always fully rigorous, manner that is suited to the time-series context.

Chapter 5 consists of worked examples to demonstrate the use of the FFT, *e.g.* amplitude and power spectra, and also to illustrate a systematic approach to calibrating the frequencies and periods in the FFT output spectrum. Chapter 6 considers the limitations of the FFT, *i.e.* what the FFT cannot do in the context of identifying periodic features, as these are aspects of the FFT that can be overlooked. Chapter 6 also describes how to construct dummy time-series to help calibrate FFT spectra, padding/truncating time-series to 'tune' for frequencies of interest, and overviews work-arounds for time-series with missing data.

The last three chapters address specific issues that arise in the preceding chapters. Chapter 7 considers non-stationary data, *i.e.* essentially time-series with varying frequency composition, and introduces basic spectrogram approaches for time–frequency analysis. Chapter 8 presents some basic characterisations of noise as likely to occur in the FFT spectra of geoscience time-series. All real time-series are noisy, *i.e.* contain random fluctuations which are not part of any deterministic or systematic 'signal' content, and, indeed, an underlying noise spectrum in a time-series can provide a reference for evaluating the statistical effect-size of 'signal' content. This provides a starting point for Chapter 9 which considers basic periodograms, as examples of LSSA techniques, and statistical effect-size and significance, both of which also link directly back to Chapter 2.

1.3.1 Examples and Exercises

With regard to examples and exercises, there is no substitute for 'playing with data' relevant to an investigation that you have in hand or are about to embark on. I cannot hope to cover all eventualities and if you have some data to 'play with', and can adapt the approaches illustrated in the worked examples to those data, then doing that will be at least as useful as working through exercises that I have included. To that end, I have deliberately kept examples to a sensible minimum which illustrate (a) typical features of geoscience time-series that you might reasonably expect to encounter and (b) generally applicable systematic ways of investigating time-series, accompanied by end-of-chapter exercises to build on the worked examples. The worked examples are presented in box-outs in the text and if you work through those then you will obtain the results and plots discussed in the text. Some of the examples and exercises are developed and extended over successive chapters so you might find it helpful to save data objects and/or workspaces as R data-files as you work through, for subsequent reloading.

1.4 What Previous Mathematics Do You Need?

This is a book on Fourier analysis, albeit an introductory one that is highly focused on one specific task, *i.e.* the detection of periodic features in time-series. Even though its core aim is to present the basic mathematical framework of the Fourier transform in accessible intuitive terms, it has to assume a starting level in terms of mathematics and statistics that the reader has at the outset. Broadly, these are:

- Basic statistics. We will start with covariance, correlation and linear regression to set the scene for the Fourier theory and then return to these in order to investigate effect-size and significance, *i.e.* correlation coefficient (R), coefficient of determination (R^2) and statistical significance (p-value). We will also take some of these ideas forward to briefly consider periodograms, which are closely related to the discrete Fourier transform.
- Fourier series. Fourier series is the starting point for Fourier theory and an understanding of Fourier series sets a solid foundation. We start our consideration of Fourier theory with a statement of Fourier series and a

review of the key concepts and properties that carry forward into more general Fourier theory, including the discrete Fourier transform.

- Circular (trigonometric) functions. Fourier series are explicitly dependent on the properties of sines and cosines as circular functions, and their derivatives and integrals, and their relationship with complex exponentials.
- Complex numbers. Complex numbers are crucial to the standard notation used for Fourier theory. Complex numbers are arguably not strictly necessary for basic Fourier theory but it would be much, much more difficult to describe, understand, use and develop without them. Complex numbers have the general two-part form $z = x + iy$ where i is the imaginary number, *i.e.* that 'imaginary' number defined according to its property $i^2 = -1$, and x and y are the real and imaginary parts respectively, and both are real numbers. Some branches of engineering and the sciences use j for the imaginary number rather than i as used in this and many other books.
- Linear algebra. Linear algebra, *e.g.* matrix-vector algebra, is not necessary for standard/continuous Fourier theory but the Fast Fourier Transform, which is what we all basically use when Fourier-transforming in practice using digital computers, is essentially an orthogonal linear transformation (linear mapping) from a time (or space) basis (frame of reference) to a frequency basis. Using concepts from linear algebra simplifies some of the material and illustrates the optimisations inherent in Fast Fourier Transform algorithms. Also, vector arithmetic is crucial to the classical (Schuster) periodogram.
- Mathematical notation. Mathematics books can be very notation-heavy. I have tried to have as light a touch as I consider useful but you will need a general familiarity with, for example, the use of lower-case and upper-case letters, Greek and Roman letters to represent mathematical constants, operators and variables, integral and summation symbols, use of superscripts/indices and subscripts.

These are all subjects and aspects of mathematics which are often included in secondary/high school A-Level, Scottish Higher, International Baccalaureate and equivalent mathematics syllabuses and, with the possible exception of Fourier series, are often included in mathematics and statistics courses taken as part of undergraduate degrees in science subjects. Don't worry if some of these are unfamiliar or out-of-practice as there are very many textbooks and online resources available at secondary/high school level. Also, many universities have support units for mathematics and statistics, as well as mathematics and statistics departments, which will be able to help.

Having outlined the sort of mathematics background that the book starts from, let us move to Chapter 2 and start with a review and extension of covariance-based techniques such as correlation and linear regression before moving on to Fourier theory. As well as being a starting point for the main subject matter of this book, variations of correlation and linear regression can provide a lot of information regarding periodic features in time-series in their own right.