Artificial Neural Network approaches and Compressive Sensing techniques for stochastic process estimation and simulation subject to incomplete data

Thesis submitted in accordance with the requirements of
the University of Liverpool for the degree of Doctor in Philosophy
by Liam Anthony Comerford

August 2015

Declaration

I hereby confirm that the results presented in this dissertation are from my own work, developed in conjunction with my supervisors, and that I have not presented anyone else's work as my own and that full and appropriate acknowledgements have been given where references have been made to the work of others.

Liam Anthony Comerford

August 2015

Acknowledgements

Great appreciation and gratitude for the help and support are extended to the following persons who have contributed in making this work possible.

Professor Michael Beer and **Dr. Ioannis Kougioumtzoglou** for their continued guidance throughout the project. They have both shown great patience and invested significant time and effort into facilitating my studies, for which I consider myself very fortunate.

Dr Edoardo Patelli and **Dr. Matteo Broggi** for helping to prepare my codes for cluster computing, and fixing numerous associated problems, despite having no direct investment in my work. The help was greatly appreciated.

Mr Adam Mannis for his general guidance and support. He knows all there is to know about writing reports and how to present them.

Mr Marco DeAngelis for always being around to bounce ideas off and acting as an efficient alternative to the Matlab help files when I needed a function quickly.

Dr. Ioannis Mitseas for his happy-go-lucky optimistic charm.

I would like to thank my Family for their support, even though they don't really know what I'm up to; and also my friends, particularly John Connor and Matthew Kent for providing frequent ant-related distractions.

I would also like to give considerable thanks to EPSRC for providing my research studentship.

Abstract

This research is themed around development of tools for discrete analysis of stochastic processes subject to limited or missing data; more specifically, estimation of stochastic process power spectra from which new process time-histories may be simulated. In this context, the author proposes three novel approaches to power spectrum estimation subject to missing data which comprise the main body of this work. Of particular importance is the fact that all three approaches are adaptable for use in both stationary and evolutionary power spectrum estimation. Numerous arrangements of missing data are tested to simulate a range of possible scenarios to demonstrate the versatility of the proposed methodologies.

The first of the three approaches uses an artificial neural network (ANN) based model for stochastic process power spectrum estimation subject to limited / missing data. In this regard, an appropriately defined ANN is utilized to capture the stochastic pattern in the available data in an "average sense". Next, the extrapolation capabilities of the ANN are exploited for generating realizations of the underlying stochastic process. Finally, power spectrum estimates are derived based on established frequency (e.g. Fourier analysis), or versatile joint time-frequency analysis techniques (e.g. harmonic wavelets) for the cases of stationary and non-stationary stochastic processes, respectively. One of the significant advantages of the approach relates to the fact that no *a priori* knowledge about the data is assumed.

The second approach uses compressive sensing (CS) to solve the same problem. In this setting, further assumptions are imposed on the nature of the underlying process of interest than in the ANN case, in particular that of sparsity in the frequency domain. The advantages being that when compared to ANN, significant improvements in efficiency and accuracy are achieved with increased reliability for larger amounts of missing data. Specifically, first an appropriate basis is selected for expanding the signal recorded in the time domain. As with the ANN approach, Fourier and harmonic wavelet bases are utilized. Next, an L1 norm minimization procedure is performed for obtaining the sparsest representation of the signal in the selected basis. Further, an adaptive basis procedure is introduced that significantly improves results when working with stochastic process record ensembles.

The final approach is somewhat different, in that it aims to quantify uncertainty in power spectrum estimation subject to missing data rather than provide deterministic predictions. By relying on relatively relaxed assumptions for the missing data, utilizing fundamental concepts from probability theory, and resorting to Fourier and harmonic wavelets based representations of stationary and non-stationary stochastic processes, respectively, a closed-form expression is derived for the probability density function (PDF) of the power spectrum value corresponding to a specific frequency. Numerical examples demonstrate the large extent to which any given single estimate using deterministic methods, even for small amounts of missing data, may be unrepresentative of the target spectrum. In this regard, this probabilistic approach can be potentially used to bound deterministic estimates, providing specific validation criteria for missing data reconstruction.

Contents

List of Figures

List of publications

A list of publications by the author that were produced as a result of the work documented in this thesis are given here.

Conference proceedings

Comerford, L.A; Kusanovic, D.; Kougioumtzoglou, I.A.; Jensen, H.A.; Beer, M., "Structural system response and reliability analysis subject to incomplete earthquake records" *Proceedings of the 12th International Conference on Applications of Statistics and Probability in Civil Engineering (ICASP).* (July 2015)

Comerford, L.A.; Kougioumtzoglou, I.A.; Beer, M., "Compressive sensing based power spectrum estimation from incomplete records by utilizing an adaptive basis" *Proceedings of the IEEE Symposium Series on Computational Intelligence (SSCI), 2014 IEEE Symposium on Computational Intelligence for Engineering Solutions (CIES).* (December 2014)

Comerford, L.A.; Kougioumtzoglou, I.A.; Beer, M., "Uncertainty quantification in power spectrum estimation of stochastic processes subject to missing data" *Proceedings of the 2nd International Conference on Vulnerability, Risk Analysis and Management.* (July 2014)

Comerford, L.A.; Kougioumtzoglou, I.A.; Beer, M., "A compressive sensing based approach for estimating stochastic process power spectra subject to missing data Spec uncertainty" *Proceedings of the 9th International Conference on Structural Dynamics (EURODYN), Reliability and robustness of dynamic systems.* (July 2014)

Comerford, L.A.; Kougioumtzoglou, I.A.; Beer, M., "A compressive sensing based approach for evolutionary power spectrum estimation subject to missing data" *Proceedings of the 7th International Conference on Stochastic Mechanics (CSM).* (June 2014)

Comerford, L.A.; Kougioumtzoglou, I.A.; Beer, M., "An artificial neural network based approach for power spectrum estimation subject to limited and/or missing data" *Proceedings of the 11th International Conference on Structural Safety & Reliability (ICOSSAR), Safety, Reliability, Risk and Life-Cycle Performance of Structures and Infrastructures, pages 1083–1090, ISBN: 978-1-138-00086-5.* (June 2013)

Comerford, L.A.; Kougioumtzoglou, I.A.; Beer, M., "An artificial neural network based approach for power spectrum estimation and simulation of stochastic processes subject to missing data" *Proceedings of the 2013 IEEE Symposium Series on Computational Intelligence (SSCI), 2013 IEEE Symposium on Computational Intelligence for Engineering Solutions (CIES), pages. 118-124, 10.1109/CIES.2013.6611738.* (April 2013) - Best Student Paper Award

Journal articles

Comerford, L.A.; Kougioumtzoglou, I.A.; Beer, M., "A compressive sensing based approach for evolutionary power spectrum estimation subject to missing data" - *Probabilistic Engineering Mechanics.* (Under review, submitted June 2015)

Comerford, L.A.; Kougioumtzoglou, I.A.; Beer, M., "Uncertainty quantification in power spectrum estimation of stochastic processes subject to missing data" – *International Journal of Sustainable Materials and Structural Systems (IJSMSS).* (Under review, submitted June 2015)

Comerford, L.A.; Kougioumtzoglou, I.A.; Beer, M., "An artificial neural network approach for stochastic process power spectrum estimation subject to missing data" *Structural Safety 2014: Special issue on Engineering Analysis with Vague and Imprecise Information. Volume 52, Part B, Pages 150-160, ISSN 0167-4730.* (January 2015)

Pedagogical work

Comerford, L.A.; DeAngelis, M.; Mannis, A.; Beer, M.; Kougioumtzoglou, I.A., "An open approach to educational resource development, with a specific example from structural engineering" *42nd Annual Conference of the European Society for Engineering Education (SEFI)* . (September 2014)

Chapter 1

Introduction

1.1 Rationale and objectives

Probabilistic engineering simulations often require models for the engineering system excitation/response processes. For stochastic process model based Monte Carlo simulations to be reliable, modelling/estimation techniques often require a significant amount of data and/or some prior knowledge of the underlying physics of the process; i.e. the more data on which a model is built, the more statistically accurate the simulation is likely to be. The data collected is of course dependant upon the process that needs to be modelled. For system excitation (particularly civil engineering structures) we are often interested in environmental stochastic processes. To model an earthquake process, accelerogram data may be collected (ground acceleration set against time); to model the effects of wind, it may be important to capture changes in pressure against time; to model ocean storms and tidal patterns, the source data will likely take the form of water height over time. All of these examples have a common factor in that they are time-dependant. Although variation in the time-domain is not a requirement of stochastic processes in general, nor does it define a restriction for the methodologies presented herein, it provides a good basis for presenting many practical problems of a similar nature and in considering analysis of frequency dependant properties. In this regard, it is important to note that a time-based model definition of a stochastic process can take many forms. It could be as simple as an estimation of the mean and variance of a process, or as complex as a complete probability density function (PDF) with correlation defined for all time. Often, especially when dealing with processes that exhibit some underlying harmonic behaviour, an estimation of the process power spectral density is key [1]. A reliable spectral model providing frequency dependant information can be of significant importance in investigating the response of an engineering system to stochastic input. However, a basic spectral model may only describe a stationary process, i.e. one in which the spectral content does not change over time. This assumption of stationarity often produces a poor approximation of the true process, as many important processes of interest are non-stationary in nature. For example, the frequency content of an earthquake induced excitation can change dramatically over its duration, and wind systems may contain short infrequent bursts that do not conform to the otherwise stationarity of the rest of the process. Hence, in many cases, realization of time-dependant properties of stochastic processes are also considered central to defining reliable spectral models [2, 3]. For these reasons, power spectrum estimation, and in particular its non-stationary form (in time) is a primary focus in this work.

No two continuous stochastic processes are identical, as is the nature of stochasticity; one wind storm will never have exactly the same properties as another; no two earthquakes will produce exactly the same vibrations. Therefore, from a statistical standpoint, the more data there is available, the better we can understand a process or class of processes. Unfortunately it is not possible to gather an infinite amount of data, from an infinite number of properties. Available data is often quite limited, an issue that makes it difficult to build a reliable model which we can be confident encompasses

all relevant features of the process. Practical reasons for having limited data include the following:

- Equipment failure

  This is a possibility whenever using sensory equipment to detect and record stochastic processes. If a sensor becomes damaged, perhaps even as a result of the process itself, data may be lost.

- Sensor limitations

  Sensors for detecting vibrations, displacements, pressures etc will operate to some threshold limits. High fidelity sensors with a wide operational range can be expensive, and so in some cases the equipment used to record a process may not be able to capture extreme features.

- Maintenance

  Sensory equipment may not be designed to operate indefinitely and as a result, for long process recordings, may require intermittent maintenance.

- Bandwidth limitation

  If sensors are used remotely then data must be transmitted wirelessly. The bandwidth required to receive uncompressed data from multiple sensors may be too great to process simultaneously. In these situations sensor data may be received selectively in bursts, producing gaps in what would otherwise be a constant data stream.

- Usage restrictions

  It may be the case that to capture relevant data on the process of interest, expensive specialist equipment is required; equipment that may need to be rented or shared. Data that could have otherwise been captured, may have to be forfeited when a third party needs access to the equipment.

- Acquisition restrictions

  It could be the case that there are practical difficulties in being able to set up equipment so that the process can be captured. For example, when using an earth based telescope to detect objects in outer space, both cloud cover and the earth's rotation could prevent relevant data capture.

- Data corruption

  When dealing with digital data, corruption is always a possibility; perhaps as a result of accidental electrical damage or even malicious attack.

These are potential real issues in many engineering applications, and which in some situations lead to highly limited and irregularly sampled data sets. In this regard, when working with limited data, standard Fourier techniques for spectral estimation, e.g. [4], can demonstrate poor performance. Fortunately there exist many algorithms and procedures in the literature that provide alternatives to standard Fourier analysis for spectral estimation in the presence of missing data. Nevertheless, most of these alternatives come with certain drawbacks and often impose several assumptions on the statistics of the underlying stochastic process. For instance, autoregressive methods can be used to fit a model to the data, most often under the assumption that the source time-history is relatively long and that the missing data are grouped [5, 6]. Further, least-squares sinusoid fitting and zero-padded gaps [7, 8, 9] offer efficient solutions for reconstructing the Fourier spectrum in the presence of missing data but suffer, in general,

from falsely detected peaks, spectral leakage and significant loss of power as the number of missing data increases. Other alternative approaches for spectral estimation in the case of non-uniform sampling may impose restrictions on the nature of the missing data; e.g., infrequent loss [10, 11] or assume that the underlying process comprises of a highly limited number of significant harmonic components [12, 13]. Additional challenges arise when dealing with non-stationary data. In this regard, to estimate the power spectrum of a non-stationary process, the Gabor transform [14], wavelets [15, 16, 17, 18, 19], chirplets [20] and the Wigner-Ville distribution [21, 22] present means of analysing the non-stationary spectral content of a signal. Nevertheless, many of the approaches for addressing missing data in the stationary case cannot be applied, at least in a straightforward manner for non-stationary cases, or assume that the process is locally stationary [23].

The limited breadth of tools available for spectral analysis of stochastic processes in cases of missing data has led the author to develop new approaches to these problems, and is the primary motivating factor for this work. The need for such research is strongly supported by academic literature, and its applications are numerous in addressing aspects of risk and uncertainty across multiple areas of engineering. In this thesis, three novel approaches to problems associated with spectral estimation under missing data are presented. Two of these centre around reconstructing incomplete/gappy data sets, providing full, uniformly sampled time-histories for further analysis. The third is a fundamentally different approach that aims to quantify the uncertainty in spectral estimates due to missing data probabilistically.

1.2 Organization of thesis

The overall structure of this thesis is represented diagrammatically in Figure 1.1.

The main theoretical background upon which this work is based is first provided in Chapter 2; where many of the ideas identified in this introduction are elaborated on in detail. Certain features of this work, particularly in the numerical examples sections, are similar across chapters. These include re-used pre-defined spectral models, numerical transforms and simulations of missing data, which are all found towards the end of Chapter 2; and are referred back to throughout the rest of the thesis. Chapters 3, 4 and 5 constitute the majority of the author's original work, which has already been peer-reviewed and benefited from the insights of other researchers, through presentations and publications at international conferences and in academic journals. Chapters 3 and 4 provide methodologies for data reconstruction via artificial neural networks and compressive sensing respectively. Chapter 5 provides an alternative probabilistic methodology for spectral estimation under missing data. Conclusions from these three approaches are consolidated and presented in chapter 6, summarizing how the novel aspects of this doctoral investigation both contribute to international scholarly work and highlight areas of further research and collaboration.

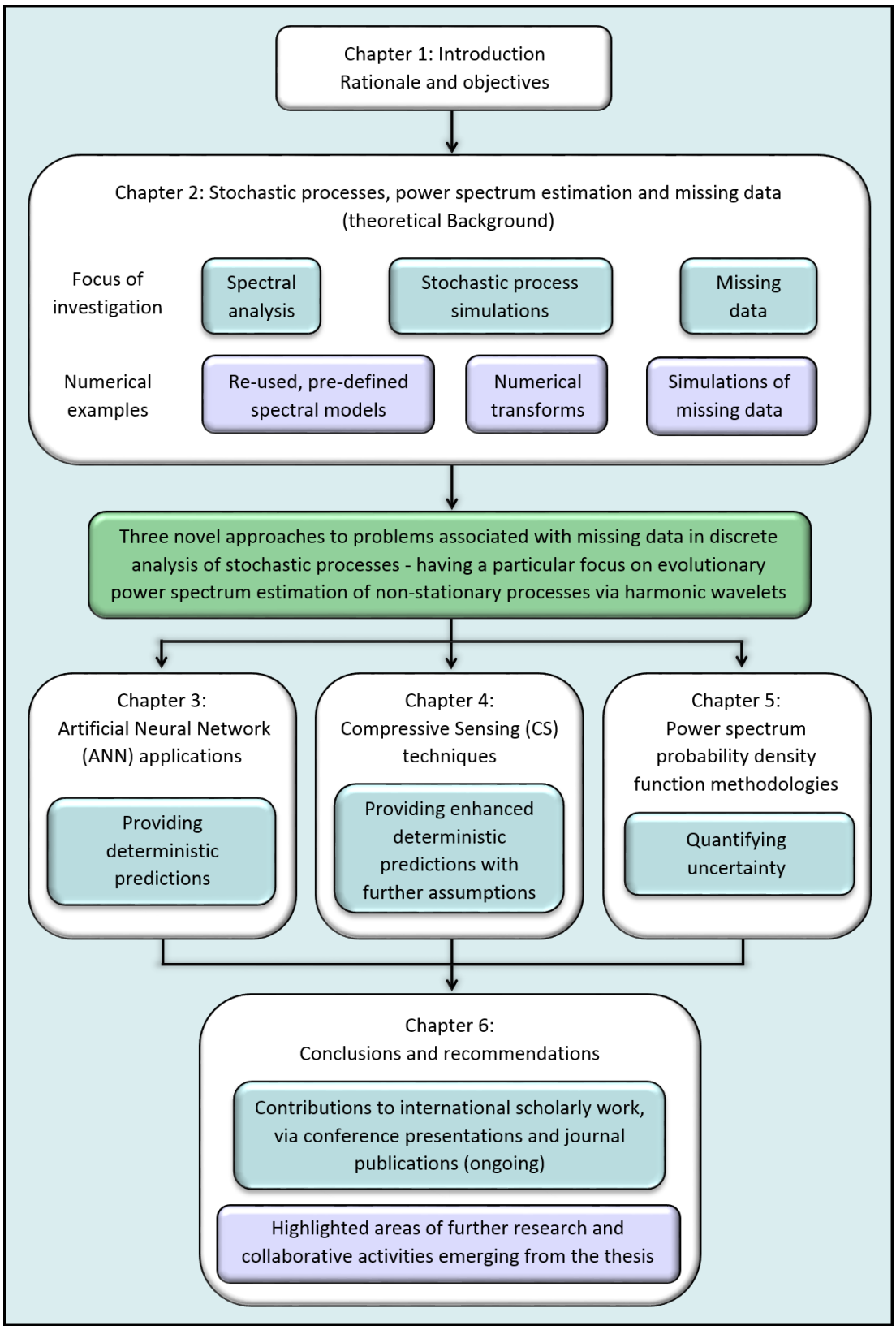


Figure 1.1: Thesis structure

Chapter 2

Stochastic processes, power spectrum estimation and missing data - an overview

2.1 Introduction

There is a wealth of information available on the subjects of stationary and non-stationary stochastic process spectral analysis, and how to mitigate problems caused by missing data, each method having its own advantages and disadvantages. As such, this section provides a brief review of relevant elements of stochastic process theory and spectrum estimation, as well as the motivations behind such analysis. This is followed by a definition of the problems associated with missing data in power spectrum estimation, along with the limitations of current spectral reconstruction methods. However, first, building blocks for stationary and non-stationary stochastic process representation applied in this doctoral investigation are discussed for completeness (this is primarily a review of probability theory, Fourier and Wavelet analysis).

2.2 Random variables & Stochastic processes

Probability concepts surrounding random variables and stochastic processes will be introduced here, providing a basis for signal transforms introduced in the next section.

2.2.1 Probability density functions

A random variable is a mathematical mapping from events on a probability space to the real line. It allows events to be represented in analytical form. If a real number, $x = X(\zeta)$ is assigned to each possible outcome, $X(\zeta)$, of the probability space, S, then X is defined on the real line and $X = a$, $X \leq a$ and $a \leq X \leq b$ may be classed as events. An example is given in Figure 2.1 where events E_1 and E_2 may be defined by limits on the real line. Random variables may be discrete or continuous. Note that most continuous events are already in numerical terms and the mapping from the event space to the real line is likely direct, I.e., weight, temperature, distance etc.

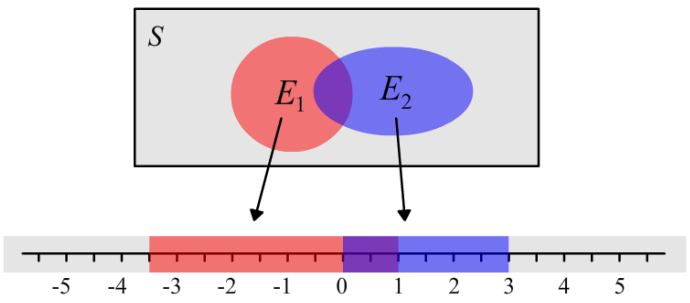


Figure 2.1: Example sample space with two events mapped to the real line

For a random variable X, we can define an event, $X \leq x$, to which some probability is assigned. When defined for all x this is known as the cumulative distribution function (CDF) denoted by,

$$F_X(x) \equiv P(X \leq x). \tag{2.1}$$

For a discrete random variable, the probability $p_X(x_i) \equiv P(X = x_i)$ for all x is known as the probability mass function (PMF). Hence the CDF for a discrete random variable may be written,

$$F_X(x) = \sum_{\text{all } x_i \leq x} p_{X(x_i)}. \tag{2.2}$$

For the continuous case, the PMF is of no use; this is because the probability of the event $P(X = x_i)$ in a continuous space is zero. Instead we define a probability density function (PDF), $f_X(x)$, representing the probability per unit over the entire number space. Hence, the probability of an event occurring in the interval $[a, b]$, is given by integrating the PDF between the interval limits,

$$P(a < X \leq b) = \int_a^b f_X(x)dx. \tag{2.3}$$

Similarly to the discrete case, the continuous CDF is written,

$$F_X(x) \equiv P(X \leq x) = \int_{-\infty}^{x} f_X(u)du. \tag{2.4}$$

In Eq.2.4 u is substituted for x to avoid its double definition as the static integral limit and variable sample value. Hence if $F_X(x)$ has a first derivative, its PDF in terms of its CDF is given by

$$f_X(x) = \frac{dF_X(x)}{dx}. \tag{2.5}$$

2.2.2 Measures of dispersion

A single random variable is completely defined in a probabilistic sense by its CDF, PDF or PMF. Other descriptors such as the central values (e.g., expectation, median, mode) and distribution moments (e.g., variance, skewness, kurtosis) are often important for a number of reasons. In practise the underlying distribution functions may not be known and it may not be possible to estimate them reliably without significant amounts of data. It is often much simpler to estimate the expectation, variance or other simple descriptor of a random variable, which may be the only information required for many practical applications. Further, some distribution functions may be completely defined by these descriptors. For example, the Gaussian distribution function (Eq.2.6) is defined by the expectation and variance of a random variable.

$$N(\mu, \sigma^2) \equiv f_X(x, \mu, \sigma) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}, \tag{2.6}$$

where μ is the process expectation, σ, the standard deviation (and σ^2, the variance); both of these are defined in the following.

For a discrete random variable, X, the expected value, denoted $E[X]$, based on its PMF is given by

$$E[X] = \sum_{\text{all}x_i} x_i p_X(x_i), \tag{2.7}$$

I.e., the sum of all outcomes, weighted by their probability of occurrence. For the continuous case the expected value is given by

$$E[X] = \int_{-\infty}^{\infty} f_x(x)dx. \tag{2.8}$$

Note also that the expectation of a function of X, $E[g(X)]$, is known as the mathematical expectation and is also given by its weighted probability average,

$$E[g(X)] = \int_{-\infty}^{\infty} g(x) f_x(x) dx. \tag{2.9}$$

Measures of dispersion give additional information concerning the arrangement of missing data. Of particular importance is the measure of how widely spread the data is about its expected value, commonly determined by its variance or standard deviation. The variance of a random variable is defined as the expectation of the square of its deviation from its expected value. Hence for a discrete random variable with PMF, $p_X(x_i)$, its variance is given by

$$E[(X - E[X])^2] = \sum_{\text{all}x_i} (x_i - E[X])^2 p_X(x_i), \tag{2.10}$$

and for a continuous process,

$$E[(X - E[X])^2] = \int_{-\infty}^{\infty} (x - E[X])^2 f_X(x) dx. \tag{2.11}$$

The square root of the variance is referred to as the standard deviation, usually denoted by σ. The standard deviation is widely used in statistical analysis as a measure of deviation from the mean as it is easy to calculate and has the same unit of measurement as the random variable. However it important to note that the standard deviation is not the same as the mean of the absolute deviations which may be a more robust measure of dispersion, especially when samples of the random variable are likely to deviate significantly from 1.

2.2.3 Multiple random variables

The concept of the random variable and its probability functions can be extended to include multiple random variables. Consider two continuous independent random variables, X and Y, each with its own probability density function. To compute the probability that both $P(X \leq x)$ and $P(Y \leq y)$ simultaneously, we consider the joint probability density function $f_{X,Y}(x, y)$. Hence,

$$P(X \leq x, Y \leq y) \equiv F_{X,Y}(x, y) = \int_{-\infty}^{x} \int_{-\infty}^{y} f_{X,Y}(u, v) dv du, \tag{2.12}$$

where u and v are substituted for x and y to avoid their double definition as the static integral limits and variable sample values and $F_{X,Y}(x, y)$ is the joint CDF. The bivariate CDF can be extended to the multivariate CDF for n random variables,

$$F_{X_1,X_2,...,X_n}(x_1, x_2, ..., x_n) = \int_{-\infty}^{x_1} \int_{-\infty}^{x_2} ... \int_{-\infty}^{x_n} f_{X_1,X_2,...,X_n}(u_1, u_2, ..., u_n) du_1 du_2 ... du_n. \tag{2.13}$$

Which may be written more conveniently in vector form,

$$F_{\boldsymbol{X}}(\boldsymbol{x}) \equiv P(\boldsymbol{X} \leq \boldsymbol{x}) = \int \int ... \int_{-\infty}^{\boldsymbol{x}} f_{\boldsymbol{X}}(\boldsymbol{u}) d\boldsymbol{u}. \tag{2.14}$$

Hence, the multivariate PDF for n variables may be defined by

$$f_{\boldsymbol{X}}(\boldsymbol{x}) = \frac{\partial^n F_{\boldsymbol{X}}(\boldsymbol{x})}{\partial x}. \tag{2.15}$$

In many practical problems of interest, random variables may not be independent, I.e., there is some relationship between them. We may determine the presence of a linear statistical relationship between two random variables X and Y by way of their covariance,

$$\mathrm{Cov}(X,Y) = E(XY) - E(X)E(Y) \tag{2.16}$$

The normalized covariance or "Pearson's correlation" is often used in practise,

$$\rho_{X,Y} = \frac{\mathrm{Cov}(X,Y)}{\sigma_X \sigma_Y} \tag{2.17}$$

and is a dimensionless coefficient with values in the range $-1 \leq \rho_{X,Y} \leq 1$.

2.2.4 Stochastic processes

We define a random variable, X as a rule for assigning a number, $x = X(\zeta)$, to every possible event, ζ from a complete set of events. If we consider instead that a function, $x(t)$, is assigned to each of these events, we are left with a family or 'ensemble' of possible realizations. This ensemble of functions is known as a stochastic process. In the same way that a random variable is, in general (depending on the mapping), different for each possible event, each function, $x(t)$, is also different. These functions make up a family or 'ensemble' of possible realizations, this ensemble of functions is known as a stochastic process. Hence, a stochastic process, $X(t)$ is a rule for assigning a function, $x(t) = X(t,\zeta)$, to every possible event in time from a complete set of events. Figure 2.2 shows three example sample realizations of a stochastic process. For each realization, $x(t) = X(t,\zeta)$, t is variable and ζ is fixed. Further, the state of the stochastic process at a given time, t is a random variable.

In the same way that two random variables are linked by their joint probability density function, the random variables that make up a stochastic process are linked by their n^{th} order probability density function.

Stationary and non-stationary processes

In many practical cases, the generating mechanism behind a stochastic process varies with time. However, depending on the type and magnitude of variation, as well as acceptable assumptions for the engineering application in question, processes may be considered stationary or non-stationary.

A stochastic process is called "Strict-Sense Stationary" (SSS) if its statistical properties are completely invariant to a shift in the origin. This means that for any process that is SSS, its joint PDF up to n time steps (its n^{th} order density) must be such that:

$$f_X(x_1, x_2 ... x_n; t_1, t_2 ... t_n) = f_X(x_1, x_2 ... x_n; t_1 + c, t_2 + c ... t_n + c) \tag{2.18}$$

Hence the marginal PDF of the process at any point is independent of time, i.e.,

$$f_X(x,t) = f_X(x) \tag{2.19}$$

This means that the joint PDF of an SSS process is independent of c. Thus, if we set $c = -t_2$, the joint density at two points in time for an SSS process may be written,

$$f_X(x_1, x_2; t_1, t_2) = f_X(x_1, x_2; t_1 - t_2) \tag{2.20}$$

So the joint density is dependent only upon the time difference $\tau = t_1 - t_2$. As a result, the autocorrelation function $R(t_1, t_2)$ is also dependent only upon τ,

$$R(t_1, t_2) = E[X(t_2 + \tau)X(t_2)] = E[X(t + \tau)X(t)] = R(\tau) \tag{2.21}$$

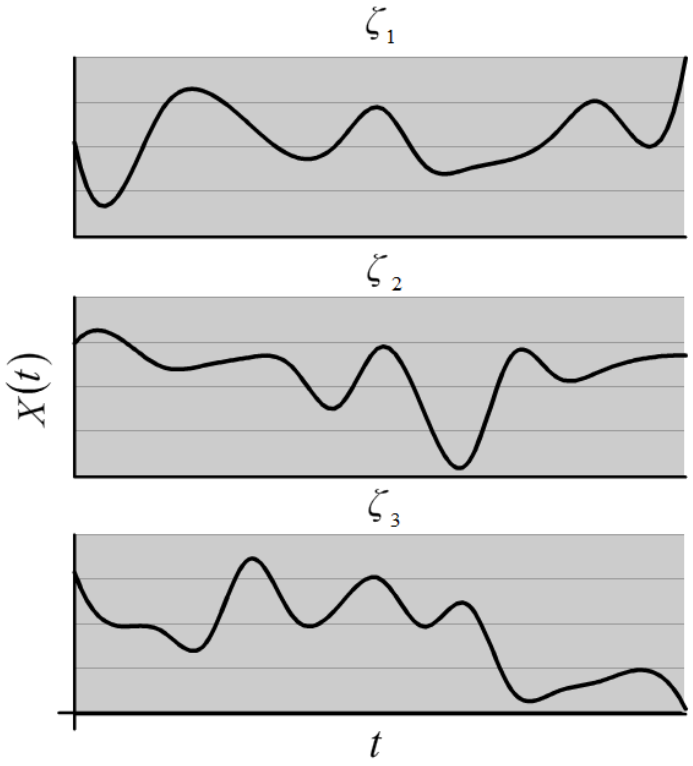


Figure 2.2: Three sample functions of a stochastic process

The requirement that the entire PDF at any point is independent of time is a very strict property and can be difficult to impose on a wide variety of real processes. A weaker stationarity condition is often assumed instead. Rather than all statistics of the PDF being dependent on time, only the first two statistical moments are fixed. I.e.,

$$E[X(t)] = E[\mathbf{X}], \tag{2.22}$$

and,

$$E[X(t)^2] = E[\mathbf{X}^2]. \tag{2.23}$$

Note here that $E[\mathbf{X}^n]$ is used in a point-wise sense, not as a matrix multiplication. Again, the autocorrelation function $R(t_1, t_2)$ is dependent only on τ, as in Eq.2.21. This class of processes are referred to as "Wide-Sense Stationary" (WSS). Processes referred to simply as "stationary" in the rest of the project may be assumed to be WSS.

Ergodicity

To find some statistical property of a stochastic process at a particular time, t, we can calculate the expectation of that property across the ensemble. For example, to simply find the expectation of the process at $t = t_1$, we would evaluate $E[X(t_1)]$.

In a real scenario with real data, clearly several recorded realizations of the same process are required to evaluate specific time dependant properties. The number of realizations needed to reliably calculate the expectation at a specific time could be significant, depending on the accuracy required and higher order statistics of the process. For a reliable estimate thousands or more process realizations may be required and it is seldom the case that this number are available. In fact, often only a single realization of a recorded stochastic process will be available for analysis. Hence an alternative practical method of estimating process statistics is required.

As an example, consider calculating the mean, not from points at $t = t_1$ across an ensemble, but for a single process at all recorded points $t_0 \leq t \leq t_p$. First the process must be stationary so that the expected result is independent of the specific times t_0 and tp. Secondly it must be assumed that all process realizations in time (if they were available) would have the same mean value when calculated over all t. Under these conditions the process is ergodic in the mean sense.

If a process is ergodic in the n^{th} central statistical moment then,

$$E[(X(t=t_1) - E[X(t=t_1)])^n] = \frac{1}{p}\int_{x=x(0)}^{x=x(p)} (x - E[X])^n dx \quad \text{holds for } p \Rightarrow \infty \quad (2.24)$$

Hence, a process that is ergodic in all statistical moments must be SSS.

2.3 Review of Fourier analysis

With concepts of stochastic processes introduced, attention is now turned to methods of transforming and representing such processes and the advantages of doing so. Transforms will be discussed first by considering deterministic signals, with stochastic processes revisited in section 2.5.

2.3.1 Linear transforms

Definitions of vector spaces, linear independence, orthogonality and orthogonal transforms are given here before introducing Fourier theory for completeness.

Vector spaces

A vector space is a set of vectors, $V = \{v_i\}$ and scalars, $A = \{a_i\}$ that have the following properties:

1. Two vectors when added together form a third vector.

2. Any vector can be multiplied by a scalar to produce another vector

With these two operations, algebraic addition and multiplication properties must hold for all $v_i \in V$ and $a_i \in A$ i.e.,

1. $v_1 + v_2 = v_2 + v_1$

2. $(v_1 + v_2) + v_3 = v_1 + (v_2 + v_3)$

3. $a_1(v_1 + v_2) = a_1v_1 + a_1v_2$

4. $(a_1 + a_2)v_1 = a_1v_1 + a_2v_1$

5. $(a_1a_2)v_1 = a_1(a_2v_1)$

Any sub-set of vectors from a full set that define a vector space may be referred to as a vector "subspace". Note therefore that a vector set may only be a subspace in the context of another larger vector set.

Linear independence

A set of vectors $\{v_1, v_2, \cdots, v_n\}$ in the vector space V over the field A is defined as "linearly independent" if

$$\sum_{i=1}^{n} a_i v_i = 0 \tag{2.25}$$

holds only if all scalars $\{a_i\}$ are equal to zero. Therefore within a set of linearly independent vectors, it is not possible to write any one vector in terms of linear combinations of others in the set.

Orthogonality

Two vectors are defined as "orthogonal" if their inner product is equal to zero. Hence for two discrete mutually orthogonal vectors v_1 and v_2 both of length N,

$$\langle v_1 | v_2 \rangle = \sum_{n=1}^{N} \overline{v_1(n)} v_2(n) = 0 \tag{2.26}$$

and for continuous orthogonal vectors,

$$\langle v_1 | v_2 \rangle = \int \overline{v_1(x)} v_2(x) dx = 0 \tag{2.27}$$

with appropriate integration boundaries (the bar indicates the complex conjugate for complex vectors). Similarly, two functions may be defined as orthogonal if their inner product is equal to zero. Orthogonality is a general feature of complex exponential signals of different frequency.

Orthogonal bases and transforms

A basis is a set of linearly independent vectors that span a given vector space. Any vector within the space may be represented by a unique combination of the basis vectors. Imposing orthogonality on a basis means that all pairs of basis vectors must be mutually orthogonal. To represent a given vector by components of an orthogonal basis which it is a part of, it must undergo a linear transformation. For a discrete signal, a basis transformation matrix is used to transform between two finite-dimensional vectors of equal length. Hence if X is a vector, Y is the same vector mapped on to a new vector space V and A is basis transformation matrix spanning V then,

$$Y = AX \tag{2.28}$$

These concepts form the foundations of Fourier and wavelet transforms which feature heavily in the rest of this text.

2.3.2 Fourier transform

Any arbitrary periodic function of time can be expressed as an infinite sum of sine and cosine functions. The resulting expression is known as a Fourier series and is demonstrated, explained or at least mentioned in any comprehensive book concerning signal processing and spectral analysis; [17, 22, 14] provide introductions to the subject.

Representing a signal as a spectrum of frequency specific components is a key procedure in many signal processing applications. Seismology, medical imaging, data compression, financial analysis and speech recognition constitute some indicative examples

where the concept of the frequency spectrum plays an important role. Spectral representation can also be a very reliable and intuitive way of characterizing excitations produced by environmental processes due to their tendency to exhibit underlying harmonic properties. Unfortunately in the case of environmental processes, although certain signal properties can be identified through a Fourier analysis, their interpretation can often be inappropriate and misleading. Aside from the well-known difficulties associated with end-effects (section 2.3.5) and aliasing (section 2.3.4) which may be mitigated with proper data collection and appropriate pre-scaling, real environmental processes of interest are commonly non stationary, as the dominating frequencies change over time. For this reason alternative spectral analysis tools are later considered that can take account of both frequency and time localization simultaneously. However, as the Fourier transform is used as a foundation for the non-stationary analysis techniques in this doctoral study, as well as being applied directly for problems of a stationary nature in many of the upcoming numerical examples, it is defined here for completeness.

For a given function $x(t)$, its Fourier transform is defined as

$$X(\omega) = \int_{-\infty}^{\infty} x(t)\, e^{-2\pi i \omega} dt, \tag{2.29}$$

Because the transform is orthogonal, its inverse can be solved given $X(\omega)$ to perfectly reproduce the original signal $x(t)$. Eq.2.29 represents a continuous transform for an analytic function. For analysis of discrete digital recordings of real process excitations, a discretized estimate of the true frequency decomposition is made; this is known as the discrete Fourier transform or DFT. If x_n is a discrete series representing the continuous function $x(t)$ sampled at N uniformed intervals, then the DFT may be written as

$$X_k = \sum_{n=0}^{N-1} x_n e^{-2\pi i k n/N} \tag{2.30}$$

Although Eq.2.30 is an estimate of the Fourier transform of the continuous time series $x(t)$, it is important to note that the discrete transform is still orthogonal, and its inverse can be computed to re-produce the original sample set x_n.

2.3.3 FFT

It may be necessary to calculate the spectra of extremely long data records or large numbers of records as in a Monte Carlo analysis. In these cases faster computational methods than the simple DFT are required, and hence the DFT is rarely calculated in practise using the method described by Eq.2.30. The Fast Fourier Transform (FFT) (e.g. [16]) is an algorithm that is able to reproduce the DFT of a signal exactly, but with significant reduction in the number of computational operations. Historically the first FFT algorithms operated on signals of length 2^n; although this is no longer a requirement, signals of length 2^n are primarily used in this project. It can be clearly seen that performing the DFT using Eq.2.30 for a record of length N will require a number of operations to the order $N \times N$. The FFT in fact reduces the number of operations to the order of $N \log_2 N$ by partitioning a signal into shorter sequences, originally through dyadic down-sampling (which is why early FFT algorithms worked most effectively on sequences of length 2^n). However other splitting algorithms can be applied for efficiency and accuracy depending on the initial data. The FFT is utilized throughout this work.

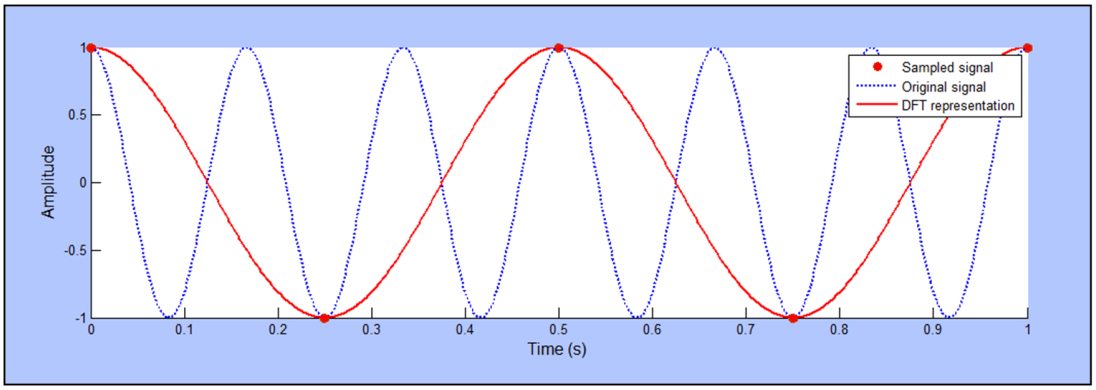


Figure 2.3: Example of aliasing

2.3.4 Aliasing

When dealing with real data it is important to be aware that any frequencies present which are higher than half the sampling rate (see Nyquist frequency [24]) will not be truly represented in the set. The DFT will still identify coefficients to fit the signal as it is, but the frequencies they represent will not be a true account of the process. For example, if a signal of frequency 6 Hz is sampled five times over a $1s$ period, the sequence would not contain enough data to describe the signal. The DFT would detect different frequency content entirely as shown in Figure 2.3.

2.3.5 End-effects

When using real or simulated discrete data for spectral analysis, all possible signals that might be chosen for analysis will be time-limited (i.e. of finite length). Without further knowledge of the underlying process from which a discrete sample signal is drawn, it is impossible to know with certainty the nature of the signal beyond the measured interval. In the discrete case of the Fourier transform, this property of signal termination is accounted for by the fact that the transform assumes that the measured signal repeats itself indefinitely. This is seldom true, because even if the signal itself is periodic within the measured interval, the interval must be a multiple of the period for the DFT to give Fourier coefficients directly representing only the original components of the signal. The consequence of this fact is sometimes known as "spectral leakage", and is demonstrated for a single harmonic signal $f(t) = \sin(5t)$ sampled from $t = 0$ to $t = 2.5\pi$ in Figure 2.4 (assumed signal) and Figure 2.5 (spectral leakage on the signal). In many cases, the end-effect problem may be mitigated to some extent by introducing a smoothing function that decays to zero at the sample ends. Specific methods used for mitigating end-effects in non-stationary signals used in this work are described in section 2.5.1.

2.4 Time-frequency analysis of non-stationary signals

A significant disadvantage of Fourier analysis comes from the fact that it provides the spectral content of a signal calculated as an average over the entire domain (in this case time) from $t = -\infty$ to $t = \infty$. Similarly, the output of the DFT produces a set of Fourier coefficients that represent the spectral composition of the entire data set. Therefore if a time-localized fluctuation in frequency dependent power occurs, from a Fourier analysis there is no way of knowing when it occurred or how long it lasted. Further, because the frequency of that particular fluctuation is represented by Fourier

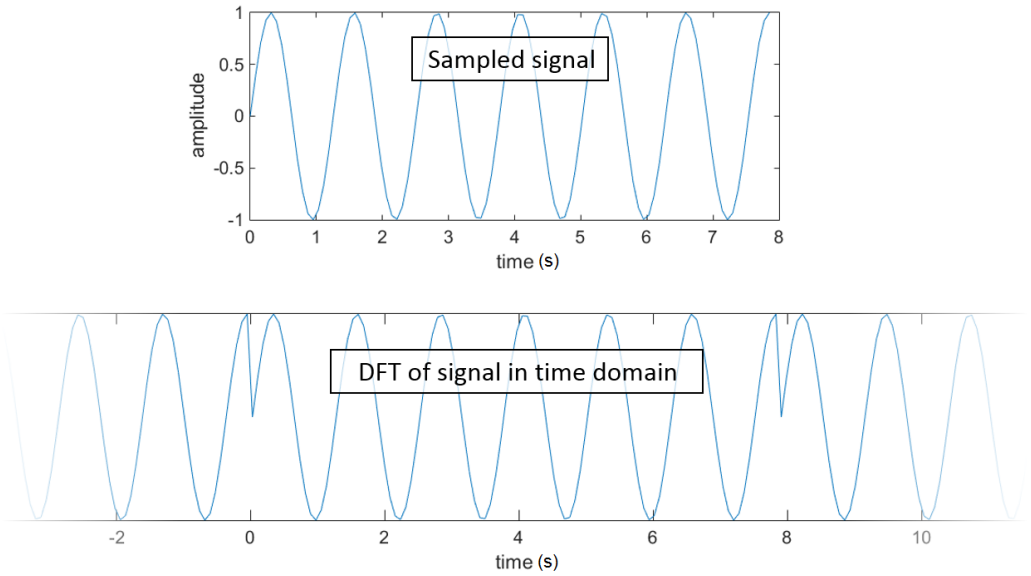


Figure 2.4: A measured sinusoid (top) and the DFT assumption (bottom)

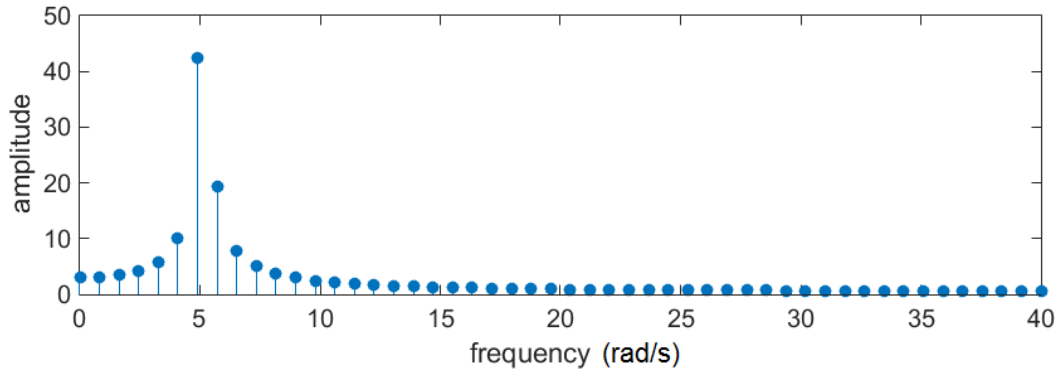


Figure 2.5: Spectral leakage around 5 rad/s in the frequency domain taken from the DFT of the signal in Figure 2.4

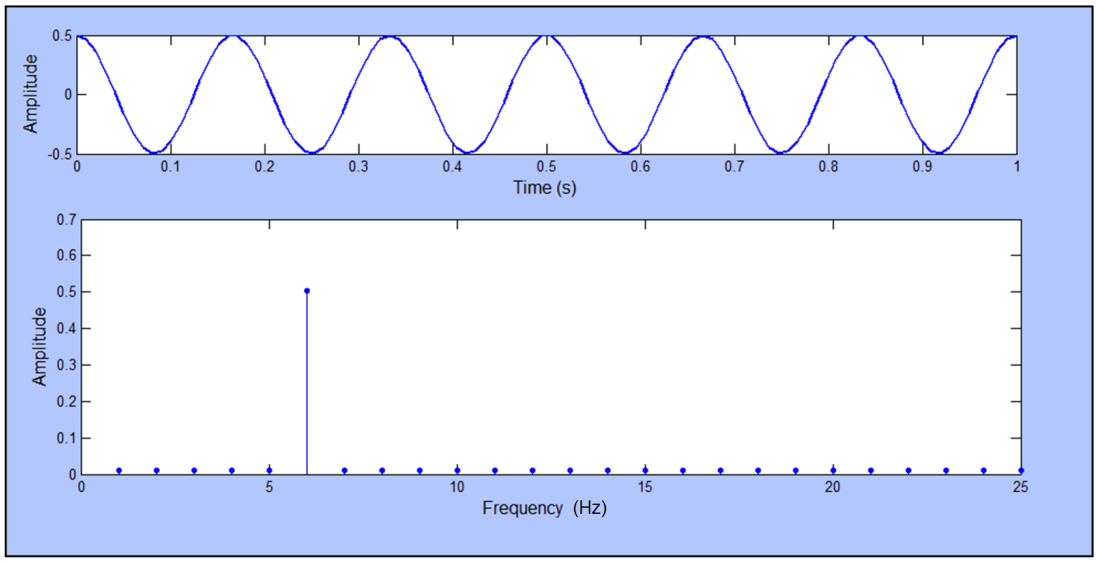


Figure 2.6: Cosine signal and its absolute Fourier coefficients

coefficients that describe the signal as a whole, the estimated power of the fluctuation will be significantly lower. When looking at Fourier coefficients it is impossible to know whether they describe the actual amplitude of harmonics lasting for the entire measured signal, or fractions of harmonic amplitudes lasting for only short intervals within the measured signal. Figure 2.6 shows a cosine of frequency 6 Hz and amplitude 0.5 above a plot of absolute value of its DFT. Notice that the DFT has perfectly identified the amplitude of the wave at 6 Hz. Figure 2.7 also shows a time-domain cosine of frequency 6 Hz, but with twice the amplitude and appearing only for half of the length of the sample. This time the DFT (Figure 2.7, bottom) has identified many unwanted frequencies; these have occurred as a result of the discontinuity within the sample. Also at 6 Hz the DFT has identified an amplitude of only half that of the actual signal (as it is 1 for half the data and 0 for the other half). If the change in frequency happened over a much shorter interval in a real signal containing noise, then it would likely be completely lost in a Fourier analysis.

As many processes of interest, especially environmental processes, harbour this time-localized or "non-stationary" behaviour, a method of estimating not only the spectral content of a record but also how this content varies in time is required.

2.4.1 Short-time Fourier transform

The short-time Fourier transform (STFT) is possibly the most intuitive way to think about conducting a spectral analysis that yields information indexed in both time and frequency. A full description of the procedure can be found in [14]. The STFT has a short time window centred at time t in the signal, hence the name "short time". The data in this window is separated and spectral coefficients are obtained based on that isolated sample. The data window is then moved along the signal in time and a new sample is isolated. The process is repeated until the window has reached the end of the entire measured data set. Hence the spectral composition of the signal may be reviewed at every instant around which the window was centred. The Discrete-time STFT may be expressed as:

$$\mathrm{STFT}\{x(n)\}(m,\omega) = \sum_{n=-\infty}^{\infty} x(n)w[n-m]e^{-i\omega n}, \tag{2.31}$$

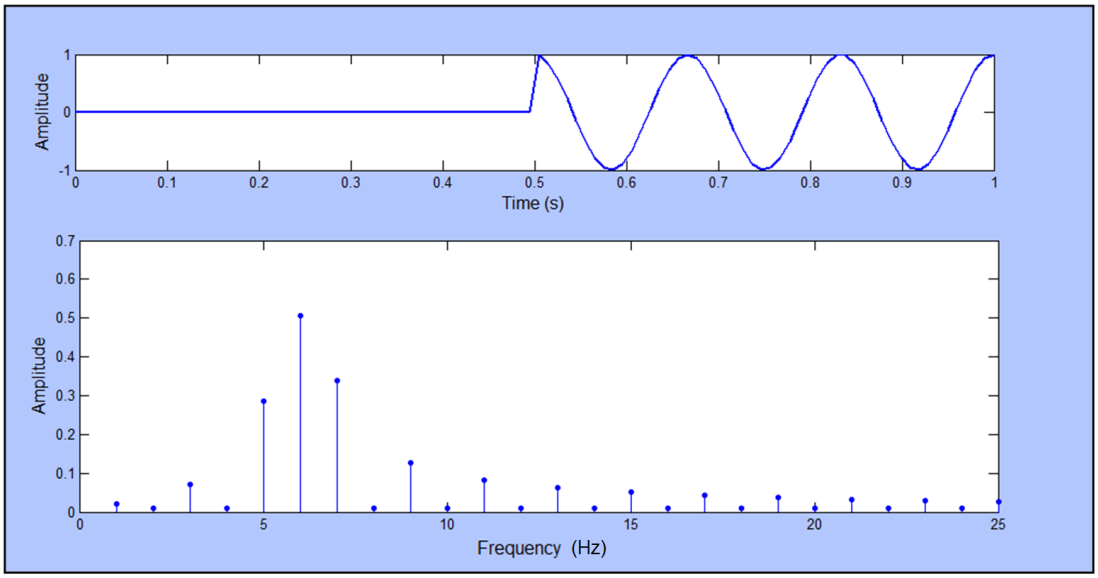


Figure 2.7: Cosine signal (lasting for half the sample) and its absolute Fourier coefficients

where w is a window function centred around m. Unfortunately high resolution cannot be obtained simultaneously in both time and frequency. This problem can be thought in a similar sense to aliasing; frequencies with wavelengths longer than the short time window cannot be identified by the DFT and will be misrepresented by higher frequencies over a mean power. Therefore this method is only effective in situations where the measured data is long in time and only high frequencies are of interest (shorter time window), or where the frequency content only changes very gradually in time (longer time window). The STFT also more prominently suffers from "end-effects" e.g. [25] which occur in the DFT as a result of its cyclic nature. As the STFT involves multiple DFT windows, these edge effects are amplified. This problem is often mitigated however by the use of smoothing window functions, though the transform is no longer orthogonal once a smooth window is employed.

2.4.2 Introduction to wavelet transforms

Wavelet analysis can be thought of as another way of representing a signal through a series of pre-defined basis functions [16]. In Fourier analysis these functions are sine waves, In a Taylor series expansion these functions are polynomials, both of which have infinite energy i.e., they span their entire domain from $-\infty$ to ∞. A wavelet basis function is fundamentally different in that it decays to zero when sufficiently far from its centre. This means that the wavelet exists only for a limited period of time, and therefore wavelet components of a signal are known to lie within certain time intervals. Unlike the standard STFT, a wavelet analysis allows time windows of varying length to be used to analyse different frequency bands while maintaining an orthogonal basis. The family of wavelets chosen will depend on the specifics of the problem. Particular wavelets might be useful for speech recognition or musical analysis for example, and others for compression/decompression of digital images or video, etc.

Continuous wavelet transform

A continuous wavelet transform is used to represent a continuous-time function as an infinite series of wavelets (in the wavelet basis). In general, the continuous wavelet

transform of a signal $x(t)$ may be expressed as

$$X(a,b) = \int_{-\infty}^{\infty} x(t)\overline{\psi_{(a,b)}(t)}dt \tag{2.32}$$

Where ψ is a wavelet function obeying a certain set of rules. The continuous wavelet transform returns a signal that is one dimension higher than the input. Hence the continuous wavelet transform or its discretised counterpart (discrete time continuous wavelet transform) will generally produce a non-orghogonal, linearly dependent set with high redundancy. In many cases, wavelet transforms are based on a "mother wavelet" function which is scaled and translated over the time domain to form a family of wavelets. Hence, in this form the continuous wavelet transform of a function $x(t)$ at scale a and translated by b may be expressed as

$$X(a,b) = \frac{1}{\sqrt{a}} \int_{-\infty}^{\infty} x(t)\overline{\psi\left(\frac{t-b}{a}\right)}dt \tag{2.33}$$

Discrete wavelet transform

The discrete wavelet transform uses a finite set of wavelets (again usually defined by their scales and translates). This can be thought of in a similar way to the DFT. The if the wavelets are orthogonal and the length of the signal and number of wavelets in the set are equal, then the discrete wavelet transform provides an orthogonal transform. Despite the dimensionality still being increased as in the continuous case, the length of each dimension is such that the total number of data remains unchanged. When used appropriately, discrete wavelet analysis can give meaningful results with efficient time-frequency trade-offs.

Orthogonal families of discrete wavelets may be generated by finding suitable coefficients $c(n)$ and $d(n)$ that satisfy the orthogonal dilation equation pair (Eq.2.34, e.g. [17])

$$\begin{array}{l} \varphi(t) = \sum_n c(n)\varphi(St-n) \\ w(t) = \sum_n d(n)\varphi(St-n) \end{array} \tag{2.34}$$

In Eq.2.34, $\varphi(t)$ is known as a scaling function, S is a scaling factor (usually chosen as 2) and $w(t)$ is a wavelet constructed from the scaling function. When $S = 2$, the resulting family of wavelets are referred to as "dyadic wavelets".

Let the scaling function $\varphi(t)$ and its translates, $\varphi(t-n)$, occupy a vector space, V_0. Also assume that scaling function and its translates on V_0 may be represented by some combination of dilated (compressed in time) scaling functions $\varphi(2t-n)$ and that these occupy the vector space V_1. If any $\varphi(t-n)$ may be represented by some combination of scaled $\varphi(2t-n)$, then it is clear that the V_0 is a subspace of V_1. Hence if V_1 encompasses V_0, then it may be represented by a combination of V_0 and an orthogonal complement set, W_0 (known as the mother wavelet). I.e., V_1 may be split into two subspaces: V_0 and W_0. Hence the space V_N may be represented by the combined spaces V_{N-1} and W_{N-1}. Therefore any function in V_N may be represented by a combination of the scaling function in $\varphi(t)$ in V_0 and the family of wavelets $w_{0,1\ldots N-1}(t)$ in $W_{0,1\ldots N-1}$.

We can now define the discrete wavelet transform as a combination of two linear transforms: one for the scaling function, and one for the wavelet set,

$$X_\varphi(j_0,k) = \sum_t x(t)\varphi_{j_0,k}(t) \tag{2.35}$$

$$X_w(j,k) = \sum_t x(t)w_{j,k}(t) \quad for \quad j \geq j_0 \tag{2.36}$$

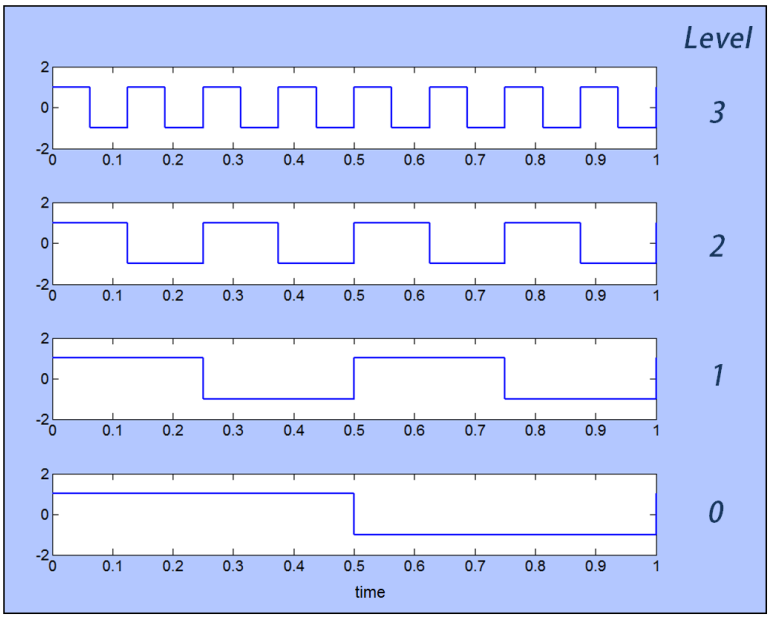


Figure 2.8: Levels 0 to 3 of the Haar wavelet

Where j represents the wavelet scale (or dilation) and k represents the translation step in the time domain.

Two discrete wavelet families are discussed to further introduce concepts surrounding discrete wavelet analysis. Then harmonic wavelets, which have some unusual properties for discrete wavelets are discussed and used extensively throughout the rest of the thesis.

Haar wavelets

The Haar transform is based on the simplest choice of discrete wavelet first proposed by Haar in 1910 [26]. A Haar decomposition involves a single step function of height 1, (known as the scaling function) and multiple rectangle functions of different scales (Figure 2.8). Note that for the transform to be orthogonal, each scale of wavelet must be half the width of the last. The scale or wavelet "level" as shown in Figure 2.8 is determined by how many wavelets fit into the unit interval $x = 0$ to 1. At level 0 there are 2^0 wavelets, at level 1 there are 2^1 wavelets etc. Below level zero, only a fraction of the wavelet would be included which could also be represented by combinations of higher level wavelets; so these are discounted.

Any signal can be broken down into its wavelet components; furthermore, using an orthogonal wavelet basis, the original measured signal can be reproduced exactly from its wavelet coefficients. This may seem difficult in the case of the Haar wavelet because of its discontinuous nature, but if a real signal record is considered of discrete measurements, when resolving down to the Nyquist frequency, a step function is indiscernible from the true function.

If we consider a signal $f(x)$ that has been measured over the interval $0 \leq x \leq 1$, its Haar wavelet decomposition can be written as

$$f(x) = a_0 + a_1 w(x) + a_2 w(2x) + a_3 w(2x-1) + a_4 w(4x) + a_5 w(4x-1) + a_6 w(4x-2) + \ldots \tag{2.37}$$

Where a_0 represents the scaling function (in this case the mean of the signal), a_1 is the amplitude of the only wavelet in level 1 (spanning the entire length of the signal), a_2 and a_3 are the amplitudes of the level 2 wavelets, and so on. In a more general form

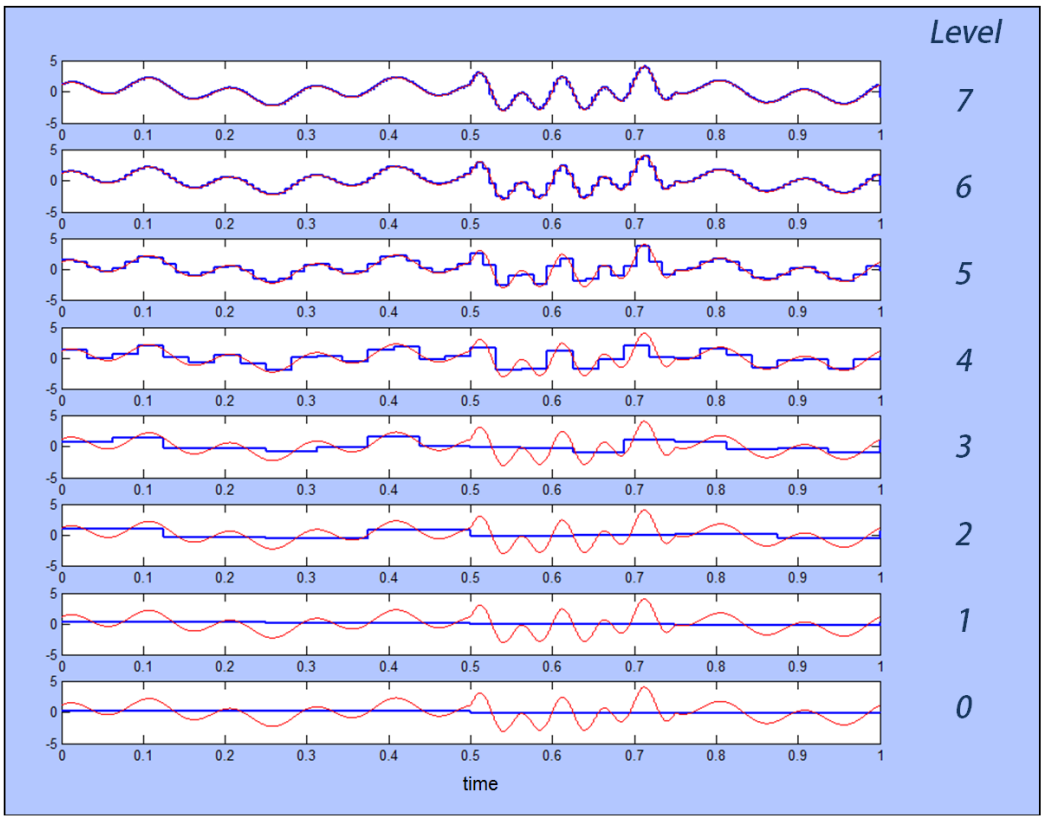


Figure 2.9: Levels 0 to 7 of a Haar wavelet decomposition of Eq.2.39

Eq.2.37 is written as

$$f(x) = a_0 + \sum_{j=0}^{\infty} \sum_{k=0}^{2^j-1} a_{2^j+k} w\left(2^j x - k\right), \quad for \quad 0 \leq x \leq 1 \tag{2.38}$$

Each coefficient has a position index along the signal, which, for a time signal, would represent wavelet amplitude over a particular time interval. The coefficients are also indexed by the level (width) of wavelet, and for a time signal this could be thought of as being similar to frequency. Figure 2.9 shows a non-stationary signal being decomposed into Haar wavelets, where the non-stationary aspect comes from a short high frequency burst added to an otherwise stationary signal between $0.5s$ and $0.73s$:

$$f(x) = \begin{cases} 2\sin(16\pi t + 1) + \sin(4\pi t) + 2\sin(36\pi t) & 0.5s < t \leq 0.73s \\ 2\sin(16\pi t + 1) + \sin(4\pi t) & otherwise \end{cases}, \tag{2.39}$$

Figure 2.10 shows the wavelet magnitudes in time at different levels for Eq.2.39

Unfortunately as the levels do not represent actual harmonic frequencies within the sample, there are limited quantitative observations that can be made from Figure 2.10. Despite this, it can be seen that between $0.5s$ and $0.73s$, signal frequency increases in levels 6 and 7 (2^6 wavelets per second and 2^7 wavelets per second respectively), so the localized high-frequency bust can clearly be identified. However, in levels 2 to 4 there are significant changes in wavelet magnitude over time, despite the lower frequency content of the signal remaining constant. These low frequency discrepancies are caused primarily in this case by poor time resolution in lower level bands when using a Haar wavelet decomposition and the discontinuous nature of the Haar wavelet. However, there are alternative wavelet families that could be considered that offer a range of time/frequency resolution trade-offs such as Daubechies wavelets (introduced

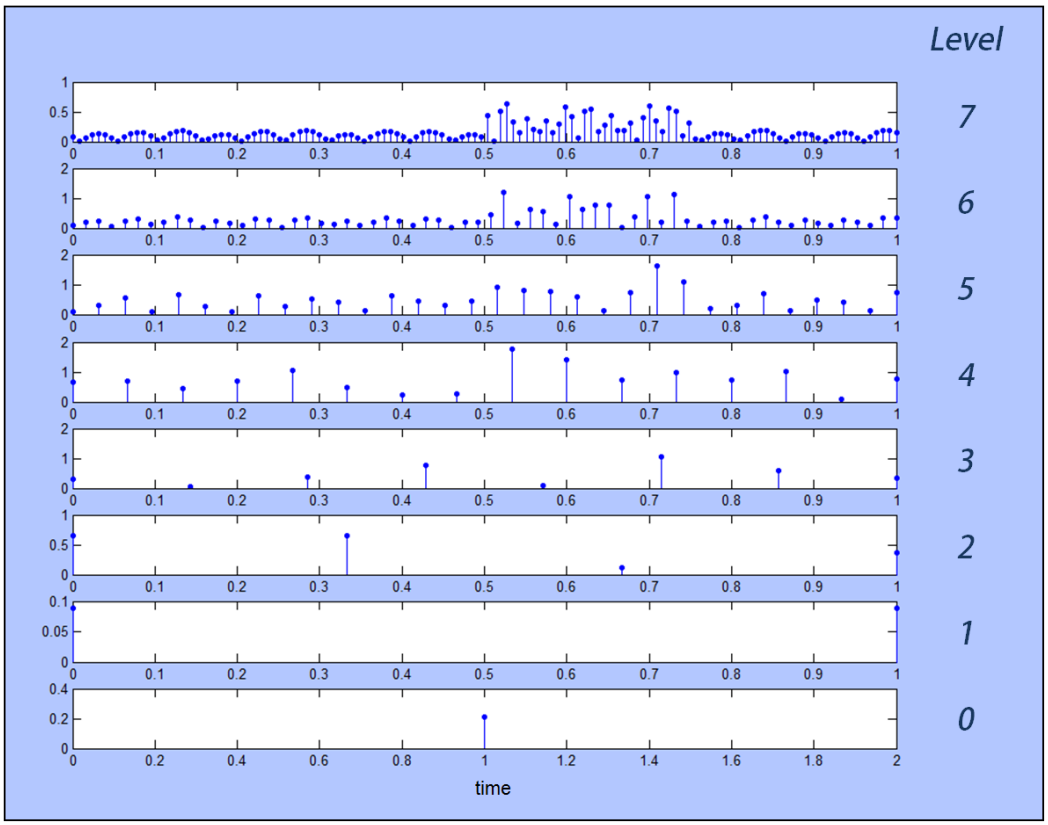


Figure 2.10: Levels 0 to 7 of Haar wavelet coefficients representing a signal

below), the Morlet wavelet (e.g.,[15], frequently used in environmental signal processing applications [27]) and Harmonic wavelets (used throughout this project) to name a few.

Daubechies wavelets

A procedure for solving Eq.2.34 (determining the coefficients $c(n)$ and $d(n)$) to generate orthogonal wavelets was developed in [28], the resulting family of functions known as the Daubechies wavelets. Daubechies wavelet families may be defined for any even number of $c(n)$ coefficients in Eq.2.34. As the number of coefficients increases, the wavelet function becomes smoother and more box-like in the frequency domain (Figure 2.11), Further, for discrete wavelet transforms produced through dyadic decomposition schemes such as Daubechies and Harr, better representations may be achieved through utilizing the wavelet packet transform e.g., [15]. Dyadic decomposition schemes discard data at each wavelet level whilst maintaining an orthogonal basis with no redundancy. The wavelet packet transform decomposes the otherwise discarded data at each level, providing a redundant set of wavelet coefficients. Some selection criteria is then used (such as minimum entropy) to choose the best wavelets from the redundant set that form an orthogonal, non-redundant basis e.g., [29].

2.4.3 Harmonic wavelets

Discrete wavelets generated via the dilation equation pair (Eq.2.34) are compact in the time domain in the sense that they have a definite beginning and ending. This means that they cannot also be compact in the frequency domain due to the Heisenberg-Gabor limit (or "uncertainty principle"). The result of which is that a function cannot be both time limited and bounded in the Fourier domain simultaneously. In contrast, Harmonic

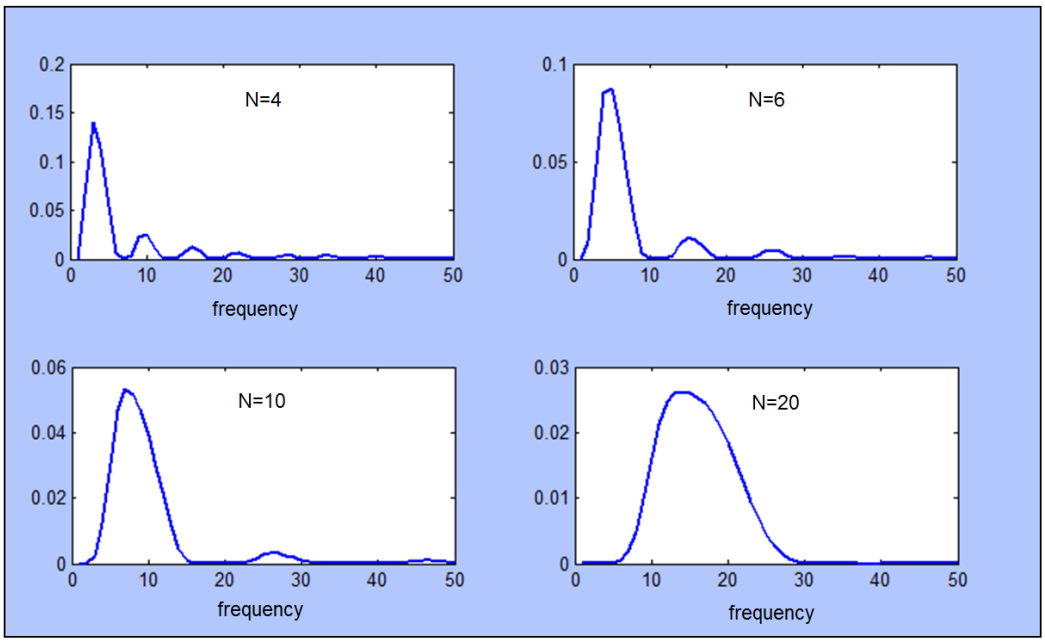


Figure 2.11: Daubechies wavelets in the frequency domain constructed from 4, 6, 10 and 20 coefficients

wavelets [17], are band limited but extend over an infinite range in time. Not only are Harmonic wavelets band limited, but they are also box shaped in the frequency domain. This makes them suitable for detecting specific time limited harmonic frequencies within signals.

Dyadic Harmonic wavelets may be defined simply in the frequency domain from Eq.2.40 as shown in Figure 2.12.

$$W(\omega) = (\frac{1}{2\pi})2^{-j}e^{\frac{-i\omega k}{2^j}} \tag{2.40}$$

The inverse Fourier transform of the frequency defined Harmonic wavelet $W(\omega)$ gives its time domain representation,

$$w(2^j t - k) = \frac{\left(e^{i4\pi(2^j t-k)} - e^{i2\pi(2^j t-k)}\right)}{i2\pi(2^j t - k)} \tag{2.41}$$

at level j and translation k. The complex form allows two real wavelets to be expressed in a single expression, much like a Fourier series. At the lowest band, $j = -1$, the inverse Fourier transform produces the scaling function,

$$\varphi(t-k) = \frac{\left(e^{i2\pi(t-k)} - 1\right)}{i2\pi(t-k)} \tag{2.42}$$

As previously stated, a key feature of these wavelets comes from the fact that the scaling function and each dilation set, $W_j(\omega)$ occupies its own unique, non-overlapping space in the frequency domain. It is therefore intuitive that this family of wavelets should form an orthogonal set. The orthogonality properties of the dyadic harmonic wavelet transform are discussed in [17].

In [30] the dyadic harmonic wavelet transform is generalized to allow for equally sized (or pre defined varying) wavelet scales. This increases the versatility of the harmonic wavelets as a basis they are no longer restricted to having high frequency, low

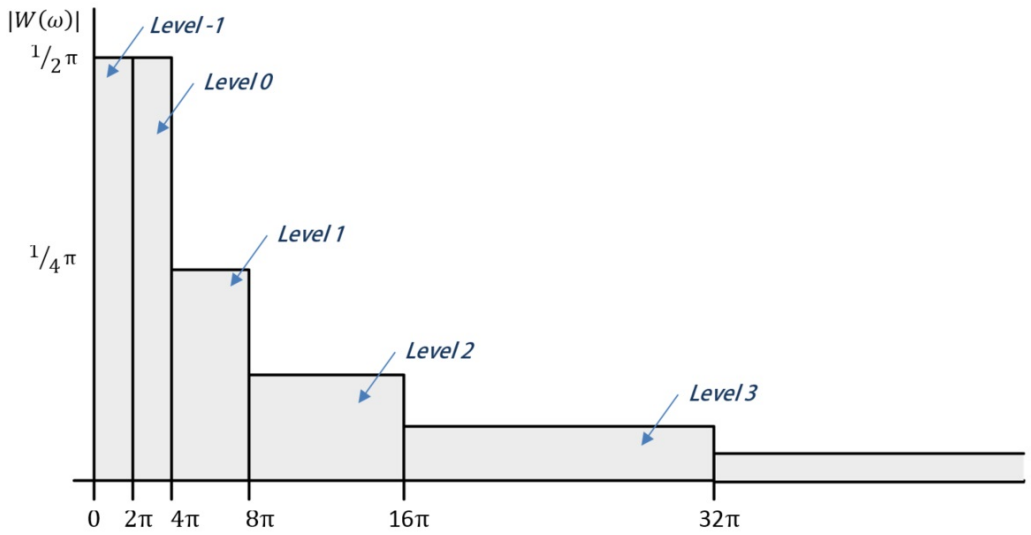


Figure 2.12: Dyadic harmonic wavelets in the frequency domain starting from Level -1

time resolution at low frequencies and low frequency, high time resolution at high frequencies. A generalized harmonic wavelet of (m, n) scale (bounding the wavelet in the frequency domain) and (k) position in time attains a representation in the frequency domain of the form

$$\Psi^G_{(m,n),k}(\omega) = \begin{cases} \frac{1}{(n-m)\Delta\omega} e^{\left(-i\omega\frac{kT_0}{n-m}\right)}, & m\Delta\omega \le \omega \le n\Delta\omega \\ 0, & otherwise \end{cases}, \tag{2.43}$$

where m, n and k are considered to be positive integers and $\Delta\omega = \frac{2\pi}{T_0}$, and where T_0 is the total time duration of the signal under consideration. Harmonic wavelets of the form of Eq.(2.43) span frequency bands defined by m and n as shown in Figure 2.13. An orthogonal set of harmonic wavelets are produced when n and m define adjacent non-overlapping intervals for all wavelets in the set. As $n-m$ nears 1, the wavelet tends towards a single harmonic (high frequency resolution); however, as $n-m$ increases, the wavelet becomes more compressed in the time domain and hence offers higher resolution in time (Figure 2.14). The analyst may choose a single value of $m-n$ for the entire wavelet set, defining a fixed time-frequency resolution for the wavelet transform or vary $n-m$ to increase, or decrease band-dependant time-frequency resolution. The inverse Fourier transform of Eq.(2.43) gives the time-domain representation of the wavelet, which is equal to

$$\Psi^G_{(m-n),k}(t) = \frac{e^{\left(in\Delta\omega\left(t-\frac{kT_0}{n-m}\right)\right)} - e^{\left(im\Delta\omega\left(t-\frac{kT_0}{n-m}\right)\right)}}{i(n-m)\Delta\omega\left(t-\frac{kT_0}{n-m}\right)} \tag{2.44}$$

Furthermore, the continuous generalized harmonic wavelet transform (GHWT) is defined as

$$W^G_{(m,n),k} = \frac{n-m}{kT_0}\int_{-\infty}^{\infty} f(t)\,\overline{\Psi^G_{(m,n),k}(t)}dt, \tag{2.45}$$

and projects the function $f(t)$ on this wavelet basis.

Redundant and non-redundant harmonic wavelet transforms

An efficient numerical implementation of the GHWT used throughout this work that makes use of the FFT is given in [31]. Instead of computing Eq.2.45 for a given

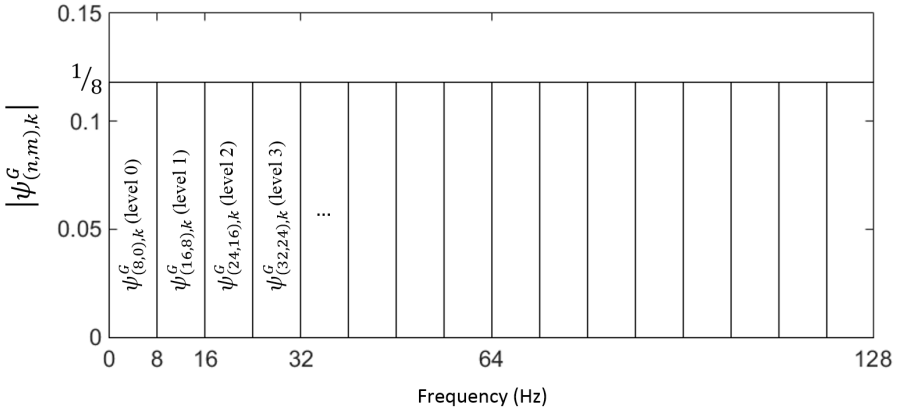


Figure 2.13: Harmonic wavelets in the frequency domain with $n-m = 8Hz$

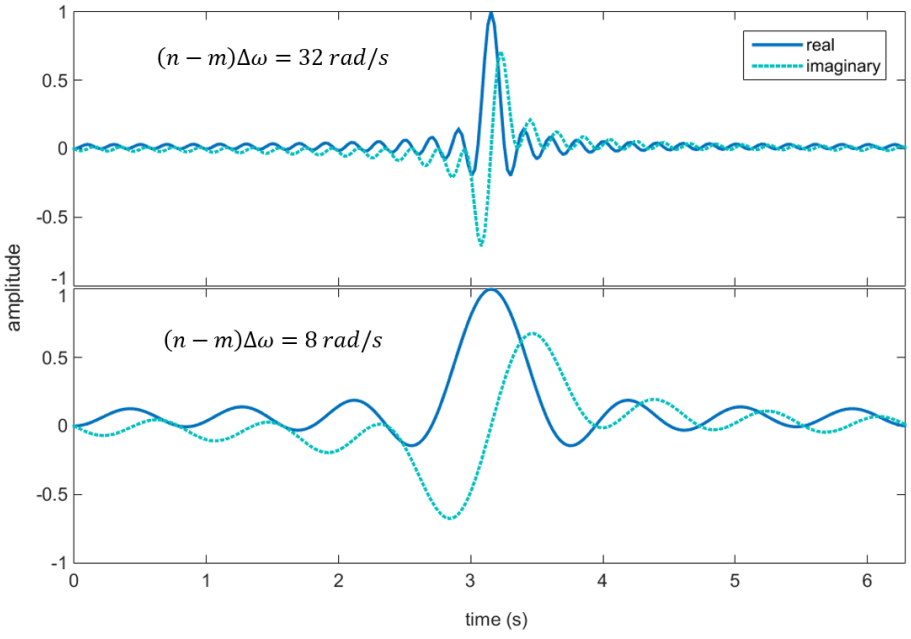


Figure 2.14: Comparison of Harmonic wavelets in the time domain for high (top) and low (bottom) resolution in time

frequency band at every translation k, the circular convolution of the signal and wavelet may be computed efficiently via the FFT. Note that taking the convolution of two signals is equivalent to transforming them to the Fourier domain, multiplying them and then inverse transforming them back to the time domain. This requires significantly less computational effort than multiplying each individual time step (i.e., operations in the order of N^2 are reduced to the order of $N\log N$). This also does not take account of the computational time required to evaluate Eq.2.44 for which the frequency domain representation required for convolution via the FFT (Eq.2.43) is significantly simpler.

The convolution of the signal $f(t)$ with the conjugate wavelet $\Psi^G_{(m,n),k}$ is performed by multiplying the FFT of $f(t)$ with the band-limited frequency domain representation of the wavelet, Eq.2.43 and then performing the inverse FFT. The procedure is illustrated in Figure 2.15. It is important to note that this scheme represents a redundant transform and does not project the source signal onto an orthogonal basis. This is because for each wavelet band, there are an equal number of coefficients to the length of the signal. Hence the entire transform produces many more coefficients than the length of the signal in which the wavelet spectrum at adjacent times is highly correlated [32]. Consider that repeating the calculation in Eq.2.45 for different integer values of k is equivalent to performing a convolution of Eq.2.44 in which $k = 0$ with the signal $f(t)$ in increments of $\frac{T_0}{n-m}$. Then the redundant wavelet transform illustrated in Figure 2.15 is equivalent to performing Eq.2.45 at N integer values of k (spanning the length of the signal) using the following time-domain representation of the wavelet,

$$\Psi^G_{(m-n),k}(t) = \frac{e^{(in\Delta\omega(t-k))} - e^{(im\Delta\omega(t-k))}}{i(n-m)\Delta\omega(t-k)}. \tag{2.46}$$

This redundant transform has the effect of smoothing the wavelet representation of the signal along the time axis and can be useful in identifying time-dependent properties of a signal. However, it is not possible to perform an inverse wavelet transform in a redundant basis (with the aim of reproducing the same original signal) as multiple reconstructed signals would be possible from a single set of wavelet coefficients. The non-redundant form of the GHWT is revisited in Chapter 5.

2.4.4 Other non-stationary spectral analysis techniques

Harmonic wavelets are used throughout this project for a number of reasons. They present a natural, intuitive extension from Fourier analysis into the joint time-frequency domain; indeed a harmonic wavelet basis in which $m - n = 1$ in Eq.2.43 is identical the Fourier series. Also, the GHWT may be applied with a range of time-frequency resolution settings (i.e. variation of $n-m$ in Eq.2.43) whilst maintaining an orthogonal basis; utilizing an orthogonal basis enables exact reconstruction of the original signal from basis coefficients. Another important point is that harmonic wavelets can be utilized to generate a spectrogram showing clear, quantifiable frequency bands rather than 'levels' or 'scales' produced by alternative wavelet bases. Finally, the GHWT was also chosen for ease of interpretation when developing missing data reconstruction methods; due to the band limited nature and orthogonal properties of the GHWT, when working with simulated data, algorithm problems and programming errors are quickly identified. This is a very useful property from a research perspective.

It is important to note that the main focus of this work concerns spectral estimation difficulties associated with missing data and that two of the primary avenues investigated herein (chapters 3 and 4) are based on reconstruction of missing data in the time domain. Therefore, although examples are demonstrated exclusively with Fourier and Wavelet based spectra in this project, the reconstruction techniques developed herein

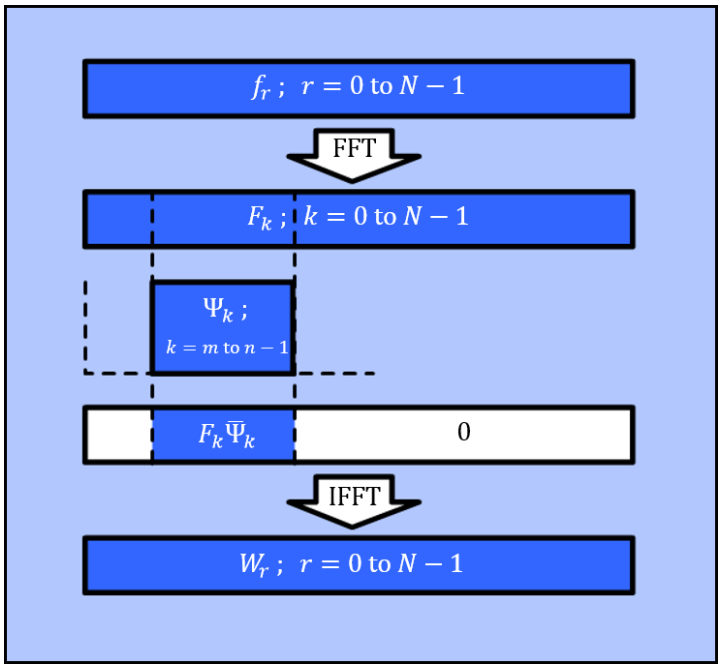


Figure 2.15: FFT algorithm for producing harmonic wavelet coefficients

(including those based on the GHWT) could be utilized before applying alternative spectrum estimation methods. Example of alternative spectrum estimation techniques include:

- The Wigner-Ville method (WVM): This is an alternative approach to both wavelet and time window based non-stationary spectral estimation, named after E.P. Wigner [33] who used it first in quantum mechanics and J.Ville [34] who first proposed its use in harmonic analysis. Explanations of the method can be found in [35, 14].

- The S-transform: This can be thought of as an extension of the STFT or as a type of wavelet analysis based on windowed sinusoids [36]. Where the STFT involves repeatedly analysing the entire frequency content of a signal over a pre-defined fixed window that is moved along the time-history, the S-transform allows for variable sized time windows.

- Chirplet transforms: A chirplet is similar to a wavelet, in that it has localized energy and is oscillatory, with the difference being that its frequency content changes over time. By using an over-defined dictionary of chirplets, the adaptive chirplet transform [14] is able to capture highly non-stationary elements without needing to reduce overall frequency resolution (as would be required to detect similar elements through a standard wavelet transform).

2.5 Stationary process representation & power spectral estimation

The power spectrum is an important concept in the analysis of stochastic processes. From a theoretical standpoint we cannot decompose a stochastic process into a finite Fourier series due to its non-periodic nature. Further, we cannot apply the continuous Fourier transform to a process of infinite length, as the condition

$$\int_{-\infty}^{\infty} |x(t)|\, dt < \infty \tag{2.47}$$

is not guaranteed (required for classical Fourier analysis, preventing infinite magnitude coefficients). Instead, the Fourier transform of the autocorrelation function is considered. This approach makes the assumption that $x(t)$ has no lasting periodic components below some minimum frequency, and hence

$$R_X(\tau \to \infty) = 0 \tag{2.48}$$

The idea is that the autocorrelation function $R_X(\tau)$ gives information about the frequencies present in a random process indirectly whilst being of finite energy. The power spectrum for the process $x(t)$ is given as

$$S_X(\omega) = \frac{1}{2\pi} \int_{-\infty}^{\infty} R_X(\tau) e^{-i\omega\tau} d\tau \tag{2.49}$$

The units of the power spectrum $S_X(\omega)$ are those of (process variance) / (unit of frequency), with the area under the entire curve being equal to the total process variance. However, it can be shown that the power spectrum density can be directly related to the Fourier transform of a stationary stochastic process, according to the theory of 'generalized harmonic analysis' [37].

For any real-valued stationary process, $X\,(t)$, there exists a corresponding complex orthogonal process $Z(\omega)$, such that $X(t)$ can be written in the form Eq.2.50, e.g. [38, 39, 4].

$$X\,(t) = \int_{-\infty}^{\infty} e^{i\omega} dZ\,(\omega) \tag{2.50}$$

Eq.2.50 is a 'stochastic integral' in which $Z(\omega)$ is the Fourier-Stieltjes transform of $X(t)$. If $Z(\omega)$ is differentiable, then Eq.2.50 reduces to the standard inverse Fourier transform. For stationary stochastic processes, $Z(\omega)$ is an orthogonal stochastic process, i.e., for non-overlapping intervals $d\omega$ and $d\omega'$, the corresponding increments $dZ(\omega)$ and $dZ(\omega)'$ are uncorrelated. The average expected power of the process within the frequency interval $d\omega$ is given by $E|dZ(\omega)|^2$. From this the power spectral density function may be defined,

$$E\left(\left|dZ^2\,(\omega)\right|\right) = S_X\,(\omega)\, d\omega \tag{2.51}$$

and

$$E\,(dZ\,(\omega)) = 0. \tag{2.52}$$

In Eq.2.51, $S_X\,(\omega)$ is the two-sided power spectrum of the process $X\,(t)$. Further, a versatile formula for generating realizations compatible with the stationary stochastic process model of Eq.(2.50) is given by [1]

$$X\,(t) = \sum_{j=0}^{N-1} \sqrt{4S_X\,(\omega_j)\,\Delta\omega}\, \sin\,(\omega_j t + \Phi_j) \tag{2.53}$$

where Φ_j are uniformly distributed random phase angles in the range $0 \leq \Phi_j < 2\pi$ and N relates to the discretization of the frequency domain.

Spectral estimation of stationary stochastic processes

Regarding estimation of the power spectrum of the process of Eq.(2.50) based on available realizations, this is given by the ensemble average of the square of the absolute

Fourier transform amplitudes of the realizations [17]; standard established FFT algorithms can be utilized,

$$S_X(\omega_k) = \frac{2\Delta T}{T} E\left|\sum_{t=0}^{T-1} X_t e^{-2\pi i k t/T}\right|^2 \tag{2.54}$$

where T is the number of data points, t is the data point index in the record, ΔT is the sampling time increment, and k is the integer frequency for ω_k (i.e. $\omega_k = \frac{2\pi k}{T_0}$ where T_0 is the total length in time of the record). For situations in which only a single process realization may be analysed, by adopting the assumption of ergodicity, $S_X(\omega)$ can be estimated by computing the temporal mean value of the square of the DFT of the available record in the form (e.g.[1, 17]),

$$S_X(\omega_k) = \lim_{T\to\infty} \frac{2\Delta T}{T}\left|\sum_{t=0}^{T-1} x_t e^{-2\pi i k t/T}\right|^2 \tag{2.55}$$

In practise, infinite length data records are not available. For finite T, Eq.2.55 is known as the Periodogram.

2.5.1 Non-stationary process representation & spectral estimation

Next, for the case of non-stationary stochastic processes, similar to Eq.(2.50), rigorous process representation of non-stationary stochastic processes is presented. In this regard, [40] developed a framework for representing non-stationary stochastic processes by utilizing a time/frequency-localized wavelet basis, as opposed to the Fourier decomposition of Eq.(2.50). The representation reads

$$X(t) = \sum_j \sum_k w_{j,k}\psi_{j,k}(t)\,\xi_{j,k}, \tag{2.56}$$

where $\psi_{j,k}(t)$ is the chosen family of wavelets and j and k represent the different scales and translation levels respectively. $\xi_{j,k}$ is a stochastic orthonormal increment sequence, similar to $dZ(\omega)$ in Eq.2.50 in that pairs of $\xi_{j,k}$ where j and k are not equal are uncorrelated. However, where the process variance was represented within the process $Z(\omega)$ in Eq.2.50, here it is normalized with the process variance represented by a separate, deterministic term, $w_{j,k}$. Specifically, the local contribution to the variance of the process of Eq.(2.56) is given by $|w_{j,k}|^2$.

The wavelet-based model of Eq.(2.56) relies on the theory of locally stationary processes (see also [41]). The aforementioned wavelet based representation can be viewed as a natural extension in the wavelet domain of earlier work related to the representation of non-stationary stochastic processes, e.g. [2, 41].

Next, utilizing the generalized harmonic wavelets, Eq.(2.56) becomes (see [42])

$$X(t) = \sum_{(m,n)} \sum_k \left(X_{(m,n),k}(t)\right), \tag{2.57}$$

where

$$X_{(m,n),k}(t) = \sqrt{S^X_{(m,n),k}(n-m)\,\Delta\omega}\,\psi_{(m,n),k}(t)\,\xi_{(m,n),k} \tag{2.58}$$

Eq.(2.58) represents a localized process at scale (m,n) and translation (k) defined in the intervals $[m\Delta\omega, n\Delta\omega]$ and $\left[\frac{kT_0}{n-m}, \frac{(k+1)T_0}{n-m}\right]$, wheras $S^X_{(m,n),k}$ represents the spectrum

$S_X(\omega, t)$ at scale (m, n) and translation (k). In [42] it has been shown that under certain assumptions, Eq.(2.58) can be written in the form

$$X_{(m,n),k}(t) = \int_{m\Delta\omega}^{n\Delta\omega} e^{i\omega\left(t-\frac{kT_0}{n-m}\right)} dZ_{(m,n),k}(\omega), \tag{2.59}$$

with the properties

$$E\left(dZ_{(m,n),k}\right)(\omega) = 0, \tag{2.60}$$

and

$$E\left(\left|dZ_{(m,n),k}(\omega)\right|^2\right) = S^X_{(m,n),k}(n-m)\Delta\omega \tag{2.61}$$

Further, it has been shown that realizations compatible with $S_X(\omega, t)$ can be generated by utilizing a generalization of Eq.(2.53) of the form (see [3])

$$X(t) = \sum_{j=0}^{N-1} \sqrt{4S_X(\omega_j, t)\Delta\omega}\sin(\omega_j t + \Phi_j) \tag{2.62}$$

In the ensuing analysis and, specifically, in the numerical examples section, stationary and non-stationary process realizations are generated by utilizing Eq.(2.53) and Eq.(2.62) respectively. Note that the power spectrum of a non-stationary process $S_X(\omega, t)$, resolved in both time and frequency, is often referred to as the evolutionary power spectrum (EPS).

Harmonic wavelets based power spectrum estimation

Regarding the problem of estimating the EPS of a non-stationary stochastic process based on available/measured realizations, a wavelet process based compatible estimation approach advocates that the EPS $S_X(\omega, t)$ of the process $X(t)$ is estimated by [43, 42]

$$S_X(\omega, t) = S^X_{(m,n),k} = \frac{E\left(\left|W^G_{(m,n),k}[X]\right|^2\right)}{(n-m)\Delta\omega}, \quad m\Delta\omega \le \omega \le n\Delta\omega, \quad \frac{kT_0}{n-m} \le t \le \frac{(k+1)T_0}{n-m}, \tag{2.63}$$

where $S^X_{(m,n),k}$ represents the EPS of the process $X(t)$, assumed to have a constant value in the intervals $[m\Delta\omega, n\Delta\omega]$ and $\left[\frac{kT_0}{n-m}, \frac{(k+1)T_0}{n-m}\right]$. Thus, the EPS can be estimated as the ensemble average of the square of the wavelet coefficients.

Mitigation of end-effects due to the Harmonic Wavelet Transform

As an extension to section 2.3.5, the presence of end-effects in the GHWT and their mitigation is discussed here. As the discrete GHWT is a circular transform (similarly to the DFT), unwanted artefacts appear in the spectral estimate for non-periodic signals. Because the GHWT resolves a signal's frequency content in time, end-effects can be seen in the wavelet domain more predominantly at the beginning and end of the sample. The area of the wavelet spectrum most effected by these end effects is referred to as the "cone of influence" (COI) [32]. As is discussed in this section, one method of preventing the signal from wrapping around onto itself is to pad the time domain signal with zeros at the beginning and end. In [32], the COI is then defined as the e-folding time for the autocorrelation of wavelet power at each scale. In this case the e-folding time is defined as the distance at which the power of the wavelet transform of a Dirac drops by a factor e^{-2}. Naturally for dyadic wavelet schemes, this COI will appear as an arc on the spectrum. I.e., the low frequency components (with short bandwidth) will have

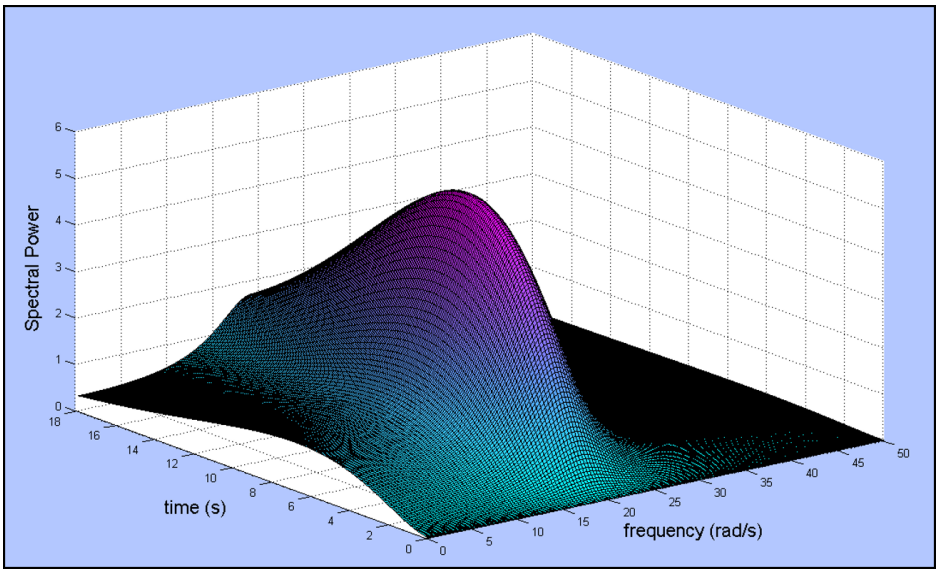


Figure 2.16: Example non-stationary power spectrum process model

a wider influence as they are characterised by broad wavelets in the time domain, and high frequency components (with wide bandwidth) will have a thinner influence as they are characterised by shorter wavelets in the time domain (see Figure 2.14). However, the GHWT allows for user defined wavelet bandwidths, usually these are chosen to be of equal width. Hence the COI will be of equal size along the time axis for all wavelet bands. In this section, as the true (source spectrum) is known, the effect of the COI is considered along side other limitations of the wavelet transform as a factor of the total error between the target and output spectra for a given case. It is not considered in significant detail for the remainder of this work, [15] provides a more detailed description of the COI and [27] provides a method of feature extraction within the COI.

Figure 2.16 shows an EPS based on a time-modulated stochastic process model (Eq.2.65) introduced in section 2.5.2. Notice that the spectrum starts at $t = 0s$ from zero power across all frequencies (i.e. no signal), and grows to its peak somewhere around $8s$, diminishing towards the end of the sample $t = 18s$. By the end of the sample however, the signal has not completely diminished, and still shows positive power across a range of frequencies. If we generate samples from this spectrum via Eq.2.62, and then estimate the original spectrum via Eq.2.63, the disparity between the power at $t = 0s$ and $t = 18s$ is averaged between the two ends. The effect is seen in Figure 2.17, most clearly at $t = 0s$ where the GHWT estimated spectrum no longer shows zero power.

An initial solution to the problem is proposed, based on the assumption that the frequency content of a non-stationary signal outside the measured data is likely to be more similar to its closest recorded frequency content in time than its furthest (furthest being the case when the signal is assumed periodic). Although this cannot be proven for any signal for which there is no information beyond the measured interval, it seems intuitively far more likely (especially if the frequency content varies slowly with time), than a chance measurement of length equal to the period of a periodic signal. The solution is simply to pad the original signal with its reverse on both sides. For example if Figure 2.18a was a recorded time history of a non-stationary random process and the padding size had been decided as one half the length of the measurement, half of each side of the signal would be copied (Figure 2.18b). These copies would then be reversed and used to pad the original signal (Figure 2.18c). By applying this reverse

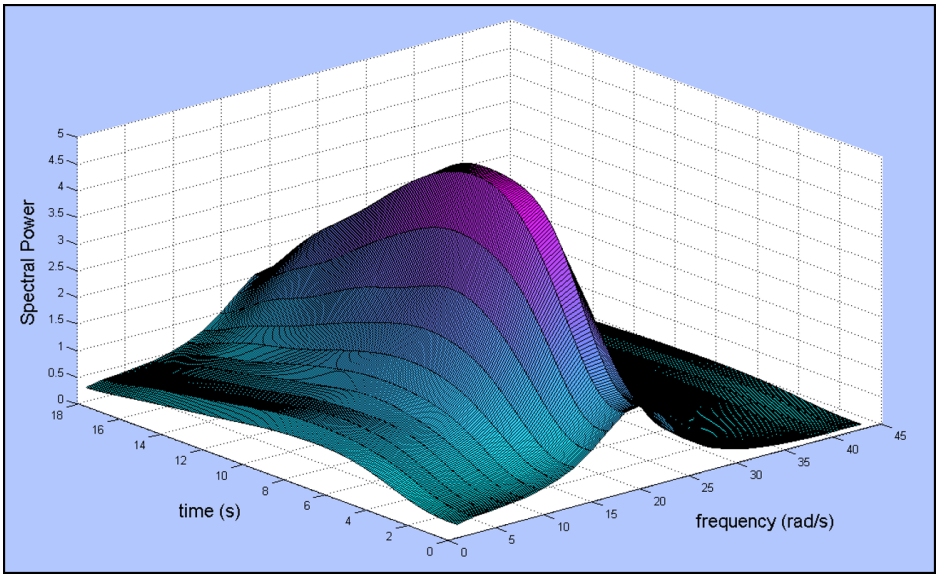


Figure 2.17: Reconstruction of Figure 2.16 via GHWT from 500 realizations

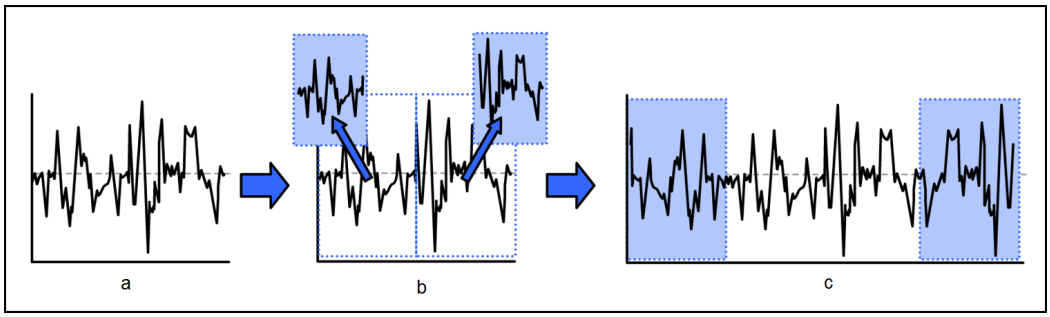


Figure 2.18: Depiction of reverse signal padding

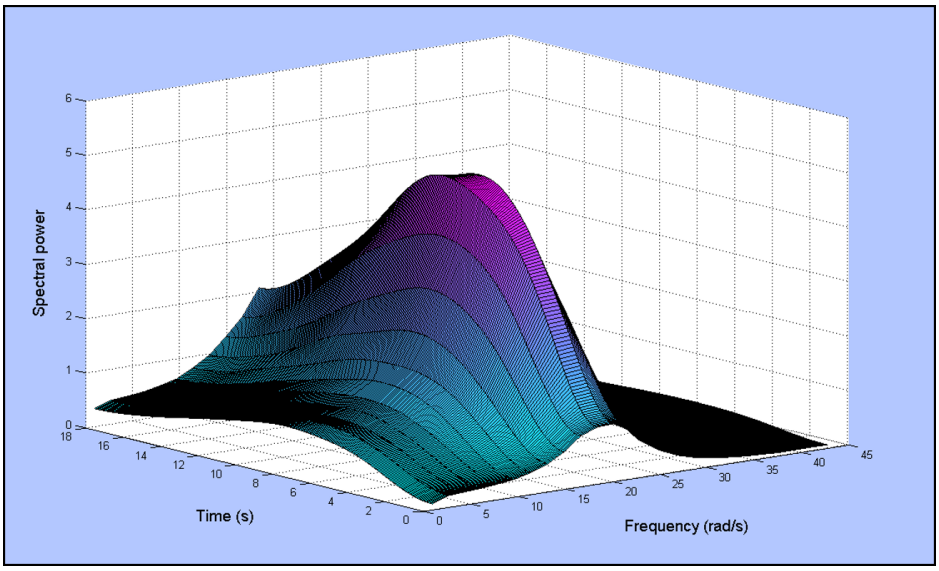


Figure 2.19: Reconstruction of Figure 2.16 via GHWT from 500 realizations with reverse signal padding

padding method to the realizations used to estimate the spectrum in Figure 2.17, the edges of the new GHWT estimated spectrum at $t = 0s$ and $t = 18s$ need not be equal, and approximate more closely to the original Figure 2.16 as shown in Figure 2.19. Another possible approach to mitigating end-effects in the GHWT is realized by considering multiple wavelet resolutions in the same basis. As explained in section 2.4.2, the choice of harmonic wavelet bandwidth presents a trade-off between time, and frequency resolution. A small bandwidth $(n - m)$ will give a low resolution in time and a wide bandwidth will give a high resolution in time. Figures 2.17 and 2.19 are reconstructed using the GHWT with $(n - m)\Delta\omega = 5$ rad/s; this may be considered a relatively high frequency resolution in this case. As a result, resolution in time is low. This resolution determines the amount of assumed signal data that leaks into the wavelet map from outside the measurement. At the expense of loosing frequency resolution at the very beginning and end of the signal, larger frequency bandwidths can be chosen for the GHWT. These can be made wider in steps for the same wavelet map representing incrementally shorter moments in time, as the transform nears the edges of the signal, potentially negating the cone of influence almost entirely. For $p = \max(n - m)$ and incrementally halving the bandwidth near the signal edges, the harmonic wavelets in the frequency domain for a signal of length N would be defined with

$$(n-m) = \begin{cases} 1, & n \leq 2 \\ \frac{n}{2}, & 2 < n \leq p \\ p, & p < n \leq N-p \\ \frac{-(n-N)}{2}, & N-p < n \leq N-2 \\ 1, & n > N-2 \end{cases}, \tag{2.64}$$

the effect is shown in Figure 2.20. For this particular spectrum, when compared to standard GHWT, reverse signal padding reduced the root mean square (RMS) error between the original and estimated spectra by 23% and by 36% when using the multi-resolution approach. These values do not mean much other than to demonstrate that significant improvements can be made by considering end-effects, even in a simple manner when estimating the wavelet spectrum. Further investigation into the mitigation of end-effects is not pursued in this work. For the majority of the examples, reverse signal padding was used in GHWT based spectrum estimation.

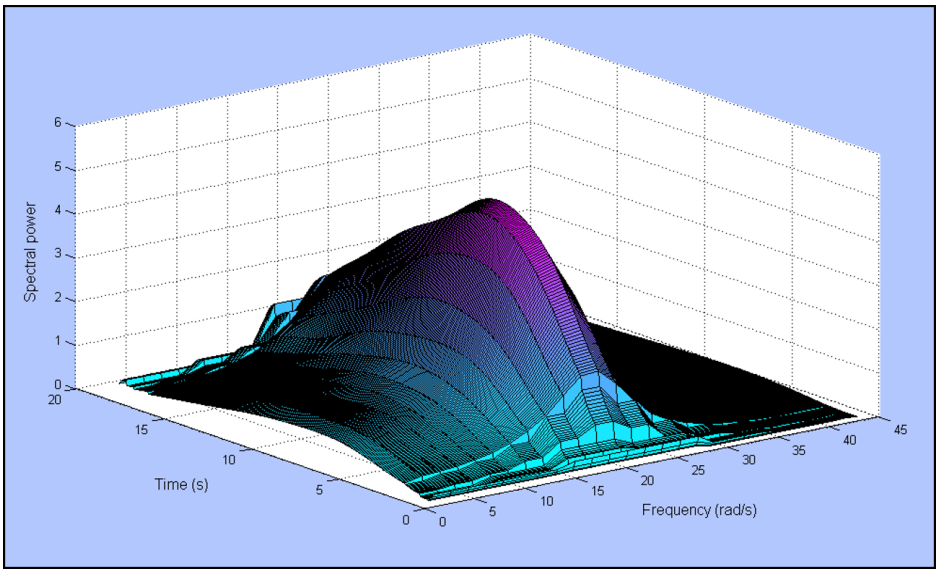


Figure 2.20: Reconstruction of Figure 2.16 via GHWT from 500 realizations with decreasing frequency resolution near the edges $t = 0s$ & $t = 18s$

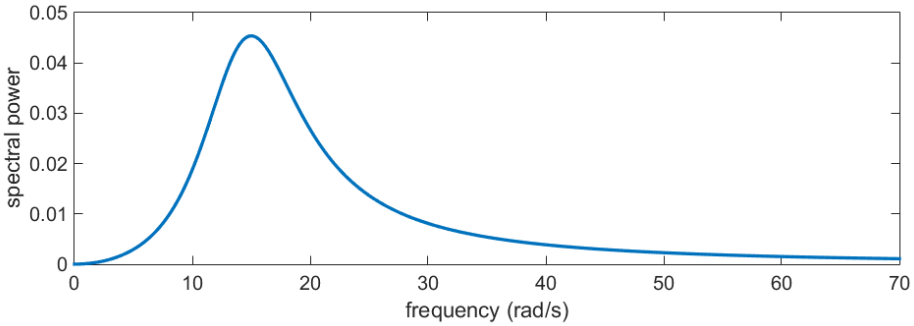


Figure 2.21: Example stochastic process spectrum from Eq.2.65 with $a = 0.5$, $\zeta = 0.35$ and $\omega_g = 15$

2.5.2 Commonly used example spectra in this work

In the following chapters, numerical examples will be shown to demonstrate various methods developed for spectrum estimation under missing data. In order to produce results, source data is needed which may be generated from a known power spectrum, via Eq.2.53 and Eq.2.62 for stationary and non stationary processes respectively. For this purpose a range of pre-defined spectral models are considered, mostly based on environmental processes. The most commonly used in this work is a simple process model, not specifically relating to a particular environmental process and is manipulated heavily to demonstrate different features of process reconstruction methods described herein. In its basic form, based on that shown in [44], this spectrum is defined as

$$S(\omega) = \frac{1 + \alpha\omega^2}{\left(\omega_g^2 - \omega^2\right)^2 + \left(2\zeta\omega\omega_g\right)^2}. \tag{2.65}$$

An example of this spectrum is shown in Figure 2.21. Other spectra that are referred back to in the numerical examples sections of the upcoming chapters are listed below.

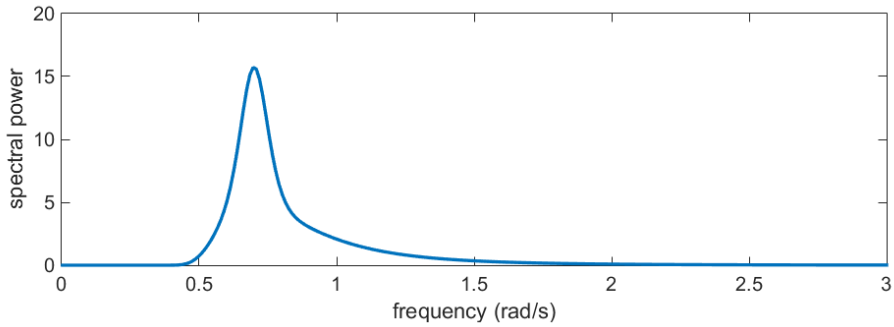


Figure 2.22: Example JONSWAP spectrum from Eq.2.66 with $\alpha = 0.03$, $\omega_p = 0.7$, $\gamma = 3.3$, $\sigma = 0.07$ for $(\omega \leq \omega_p)$, and $\sigma = 0.09$ for $(\omega > \omega_p)$

JONSWAP

The JONSWAP spectrum is a stationary sea wave model based on data collected during the Joint North Sea Wave Observation Project [45]. This spectrum typically has a very sharp, strong peak and occupies a relatively narrow band in the frequency domain. The JONSWAP spectrum has the form,

$$S(\omega) = \frac{ag^2}{\omega^5} e^{-\frac{5}{4}\left(\frac{\omega_p}{\omega}\right)^4} \gamma^r; \quad r = e^{-\left(\frac{\omega-\omega_p}{2\sigma\omega_p}\right)^2} \tag{2.66}$$

An example of a JONSWAP spectrum is shown in Figure 2.22

Clough-Penzien

The Clough-Penzien stationary spectral model [46] is a modified Kanai-Tajimi [47, 48] spectrum. The Kanai-Tajimi stationary spectral model and its many variants are widely used in earthquake engineering. The spectrum describes strong ground vibrations caused by an earthquake, and is usually arranged to have a broader band than a JONSWAP process. The purpose of the Clough-Penzien modification is to generate an earthquake process with zero mean and no ultra low frequencies, but with a steep gradient toward the peak harmonic band.

$$S(\omega, t) = S_0 \frac{\omega^4}{\left(\omega_f^2 - \omega^2\right)^2 + 4\zeta_f^2\omega_f^2\omega^2} \cdot \frac{\omega_g^4 + 4\zeta_g^2\omega_g^2\omega^2}{\left(\omega_g^2 - \omega^2\right)^2 + 4\zeta_g^2\omega_g^2\omega^2}, \tag{2.67}$$

An example of a Clough-Penzien spectrum is shown in Figure 2.23.

Non-stationary spectra

So far the spectral models listed have been stationary only. To generate non-stationary processes a time-dependant spectral model is required. In many cases stationary spectral models may be simply modulated by a second function in the time domain to produce a non-stationary model. Eq.2.68 is an example of such a function,

$$g(t) = ke^{-at} - e^{-bt}. \tag{2.68}$$

Using an envelope function in this way modulates all frequencies equally, and the resulting spectrum is referred to as "separable" (i.e. time and frequency components evolve independently). Figure 2.24 shows an example envelope function. This is used

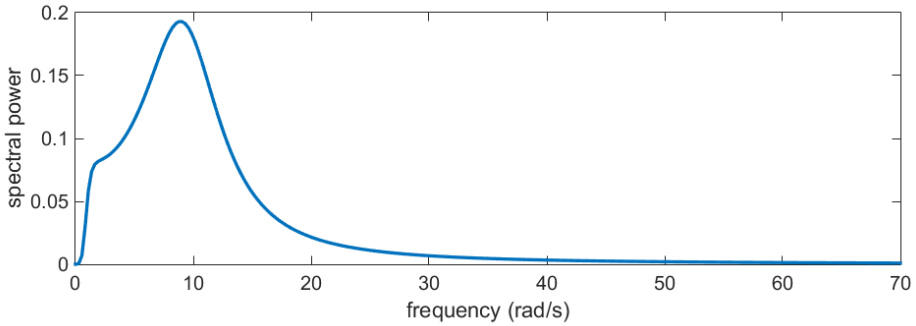


Figure 2.23: Example Clough-Penzien spectrum from Eq.2.67 with $S_0 = 0.07$, $\omega_f = 1$, $\zeta_f = 0.6$, $\omega_g = 10$ and $\zeta_g = 0.4$

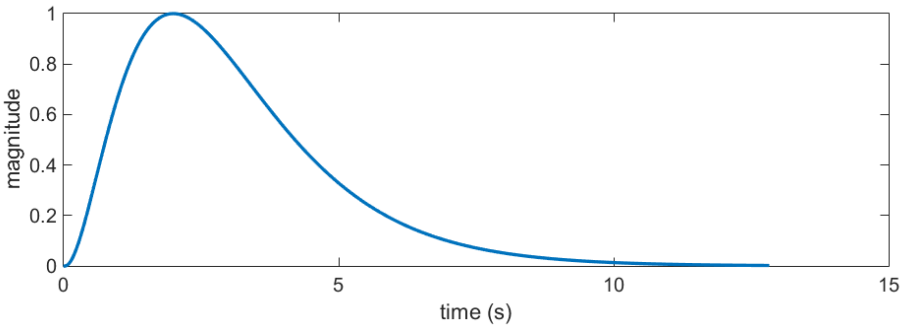


Figure 2.24: Example envelope function from Eq.2.68 with $k = 4$, $a = 0.3$ and $b = 0.6$

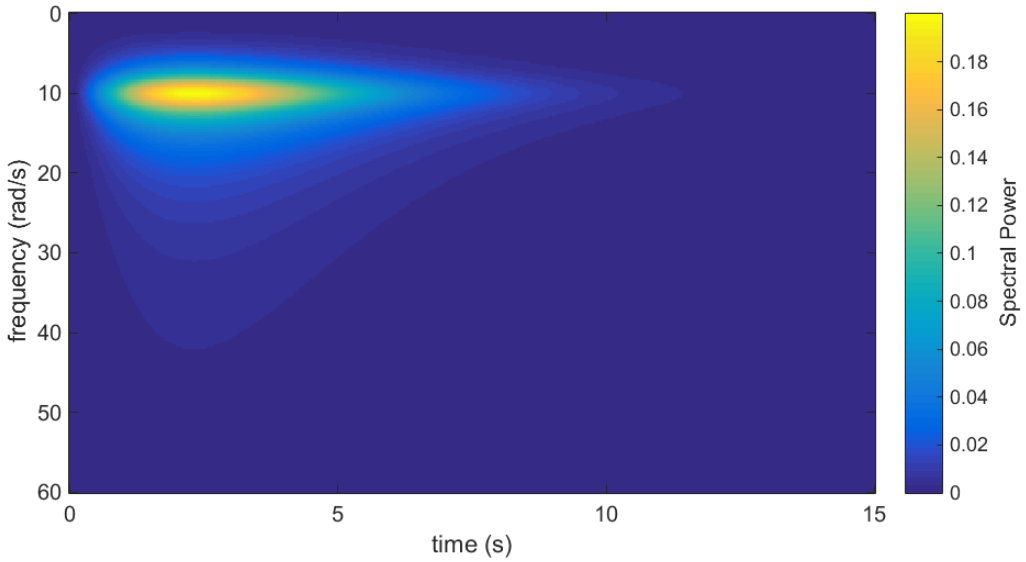


Figure 2.25: Example time-modulated stochastic process model based on Eq.2.65

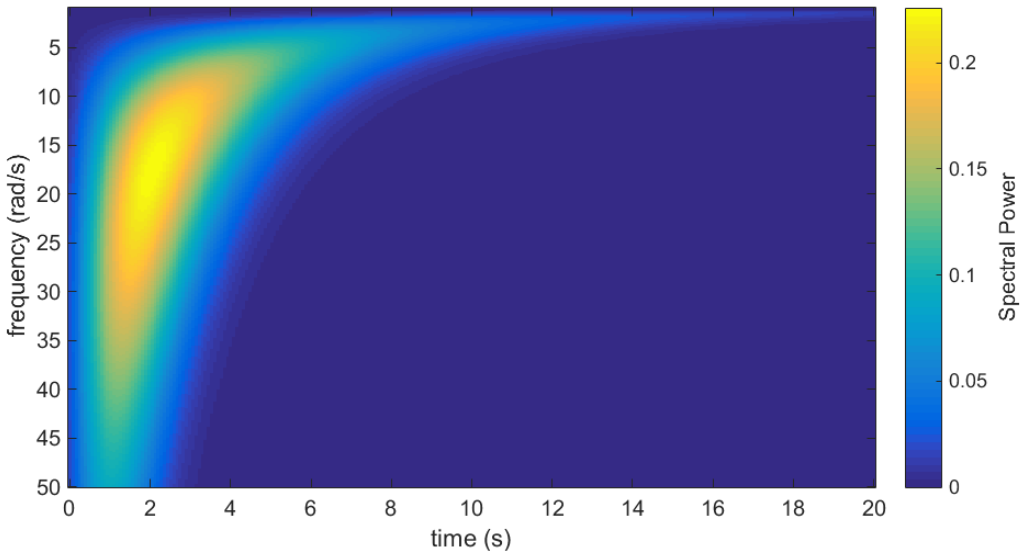


Figure 2.26: Example non-stationary, non-separable spectrum from Eq.2.69

to modulate Eq.2.65 to produce the separable non-stationary spectrum shown in Figure 2.25.

Finally another non-stationary earthquake power spectrum is used, taken from [43] which is also non-separable. For many real non-stationary processes, frequency dependent power does not vary with time uniformly across the spectrum. Eq.2.69 comprises some of the predominant features of seismic shaking, such as decreasing of the dominant frequency with time [49] (example shown in Figure 2.26),

$$S(\omega, t) = \left(\frac{\omega}{5\pi}\right)^2 e^{-0.15t} t^2 e^{-\left(\frac{\omega}{5\pi}\right)^2 t}. \tag{2.69}$$

2.6 Definition of missing data & associated issues

When analysing real excitation data gathered from environmental processes, coupled with the problem that there may be a limited number of samples or shorter than ideal sample lengths, is the common and major issue of missing data. In the literature, "missing data" is generally defined as a loss of discrete datum values in a regularly sampled time-series e.g., [23, 5, 6]. This is not the same as the more general case of "uneven sampling" where missing data do not need to occur on some regular sampling grid e.g., [50]. Examples in this work all conform to the former definition of missing data with the possibility of extending to the more general case of uneven sampling, however, the methods given in Chapter 4 are immediately applicable to both cases. Figure 2.27 (top) shows an arbitrary process sampled at regular intervals joined with a solid line. Figure 2.27 (bottom) shows that same time-series with random missing data (the few points that appear next to each other at the original sampling frequency are joined by dotted lines). One of the primary aims of this project was to look into methods of estimating a power spectrum model (stationary or non-stationary) from discrete-time data suffering from gaps in the manner shown in Figure 2.27. Due to the breadth of possible sources of missing data in practical settings (as discussed in the main introduction), the topic of spectrum estimation from gappy data is not new. As

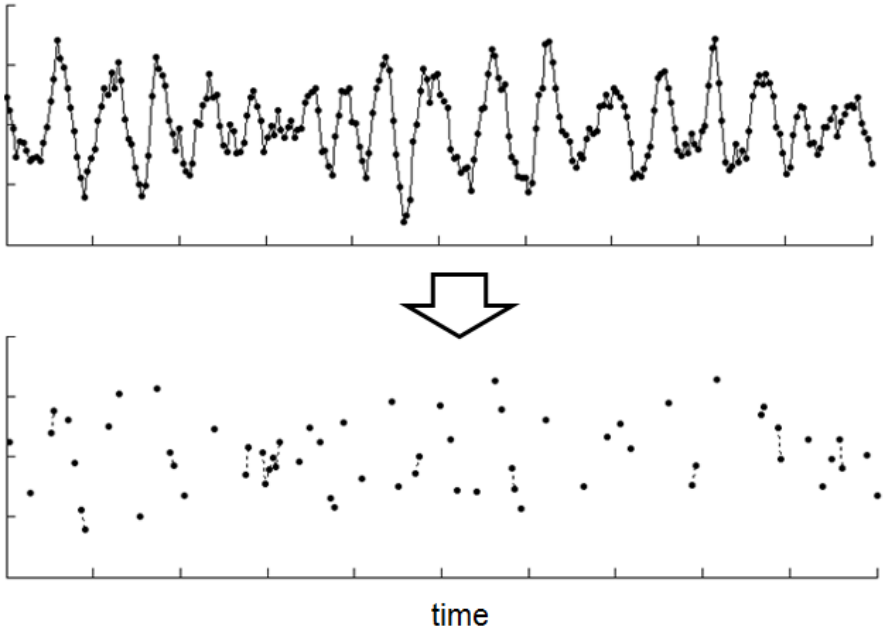


Figure 2.27: Complete regularly sampled time-series (top) compared to the same signal with missing data (bottom)

such there are tools available for dealing with these problems, though many come with restrictions and assumptions concerning the nature of the original signal.

In general, if a measured realization of a stochastic process is available for analysis with no further information than that contained within the sample, it is impossible to predict with certainty what lies beyond the known time interval. Similarly, if data points are missing within the same sample, it is impossible to predict them with certainty. Further, if the focus is not on predicting exactly what happens within intervals of missing data, but instead, on determining a complete time history to be used in a Monte Carlo simulation analysis for instance, then predicting the missing data in an "average sense" can be satisfactory.

For instance, in a Monte Carlo analysis it is often required that millions, or even billions of samples are considered to formulate reliable probabilities for events [51]. Obviously, it is not always possible to collect that many real time histories; thus, they must be simulated in an appropriate manner. In this regard, an appropriate spectral analysis can be conducted on the few available measured time histories, and then new time histories with the same statistical characteristics can be further generated for a larger scale analysis. If the available measured time histories contain gaps, the problem is not that the time histories no longer contain important underlying process information (and, thus, they should be discarded), but that traditional spectral analysis techniques are not equipped to identify this information (i.e. they require a uniformly indexed time series input). The DFT is one of these cases; in order to perform a Fourier analysis when presented with such data it is necessary to re-sample onto a uniform grid. There are different ways in which the data may be re-sampled. The method of re-sampling onto a uniform grid can be chosen based on speed, accuracy and also on any knowledge of the underlying process from which the signal was generated. For example if it is known that a time-history containing gaps is generated from an underlying process containing only a single harmonic within a known frequency range, then a simple regression of a sine wave onto the available data should provide an acceptable means of interpolation over the gaps. This specific problem is however

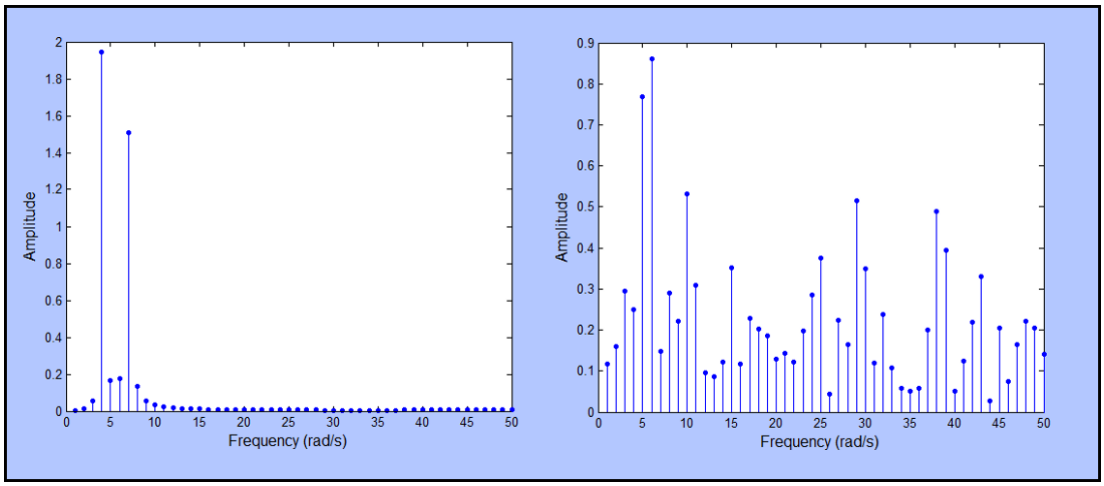


Figure 2.28: Fourier transform of a signal containing two sinusoids from a complete, uniformed sample set (left) and with 40% of the data missing, filled with zeros (right)

very simple; by performing the regression to fill the gaps, to give a uniformly sampled signal, taking the DFT is no longer necessary (except possibly to detect low power high frequencies hidden in noise along the main harmonic). In many real problems involving missing data, insufficient information is known about the underlying process to fill these gaps with a high level of confidence; otherwise there would be little need to conduct a spectral analysis in the first place.

2.6.1 Available tools for working with missing data

Selected methods currently in use for spectral estimation under missing data are described here, along with their advantages and limitations.

Zero-padding

One of the simplest and perhaps most common approaches to dealing with missing data in Fourier analysis is to fill gaps in the time series with zeros and then perform the DFT [22]. When the ratio of zero padding to recorded data is small (around 5% - 10%), the power spectrum is not usually greatly effected. Therefore this method is suitable not only for padding the ends of the signal, but also for replacing missing data in these cases. In fact, in the case of a Fourier analysis, as the signal is resolved into its non-time dependant frequency components, there is no real difference in placing zeros at the end of a record or within the body of the record. Unfortunately for larger numbers of missing data, this method introduces significant levels of noise into the spectrum estimation, to the extent where important frequency peaks cannot be identified. Figure 2.28 shows the DFT for a dual-sinusoid signal from a single time-history sampled regularly against one with 40% of the data, removed and replaced with zeros (in uniformly distributed random locations).

Lomb-Scargle periodogram

Lomb [7] and Scargle [8] presented a method of least-squares spectral analysis that can be used to calculate the power spectrum for unevenly spaced data, known as the Lomb-Scargle periodogram. Sine and cosine waves are matched with the signal via least-squares which negates the requirement for uniformed samples. Unfortunately as

each power is calculated independently, the total power at the end of the transform is not equal to the total power of the original signal, and particular frequencies can be easily over or under represented. A similar method of least-squares spectral analysis was presented by Vaníček [9] which involves solving a single linear equation matching all frequencies simultaneously. Power is conserved in this case, assuming the least-squares fit is near perfect. A disadvantage of this is that the number of frequency components must be less than half the number of available data points (otherwise there are multiple solutions). Appropriate frequencies are chosen via matching pursuit, rendering this a computationally intensive solution. Both of these least-squares approaches however could be modified to work from a wavelet basis rather than a Fourier basis for non-stationary analysis. The Lomb-Scargle periodogram estimate of the spectral power at a single frequency is given by

$$S_X(\omega) = \frac{1}{2}\left(\frac{\left[\sum_j X_j \cos\omega(t_j-\tau)\right]^2}{\sum_j \cos^2\omega(t_j-\tau)} + \frac{\left[\sum_j X_j \sin\omega(t_j-\tau)\right]^2}{\sum_j \sin^2\omega(t_j-\tau)}\right) \tag{2.70}$$

with the time offset τ defined by

$$\tan 2\omega\tau = \frac{\sum_j \sin 2\omega t_j}{\sum_j \cos 2\omega t_j}. \tag{2.71}$$

It is important to note here the difference between "uneven sampling" and our definition of missing data. We have introduced missing data with the assumption that the original or target data was evenly sampled. If the data is sampled N times randomly over an interval of continuous time, the Lomb-Scargle periodogram presents a useful tool for spectrum estimation. However, if it is possible to pad missing points with zeros to produce a uniformly sampled time-history (as in the previous section), then the DFT of this re-sampled signal will give a similar result to the Lomb-Scargle periodogram. Most of the results in this thesis are based on the more restrictive definition of missing data given at the start of this section. It is worth noting however that, the methodology used in Chapter 4 could be easily modified for completely unevenly sampled data.

CLEAN deconvolution algorithm

A method of spectral analysis with incomplete data that has been shown to work with seismological data [12] is presented by [13], based on the CLEAN deconvolution algorithm [52]. The CLEAN algorithm provides an iterative method of removing unwanted artefacts in the Fourier domain that occur as a consequence of performing the DFT with zeros in place of missing data. The method stems from the fact that in the discrete case of the Fourier transform, the spectrum of the original signal is not produced, but in fact the original signal multiplied by a sampling function. For evenly spaced data with no gaps this is a series of Dirac deltas, for data with gaps this sampling function is a mix of deltas and zeros (corresponding to where the sample is known and not known). The Fourier transform of a function sampled at s is therefore

$$\mathrm{DFT}(f_s(t)) = \mathrm{DFT}(f(t)s(t)) \tag{2.72}$$

where $\mathrm{DFT}(f(t))$ is the true spectrum and $\mathrm{DFT}(f_s(t))$ is known as the 'dirty' spectrum. The product of the function with the sampling function $f(t)s(t)$ becomes a convolution in the frequency domain. The CLEAN algorithm attempts to perform an iterative deconvolution of the dirty spectrum with the known spectrum of the sampling function. By doing this it can separate artefacts generated by the sampling function from the real

data. It has been shown to work well on signals with few harmonic components and with small levels of noise [12], but as the number of harmonics increases, the algorithm becomes less likely to be able to identify the correct elements to remove [13]. Therefore this method may be useful for identifying highly sparse spectra with sharp peaks, but not so applicable for wider band processes.

Interpolation methods

Another intuitive approach is to interpolate between the points directly. Basic methods include linear and polynomial interpolation. For time-histories with uniformly scattered missing data, interpolation can give superior results to zero-padding with little computational effort. Interpolation has the potential to conserve some of the power of the original signal that was lost, where as zero-padding does not. However, there are two significant drawbacks when using interpolation methods. Firstly, high frequency power components are likely to be underestimated in the spectrum. Interpolation has a smoothing effect on the signal and as a result, boosts the low frequency components and dampens high frequency components. Secondly, if the missing data is grouped densely over constant time intervals, interpolation will give a poor approximation to the original signal, introducing a low frequency discontinuity and likely altering the signal mean.

Autoregressive estimation

Another approach to filling in the gaps in a sample is given in [10] and is based on an autoregressive model of the process. Observed data that is assumed to come from a stationary harmonic process with noise can be approximated by an autoregressive model of the form

$$x_t = \sum_{j=1}^{L} a_j x_{t-j} + e_t. \tag{2.73}$$

Each time step value x_t can be given by a linear combination of previous values $a_j x_{t-j}$. e_t is the error of each prediction (in this case the forward prediction error, as the formula predicts new steps from previous ones). This error is minimised through an iterative least-squares procedure to give the optimal values for a_j. L denotes the order of the model (how many previous time steps are taken into consideration when generating the next). The proposed algorithm fills gaps by fitting autoregressive models to available data and using them to estimate the unknowns. A severe restriction of this method is that the final order of the autoregressive model must be shorter than the shortest data segment. This means that the method could be virtually unusable depending on the spacing of the data, since short autoregressive models are not able to capture low-mid range frequencies in the data.

2.6.2 Simulation of missing data

In the following chapters, novel solutions to the problem of missing data in spectral analysis are demonstrated with numerical examples. In this regard, missing data must be simulated. Two different arrangements of missing data are applied. The first case simulates missing data at random locations drawn from a uniform distribution of the time index,

$$f_0(t) = \begin{cases} f(t), & r_a \geq m \\ 0, & r_a(t) < m \end{cases} \tag{2.74}$$

where $f_0(t)$ is the sample time history with missing data, $f(t)$ is the original sample generated from its power spectrum, r_a is a vector of N_0 equally spaced numbers from

0 to 1 arranged in random order, and m is the fraction of missing data. The second case simulates missing data that occur in groups, positioned at random locations, again drawn from a uniform distribution of the time index,

$$f_0(t) = \begin{cases} f(t), & v(t) = 1 \\ 0, & v(t) = 0 \end{cases} \tag{2.75}$$

where $v(t)$ is given by

$$v(u) = \begin{cases} M, & r_b(u) \leq 1/k \\ 0, & r_b(u) > 1/k \end{cases} \tag{2.76}$$

and k is the number of intervals, M is a vector of ones of length $N_0 \times m/k$, and r_b is a vector of $N_0 - (N_0 \times m - k)$ equally spaced numbers from 0 to 1 arranged in random order.

Process ensemble missing data

When simulating missing data for multiple process records used for the same spectral estimate, the missing sample points are re-generated randomly for each sample unless otherwise stated. In chapter 3 examples of missing data occurring in prolonged intervals are given in which the missing data occurs in the same locations for each record in the set.

2.7 Chapter Summary

The theoretical background to the doctoral investigation has been provided in this chapter. Here, the author has drawn upon an extensive literature review of the issues associated with stochastic process power spectrum estimation and missing data; alongside a critical appraisal of relevant methods, which included a review of their advantages and limitations. This chapter has also benefited from the wide-ranging insights the author has gained throughout the doctoral study: from not only supervisors and colleagues involved in risk and uncertainty work at the host university, but also from discussions with other researchers and external collaborators at national and international meetings and conferences. This has placed the thesis into an emerging area of scholarly outputs, and affirmed the author's choice of novel approaches for investigation, herein to follow as original research in Chapters 3, 4 and 5.

Chapter 3

Artificial neural network approaches for power spectrum estimation subject to missing data

3.1 Introduction

In this chapter, an Artificial Neural Network (ANN) approach, e.g. [53] is developed for estimating the power spectrum of stationary and non-stationary stochastic processes subject to missing data. First, an appropriately defined ANN is employed to capture the stochastic pattern in the available data in an "average sense". Next, the ANN, having stored process trends within its connection weights, is exploited for generating new data to fill sampling gaps fitting with the underlying stochastic process. Finally, power spectrum estimates are derived by utilizing standard Fourier analysis (stationary case), or wavelet based EPS estimation approaches (non-stationary case). Several numerical examples are included to demonstrate the reliability of the approach.

3.1.1 Introduction to artificial neural networks

In a general sense, a neural network is a tool designed to model the way a biological brain solves problems. Even small brains of rodents are able to perform complex perceptual recognition tasks in meagre fractions of the time it would take a modern computer. Therefore by attempting to simulate some of the known functions of a biological brain it is possible to perform certain computational tasks more efficiently. The two main simulated features of any ANN are the ability to acquire knowledge through some learning process and then store that knowledge. To achieve this, ANNs employ large numbers of simple computing junctions commonly referred to as "neurons", which exhibit interconnectivity by way of weighted neuron links or "synaptic weights". The weights allow the network to store acquired knowledge which may change fluidly based on new information - as in nature.

Neurons and network learning

A basic model of a neuron with attached synaptic weights is shown in Figure 3.1. Each of these individual processing units exhibit features that are common amongst most neural network frameworks.

1. Inputs to the neuron are passed through synaptic weights. These weights are basically gains that multiply the input signals. The value of the gain for each link can be changed independently of the others and may be updated as the network is tuned to perform a specific function.

2. The weighted inputs are then passed through a summing junction which also includes a bias. The bias may also be updated as the network is tuned to control the magnitude of the effect of the inputs on the network output.

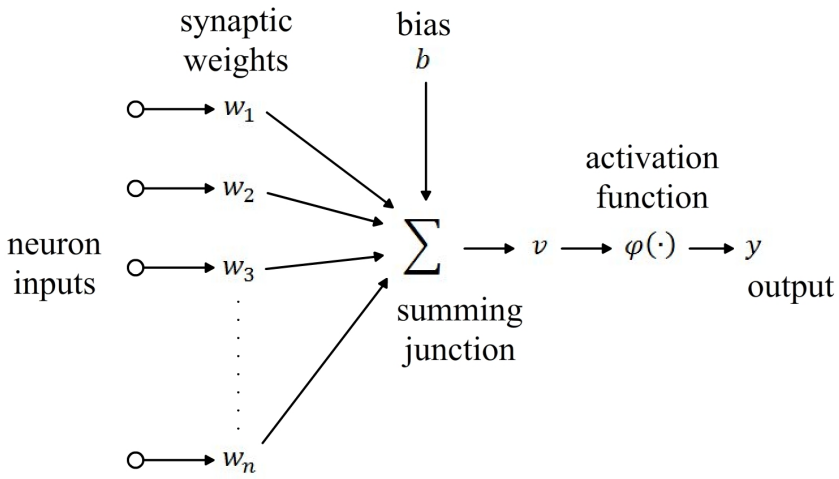


Figure 3.1: Model of ANN neuron with synaptic weights

3. After the summing junction, the neuron will typically pass the signal through an activation function. This tends to take the form of a limiting function over a closed interval - typically normalized to $[0, 1]$ or $[-1, 1]$.

In order for an ANN to learn, it requires some learning algorithm that alters the synaptic weights between neurons. Networks with pre-defined topologies may be tuned to perform specific tasks by simply altering their synaptic weights. However, some networks may also modify their own topology during the learning process, killing off weak connections and growing new ones.

Benefits of neural networks

ANNs are able to perform tasks that exhibit a high level of non-linearity. A single neuron component of an artificial network is typically non-linear. When these neurons are set in a interconnected web, the network itself will also be non-linear. By having distributed non-linearity throughout the network, the ANN is suitable for processing highly non-linear signals.

A neural network is able to learn to map a set of inputs to a set of outputs in a 'model-free' sense. This means that no prior assumptions need to be made on a statistical model for the input data [53]. The network learning algorithm modifies synaptic weights based only on a given set of training samples. During a single learning cycle, the network is presented with one of the training samples, containing both input and desired output data. The weights are altered to minimize the difference between the output that the network produces and the desired output. By repeating this procedure multiple times on the training set, the network will reach a steady state, where changes in weights become negligible.

3.2 Artificial neural network based stochastic process simulation

An approach for simulating stationary stochastic processes given a short input sample with the aid of artificial neural networks (ANN) has been developed in [44]. In this regard, the short recorded sample is used to 'train' a neural network to recognize the pattern, so as to capture the properties of the underlying process. Once trained, the

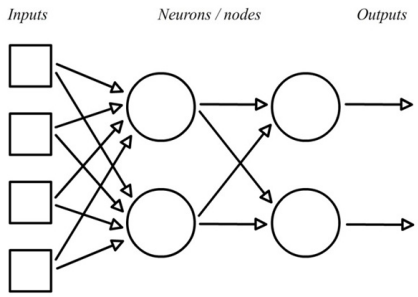


Figure 3.2: Example feed-forward ANN architecture (4-2-2)

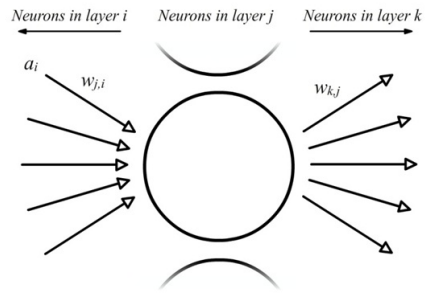


Figure 3.3: A single neuron (or perceptron) in an ANN

network can then be used to generate more samples with similar properties. Although this procedure does not account for gaps in the samples, it does present a solid starting point open to modification.

In [44] it was shown that ANN based procedures already provide a suitable tool in cases of limited (short length) sample data and are able to simulate stochastic processes. Further, as the neural network operates model-free, realizations are generated based only on perceptions gained from the available data. Hence, an ANN approach to stochastic process simulation can be applied more generally than other methods such as Auto Regressive (AR) and Auto Regressive Moving Average (ARMA). This is shown in [54] for short term rainfall forecasts and in [55] for drought prediction where ANNs were compared with traditional regression techniques including AR and ARMA models and shown to be superior.

3.2.1 Network architecture

The manner in which ANN neurons are linked is referred to as "network architecture" or "network topology" and is closely linked with the type of learning algorithm employed in training the network. The ANNs employed in this chapter are based on a multi-layer feed-forward structure. In this setting data flows from the input layer to the output layer through multiple layers of neurons but not in the opposite direction. This is also a common layout for function approximation and an exapmple is shown in Figure 3.2 (each of the circular elements represents an individual neuron as shown as part of the network in Figure 3.3 and individually in Figure 3.1). The logistical sigmoid function is used as the activation function for each neuron,

$$\varphi = \frac{1}{1+e^{-x}}. \tag{3.1}$$

Eq.(3.1) is chosen, since it is a continuous non-linear function with limited outputs and is easily differentiable (required for gradient descent network training explained below); therefore the output of each neuron is given by,

$$y_{j,n} = \left(1 + e^{-\sum_{m=0}^{N_{j-1}}\left[w_{j,j-1}(n,m)y_{j-1,m}\right]}\right)^{-1}, \tag{3.2}$$

where $y_{j,n}$ is the output of the n^{th} neuron in the layer j, j, N_{j-1} is the number of neurons in layer $j-1$ (plus the bias), $w_{j,j-1}(n,m)$ is the weight of the connection from neuron m in layer $j-1$ to neuron n in layer j, and $y_{j-1,m}$ is the output of the m^{th} neuron in layer $j-1$.

3.2.2 Back-propagation learning

As previously stated, the objective of a supervised network learning algorithm is to minimize the difference between the network output and desired output for a given set of training data. To this end, the networks are trained via a method of error correction through gradient descent, for multi-layer networks this is known as the back-propagation algorithm (as the error contribution from each weight is propagated back through the network); see [53]. The contribution of each neuron to the error in the output is calculated and used to alter synaptic weights, minimizing the error. The error is given by:

$$e_k(n) = d_k(n) - y_k(n) \tag{3.3}$$

where $e_k(n)$ is the error of the n^{th} output at level k. As in a lest squares optimization, the value to be minimised is the error energy, $E(n) = \frac{1}{2}\sum_{k\in C} e_k^2(n)$ (where C is the set of output neurons and the scale of $\frac{1}{2}$ serves to simplify the expression when differentiated). The amount by which the weights must be altered is based on a fraction of their contribution to the error. The contribution of a single weight to the error of an output neuron it is feeding is defined as the change in output error with respect to its input weight. The change in error for a single neuron can be written as a product of partial differentials,

$$\frac{\partial E(n)}{\partial w_{kj}} = \frac{\partial E(n)}{\partial e_k(n)} \cdot \frac{\partial e_k(n)}{\partial y_k(n)} \cdot \frac{\partial y_k(n)}{\partial v_k(n)} \cdot \frac{\partial v_k(n)}{\partial w_{kj}(n)}, \tag{3.4}$$

- $\frac{\partial E(n)}{\partial e_k(n)}$ is the change in error energy with respect to the actual error and is equal to $e_k(n)$.
- $\frac{\partial e_k(n)}{\partial y_k(n)}$ is the change in error with respect to the neuron output and is equal to -1.
- $\frac{\partial y_k(n)}{\partial v_k(n)}$ is the change in neuron output with respect to the summed neuron input. This is the differential of the activation function. For the sigmoid chosen (Eq. 3.1), its derivative is $\varphi(v_k) \cdot (1 - \varphi(v_k))$.
- $\frac{\partial v_k(n)}{\partial w_{kj}(n)}$ is the change in summed neuron input with respect to the single weight feeding into it. This is the output of the previous layer, $y_j(n)$ that passes through the weight $w_{kj}(n)$.

Therefore the change in weight on a single connection for a single training sample is given by:

$$\Delta w_{kj} = \eta e_k(n) \cdot \varphi(v_k) \cdot (1 - \varphi(v_k)) \cdot y_j(n) \tag{3.5}$$

Where η is the learning rate usually set between 0 and 1. From Eq.3.4, the sign of Eq.3.5 should be negative, however this is altered to account for the desired direction of the gradient decent (i.e., to reduce the error, not increase it). Eq.3.5 may be used directly for output layer neurons but for hidden layers, the individual errors are not known. Therefore the error at the output layer must be propagated back through the network to determine the weight changes at hidden neuron connections. For a hidden layer that is one layer behind the output layer, the weight change is calculated by,

$$\frac{\partial E(n)}{\partial w_{ji}(n)} = \frac{\partial E(n)}{\partial y_j(n)} \cdot \frac{\partial y_j(n)}{\partial v_j(n)} \cdot \frac{\partial v_j(n)}{\partial w_{ji}(n)} \tag{3.6}$$

where,

$$\frac{\partial E(n)}{\partial y_j(n)} = \sum_k \left[e_k \frac{\partial e_k(n)}{\partial y_k(n)} \cdot \frac{\partial y_k(n)}{\partial v_k(n)} \cdot \frac{\partial v_k(n)}{\partial y_j(n)} \right] \tag{3.7}$$

Therefore the change in weight for a single connection on the last hidden layer is,

$$\Delta w_{kj} = \eta \left(\sum_k e_k \cdot \varphi(v_k(n)) \cdot (1 - \varphi(v_k(n))) \cdot w_{kj}(n) \right) \cdot \varphi(v_j(n)) \cdot (1 - \varphi(v_j(n))) \cdot y_i(n) \tag{3.8}$$

Weight changes for deeper hidden layers are calculated in the same manner once all of the weight changes are known for the layer above.

To simplify notation, particularly when dealing with more than one hidden layer, a local gradient is defined, δ, pointing to changes for individual weights. This is given by Eq.3.10 for output layer weights and Eq.3.9 for hidden layer weights where subscript m denotes the layer of any required hidden weight. m is used in the hidden case to avoid confusion with the specifically referenced layers i,j & k.

$$\delta_m = \frac{\partial y_m(n)}{\partial v_m(n)} \sum_{m+1} \left[\delta_{m+1} \frac{\partial v_{m+1}(n)}{\partial y_m(n)} \right] \tag{3.9}$$

$$\delta_k = e_k(n) \frac{\partial v_k(n)}{\partial y_j(n)} \tag{3.10}$$

The change in weight at any neuron in any layer is therefore given by,

$$\Delta w_{m,m-1}(n) = \eta \delta_m y_{m-1}(n) + \alpha \Delta w_{m,m-1}(n-1) \tag{3.11}$$

where α is the 'momentum factor', a fraction of the previous weight change added to the current. This has the effect of causing the network to 'centre in' on local minima quickly, greatly increasing the rate of error convergence.

Initial weights are chosen from a uniform distribution of random numbers, scaled to the magnitude within which they are expected to operate (depending on the average magnitude of input and bounds of the activation functions). The weights are updated multiple times for all available training data.

A dual hidden-layer, single-output network was found to work well with the majority of simulated Gaussian stochastic processes tested (based on power spectrum models introduced in section 2.5.2).

3.2.3 Network application scheme

The network is used to output the next point along a time-history, given some number of previous values. This allows the network to learn a process independently of time as a whole; (as the time between inputs is constant the network can learn time-dependent correlations, but not how behaviour may change beyond the end of the sample). The main design features of the network are based on those described in [44], a short summary of which is given in this section.

Assuming there is at least one time history available on which to train the network, the network must ultimately be capable of producing future time steps within the time history given a set of inputs. The network must be set up to receive these inputs, corresponding to a short interval of the available time history, and contain a number of hidden layers, neurons and a single output predicting the next value after the input interval. The choice of interval input is a trade-off between size and number of available unique training sequences. A large interval will contain more information about the process than a short interval; however if large intervals are used, considering that the training data is limited, there will only be a few unique training sets. The network is less likely to learn the underlying process with few training sets, and will perform poorly when tested on new data. Also if the interval is too small, the network will not have enough information required to learn wider correlations.

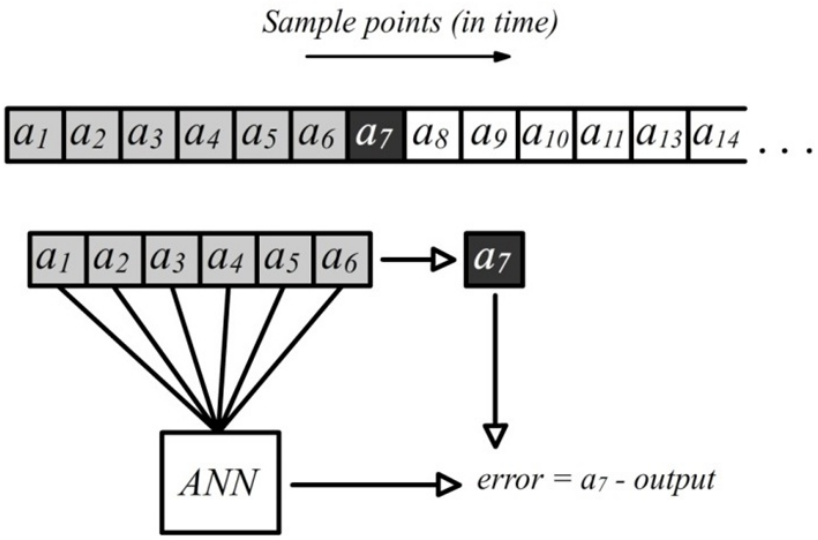


Figure 3.4: ANN input selection for training

The inputs must always be presented in the same order. The number of training sets is therefore equal to the number of sample points minus the number of inputs. Each training set contains a small sample of the whole, with the specified number of inputs plus one to check against the output (Figure 3.4).

The order of the training sets is then randomized before each is fed into the network. Once all have been passed through the network, the process is repeated with the sets in a different, random order.

When utilizing an ANN for function approximation (rather than stochastic process simulation), the training is commonly terminated based on some validation criteria. This can take the form of a training set that is kept separately; it is not used to train the network, but its error is still calculated. During the training procedure, this is likely to pass a minimum threshold, determining the point at which the training stops. The network is unlikely to fit tightly to the training data (errors near zero), but instead gives the best overall approximation to the function in the presence of noise. This is counter-productive in the case of simulating stochastic processes; a tight fit is important so that all of the features are taken into account as part of the process, and this includes fine features that may appear as noise. Therefore the network is instead trained until the error is considered to be sufficiently small (several orders of magnitude lower than the process variance). Once the network is trained, a set of new input data can be used for which the output is not known. The output can then be attached to the end of the input sequence, and the new input shifted one step forward in time (discounting the first of the last input and including the last prediction output) (Figure 3.5). This process can be repeated indefinitely to produce a new time series of any length.

3.3 Stationary stochastic process simulation subject to missing data

Changes to the network application scheme described previously are required if the same ANN process estimation approach is to be considered in the presence of missing data. Gaps in a training sample cannot simply be removed and the sample shortened, because for the network to learn the process the inputs must be evenly spaced in time.

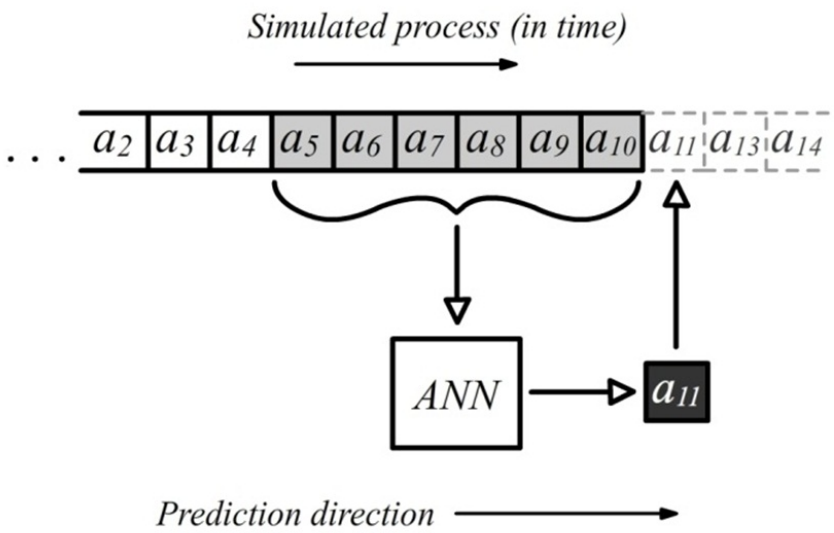


Figure 3.5: ANN process prediction procedure

It is proposed that the missing data points are filled with random values drawn from a random probability distribution (Figure 3.6).

This distribution is estimated empirically based on the known input data. First, a histogram is drawn based on the available data, then arranged cumulatively to build a discrete CDF. To approximate the continuous CDF of the process, the individual CDF points are joined via a polynomial spline. Random values fitting with the original process may then be drawn by passing generated uniform random numbers through the approximated CDF, i.e.,

$$a_i = F^{-1}(X)\left(U_{(0,1)}\right) \tag{3.12}$$

where F^{-1} is the inverse CDF of the known data, and $U_{(0,1)}$ is a random value drawn from a uniform distribution between 0 and 1.

Because the function of the network is stored over all of the weights instead of certain components in separable areas [53], the output should not be drastically affected if a few inputs are not accurate. However this method alone would, over many training cycles, generate a network which predicts an entirely new process based as much on the randomly generated missing data as on the known data. This is why, for each training cycle, the random values filling the missing points are replaced with new random values. By doing this, the network becomes sensitive only to the known data.

The number of training sets will now be smaller than if there were no missing data. While useful to fill input gaps to the network with random values, checking the output against a random value would serve no purpose and only hinder network training. Therefore all of the training sets that would have previously been valid, but for which the final value in the set is missing, are discounted (Figure 3.7).

Training a network with random inputs prevents the overall prediction error from decreasing indefinitely (as a network with sufficient neurons would if trained on a complete set). The error will instead eventually decrease to some mean value around which it fluctuates with constant variance, as shown in Figure 3.8.

This mean value and error variance is dependent on the number of missing data points. Once these error properties have settled, network training can be stopped, as the random variables are preventing the network from differentiating between properties of the process itself and properties induced by the fluctuating points. The trained network can then be used to fill in the missing data points as in Figure 3.9, or to

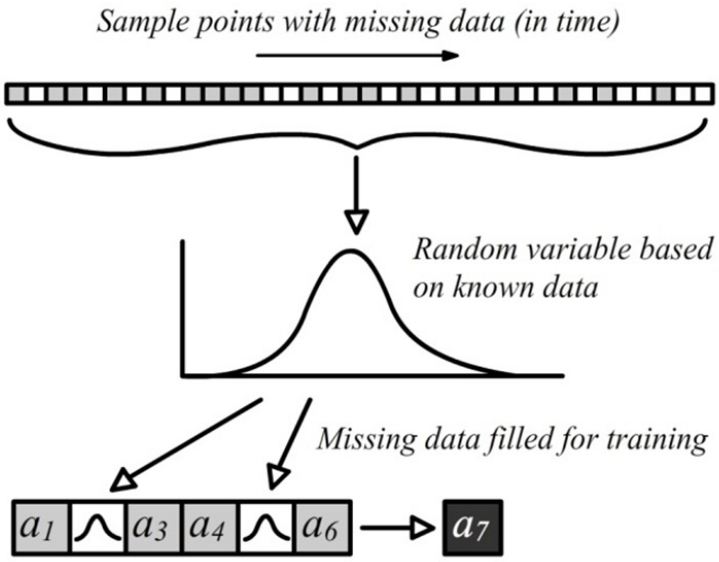


Figure 3.6: Procedure for filling missing data during training on a stationary process

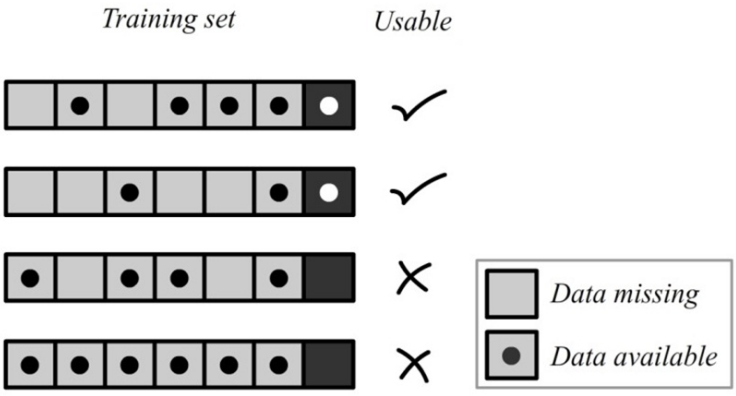


Figure 3.7: Examples of usable training sets when training with missing data

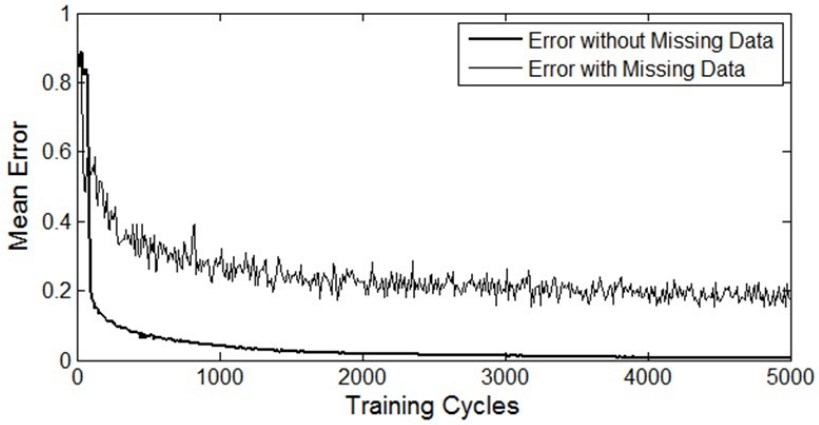


Figure 3.8: Error convergence of ANN trained on samples with and without missing data

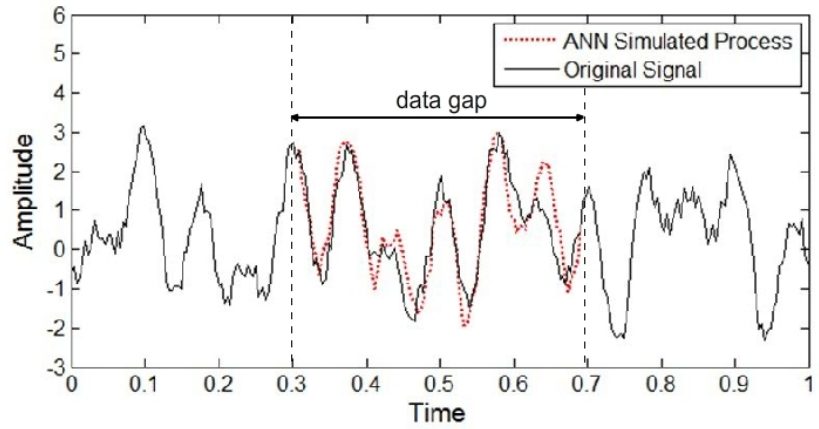


Figure 3.9: ANN trained on data with large gap, then used to fill the gap to give a complete time-history

generate entirely new processes.

3.4 Non-stationary stochastic process simulation subject to missing data

The ANN based approach to process simulation presents a general and highly customizable framework, hence the extension to simulation with missing data. Despite the problem of missing data in non-stationary processes being more complex and unworkable for the majority of irregular data-spectral estimation procedures, the ANN can be adapted to account for changes over time. There are two major differences between the ANN training procedure for stationary and non-stationary processes, these are as follows:

1. The random variable used to fill all missing points in the time-history is no longer valid, as it is drawn from a single distribution covering the entirety of the known process. To overcome this problem, multiple distributions can be drawn that are dependent on time, i.e. they are only based on locally windowed subsets of data (Figure 3.10), making a temporary assumption that the process is locally stationary. This does add in an additional user specification to the network, as the

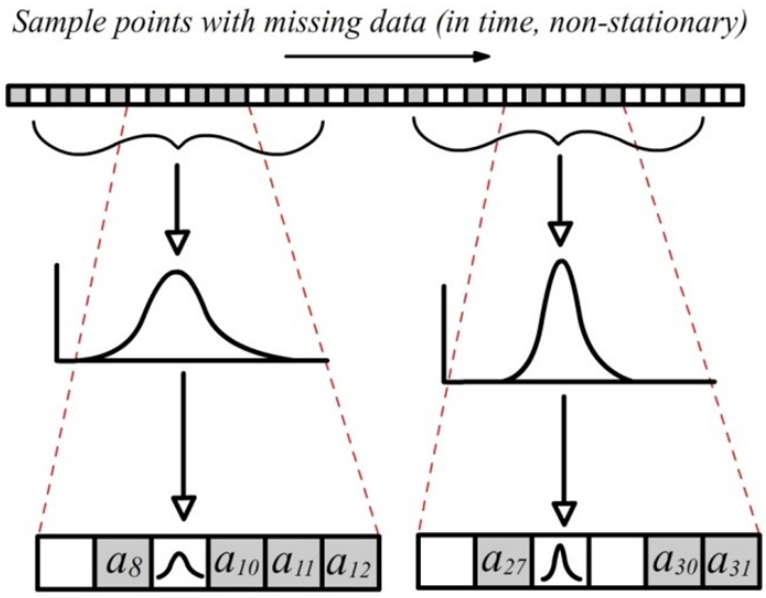


Figure 3.10: Procedure for filling missing data during training on a non-stationary process

window size must be known (how locally stationary the process is assumed to be). Choosing a window that may be too long or even using the distribution of the entire sequence will not prevent the network from simulating a non-stationary process, though training convergence will likely be at a higher mean error. Choosing a window that is too short however is more likely to cause a significant undesired bias, as it tends towards a solution that is based on a single set of random values. The problem of choosing a window size and the assumption of local stationarity can be avoided by using a distribution built from the ensemble of available process samples in time. Although using this method alone would only be appropriate when the number of available samples is large enough to build a full distribution, a trade-off is also possible by using a combination of ensemble points over a tighter window. In this case, the assumption of local stationarity would still exist, but over much smaller local intervals.

2. The second major difference is that in the case of the non-stationary process training, an additional input is added into the ANN. It was found that by adding the time index associated with the current training sample as an extra input, the network was able to better simulate the non-stationary trends in the process. Because each of the training samples have a unique time signature, it is possible to add an extra input at the end (or at any position as long as it is consistent) to represent this. Although the sample order is randomized each time the network is trained, the network is still made aware of the sample position in time due to this identifier. In the case of training a network to learn a stationary process, the time index is not required, as the output is dependent only on a given set of inputs, regardless of time. With these changes, the network is capable of filling gaps in non-stationary process histories with time dependent simulated data.

3.4.1 Mechanization of the approach

To further elucidate the main steps of the ANN approach developed in section 3.3, a flowchart is provided, (Figure 3.11), summarizing the main features.

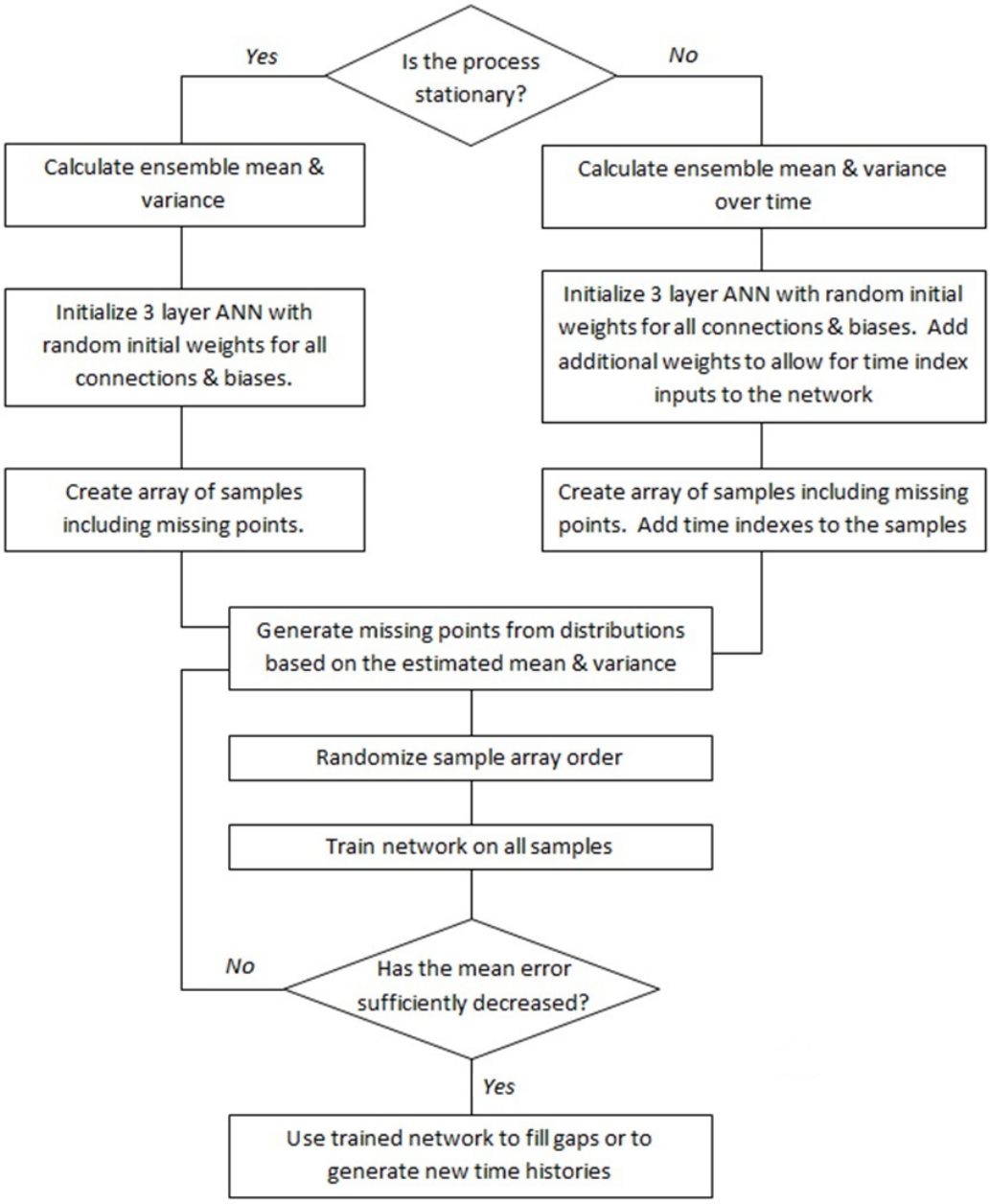


Figure 3.11: Flowchart depicting step-by-step ANN approach to process learning subject to missing data

3.5 Numerical examples

In the ensuing examples, the aim of the ANN simulation is to re-draw the original power spectrum governing a process (which given real data would be unknown). The estimated spectrum is then not only useful for generating new processes for a Monte-Carlo analysis, but also for conducting a spectral analysis for other applications (e.g. system identification / damage detection). The spectral representation of a process however does not provide a complete picture. In the context of harmonic signals, the phases or differences in phase have been lost. As in the following example, time histories are based solely on the power spectrum, this is not a problem (as the phase differences are uniformly randomized). However in many cases, given real data, there could be important phase correlations that need to be simulated in the Monte-Carlo analysis. The ANN, to an extent, is capable of learning these phase differences as well as the frequencies and power shown in a spectral plot. Therefore, processes generated directly by the network, (after feeding an initial set of random inputs fitting with the statistics of the known data), may provide a more reliable simulation set for Monte-Carlo analysis. Also in cases where there are too few processes to be able to effectively estimate the spectrum (i.e., 10 or less), it would still be possible to use the trained networks to generate many new processes. In both stationary and non-stationary examples, 25 sample time histories are utilized for each. This is a large enough number to clearly show the benefit of using ANN missing data reconstruction over zero-padding, whilst being of a reasonable number that one might expect to have available in a real scenario. Convergence of the ANN reconstruction beyond 25 samples is shown in section 3.5.4.

3.5.1 Stationary example

The initial implementation of the neural network for simulating stationary stochastic processes is based on a 3-layer architecture, 15 inputs, 13 neurons in the first hidden layer, 7 neurons in the second and 1 output. This architecture is a product of trial and error to a large extent, the aim being to find the smallest number of neurons (fastest training time) that can learn the process. There are suggested methods for choosing architectures such as evolving ANNs [56] that are able to create / remove connections during training, but in this context for this problem the given architecture performs sufficiently.

The effectiveness of the procedure is demonstrated by training the network using 25 sample paths generated via Eq.(2.53) compatible with the example stationary process power spectrum, described in section 2.5.2 of the form Eq.2.65, where the natural frequency, $\omega_g = 10 rad/s$, damping ratio $\zeta = 0.25$ and $\alpha = 5$. Two tests are performed, one with 30% of the data removed (Figure 3.12) and one with 50% of the data removed (Figure 3.13), both at random locations drawn from a uniform distribution of the time index (Eq.2.74). Once the networks are trained they are used to fill the gaps in the original training samples to give full, uniformly spaced time histories upon which the FFT based power spectrum estimation can be performed. These individual sample-path spectra are then averaged to give the estimated power spectrum. As previously stated, one of the most common model-free approaches to spectral analysis when dealing with missing data is to simply fill the gaps with zeros; the FFT is then performed and scaled up relative to the number of missing points. This method is applied as well, and is shown in comparison to the proposed ANN approach, demonstrating a clear difference in spectral representation effectiveness.

In the 30% missing data case, the ANN already gives a superior estimation to the scaled zero-filled spectrum; the peak power is clearer and the tail has fewer artefacts at higher frequencies. In the 50% missing data case, the ANN still gives good performance

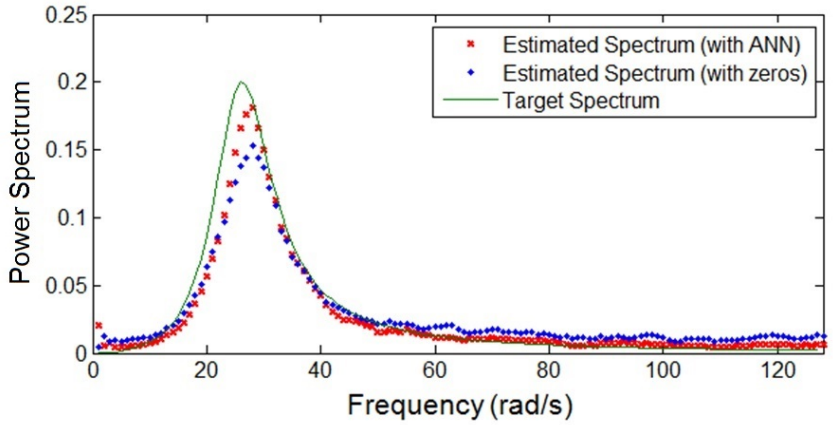


Figure 3.12: Target and estimated spectrums for 25 averaged samples using FFT directly with zeros and FFT of ANN predictions - 30% missing data at random locations

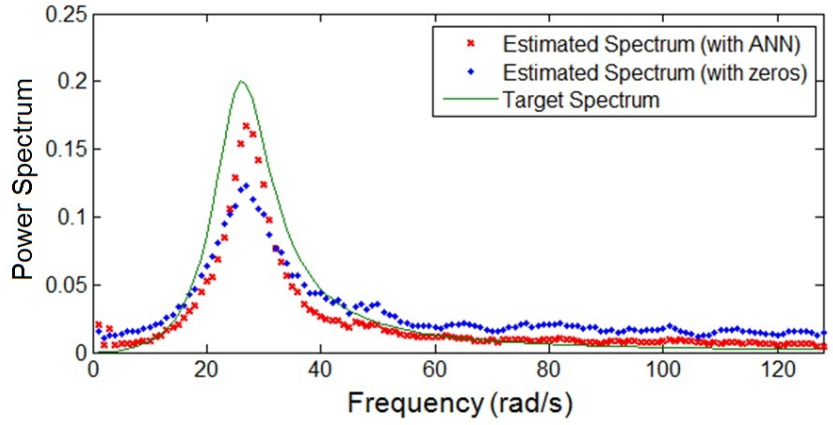


Figure 3.13: Target and estimated spectrums for 25 averaged samples using FFT directly with zeros and FFT of ANN predictions - 50% missing data at random locations

with a well-defined peak spectral power at 28 rad/s. The scaled zero-filled approach has lost significant power at the required spectral peak, and taken on a noisy tail at higher frequencies. These results show that the ANN has indeed learned the underlying spectral properties of the process, and in both cases gives a superior estimation when compared to the zero-filled time-history approach.

3.5.2 Non-stationary separable example

To demonstrate the efficiency of the ANN approach for the case of non-stationary processes, a separable power spectrum of the form,

$$S(\omega, t) = g(t)^2 S(\omega), \tag{3.13}$$

is considered next, where $g(t)$ is the envelope function given by

$$g(t) = b\left(e^{-ct} - e^{-2ct}\right) \tag{3.14}$$

with $b = 4$ and $c = 0.8$. $S(\omega)$ is a modified stationary spectrum based on Eq.2.65 given by Eq.3.15. It can be readily seen that $S(\omega)$ of Eq.3.15 has an additional high frequency spectral peak in comparison with Eq.2.65. This is added to demonstrate the reliability of the ANN approach when simulating processes with multiple dominating frequencies. The resulting EPS is plotted in Figure 3.14.

$$S(\omega) = \frac{1 + \alpha\omega^2}{\left(\omega_g^2 - \omega^2\right)^2 + \left(2\zeta\omega\omega_g\right)^2} + \frac{1 + \alpha\omega^2}{\left(\omega_{g_2}^2 - \omega^2\right)^2 + \left(2\zeta\omega\omega_{g_2}\right)^2} \tag{3.15}$$

In Eq.3.15 above, the natural frequency of the second peak, $\omega_{g_2} = 35 rad/s$. Unlike in the stationary tests, where 25 full (no missing data) time-histories would be enough to give a very close approximation to the target spectrum via an FFT based estimation approach, the GHWT based estimation of Eq.3.15 requires many more sample paths (up to ten times as many) so that a reliable estimate is achieved. Even with a surplus of time histories, end-effects are still present due to the assumed cyclic nature of the process [25]. Because of these factors, a good approximation of the target spectrum (Figure 3.14) via GHWT, averaging for 25 sample paths, would appear as in Figure 3.15. The spectrum has reduced resolution in time and frequency, and the shape is far noisier. It still presents a valid and useful output; however considering these limitations it should, therefore, be noted that even if the ANN were able to perfectly reconstruct the original signal, the best approximation of the target spectrum available utilizing Eq.(2.63) is shown in Figure 3.15, not Figure 3.14.

Two tests are performed, both with 50% of the data removed, one at random locations drawn from a uniform distribution of the time index, and the other at two fixed intervals. 25 individual networks are trained on separate samples of length 512. As with the previous tests, the ANN generates missing data for each time-history, and the estimated power spectrum is obtained (this time by averaged GHWT). These are compared against the output of performing the same analysis with the missing data set to zero. For the first test, an additional comparison was used to show the effect of interpolation to fill the gaps (Figure 3.18). In some cases interpolation may be more appropriate than filling gaps with zeros, but when the higher frequencies present in the signal are near the Nyquist sampling rate (as in these examples), interpolation has the effect of removing them almost completely and introducing low frequencies that should not exist.

When compared with the best possible estimate given 100% of the data (Figure 3.15), Figure 3.16 for 50% missing data in random locations shows that the network

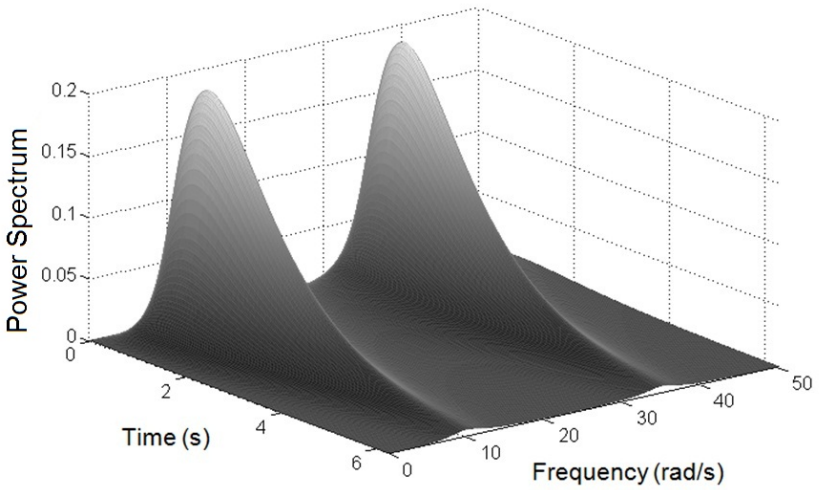


Figure 3.14: Non-stationary target spectrum

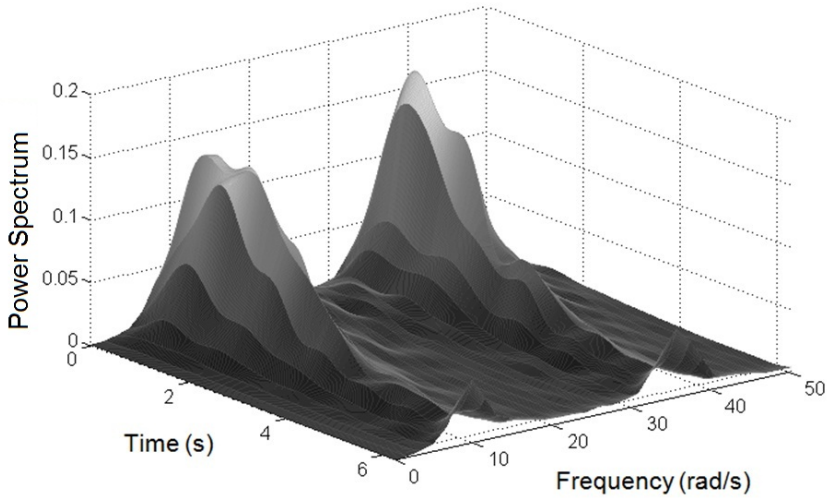


Figure 3.15: GHWT estimated spectrum with no missing data using 25 averaged time-histories

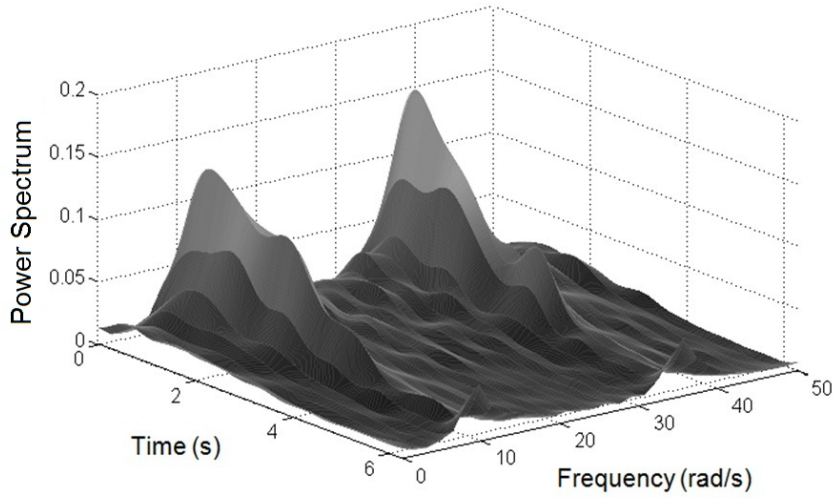


Figure 3.16: GHWT estimated spectrum with 50% missing data at random locations for 25 averaged time-histories using ANN

has captured the process to a large extent. The network is clearly producing a non-stationary processes fitting with the shape of the target spectrum over time. The total spectral power is smaller when compared with Figure 3.15, and both frequency peaks are lower, while higher levels of noise can be seen across all frequencies. However, without the ANN simulated time histories, instead applying zeros (Figure 3.17), the peaks are shown at much lower power, the shape of the spectrum over time is not as clearly identified and the noise, although not seemingly larger, has more prominent effect.

For the fixed missing intervals (of which there are two), the ANN has once again given well-shaped power spectrum (Figure 3.19), similar to Figure 3.15. This demonstrates that the ANN process simulation is relatively unaffected by arrangement of missing data. The fixed interval missing data test also highlighted an interesting flaw in the zero-filled time-history approach. Notice that in Figure 3.20, there are two large initial peaks at the beginning of the process which are cut off after around 1 second, re-appearing towards the end, before they are again cut off. This is because the missing data were consistently located at the same intervals in time; the low power bands correspond to these intervals.

3.5.3 Non-stationary non-separable example

Note that all previous non-stationary examples have been based around a separable process, making the assumption that the frequency content remains constant with time (only the intensity varies). In the case of a real earthquake, however, the frequency content changes with time as well. Next, consider the non-separable EPS given by Eq.2.69 (Figure 3.21). Reconstructions for all available data, ANN and zero gap filling for 50% missing data in random locations for 25 time-histories are shown in Figures 3.22, 3.23 and 3.24 respectively. As the non-separable spectral shape is more difficult to identify when re-constructed with harmonic wavelets, the outputs are shown in 2-D with a normalized colour map. The ANN reconstruction again maintains a strong spectral power, roughly 20% lower than the target, and clearly reproduces the decoupling of time and frequency in the process. As with Figure 3.17 for the separable process, filling with zeros is able to reproduce the spectral shape to some extent, but with much lower power and higher noise than the ANN solution.

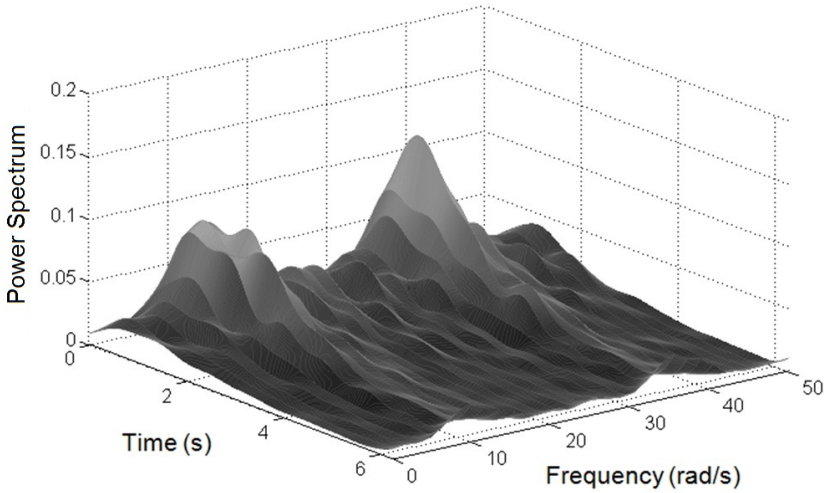


Figure 3.17: GHWT spectrum with 50% missing data at random locations for 25 averaged time-histories using zero-filled gaps

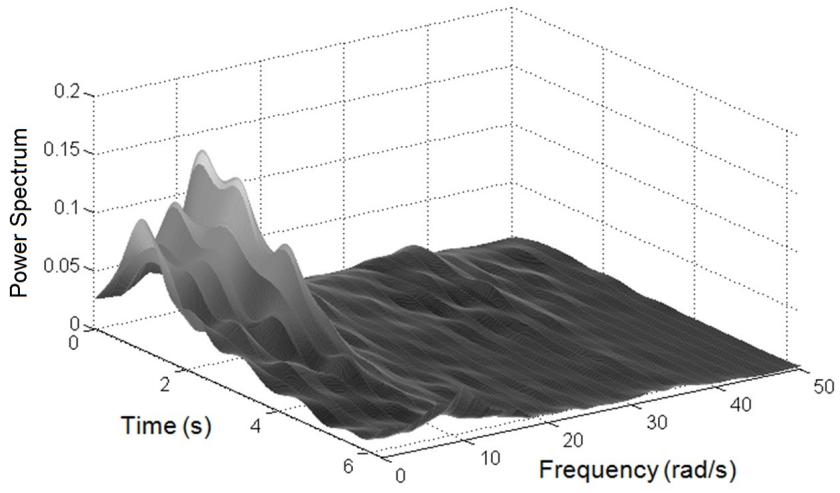


Figure 3.18: GHWT estimated spectrum with 50% missing data at random locations for 25 averaged time-histories using linear interpolation

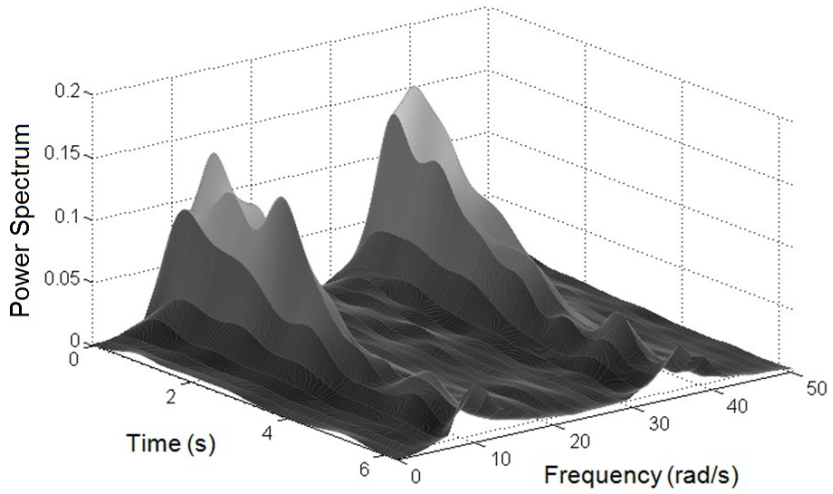


Figure 3.19: GHWT estimated spectrum with 50% missing data at two fixed-interval locations for 25 averaged time-histories using ANN

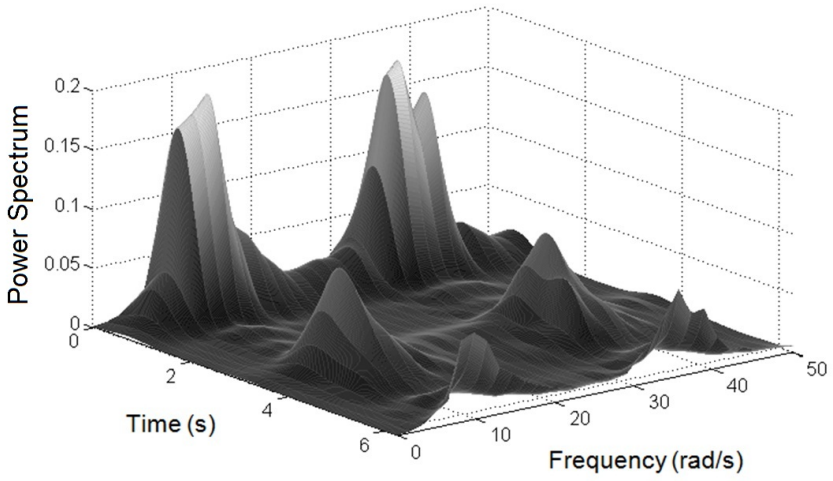


Figure 3.20: GHWT estimated spectrum with 50% missing data at two fixed-interval locations for 25 averaged time-histories using zero-filled gaps

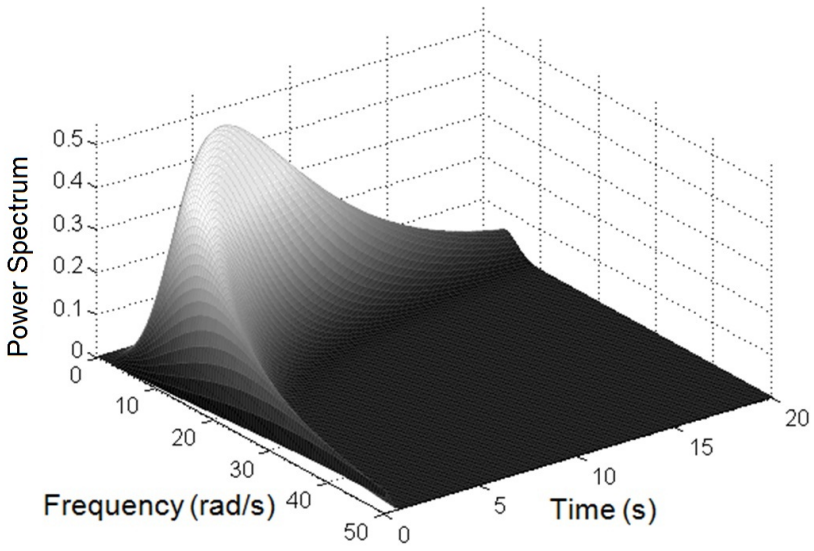


Figure 3.21: Non-separable evolutionary power spectrum

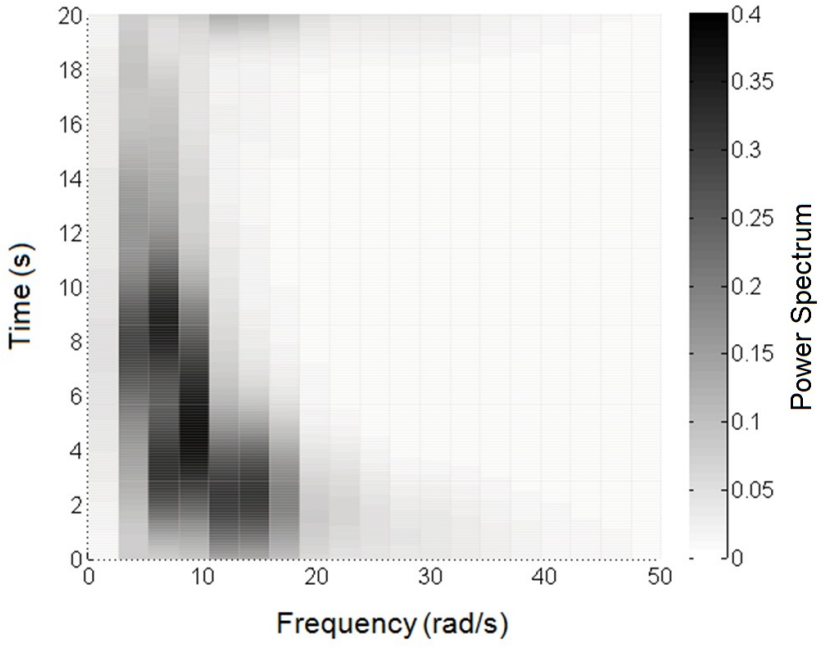


Figure 3.22: GHWT estimated non-separable spectrum with no missing data for 25 averaged time-histories

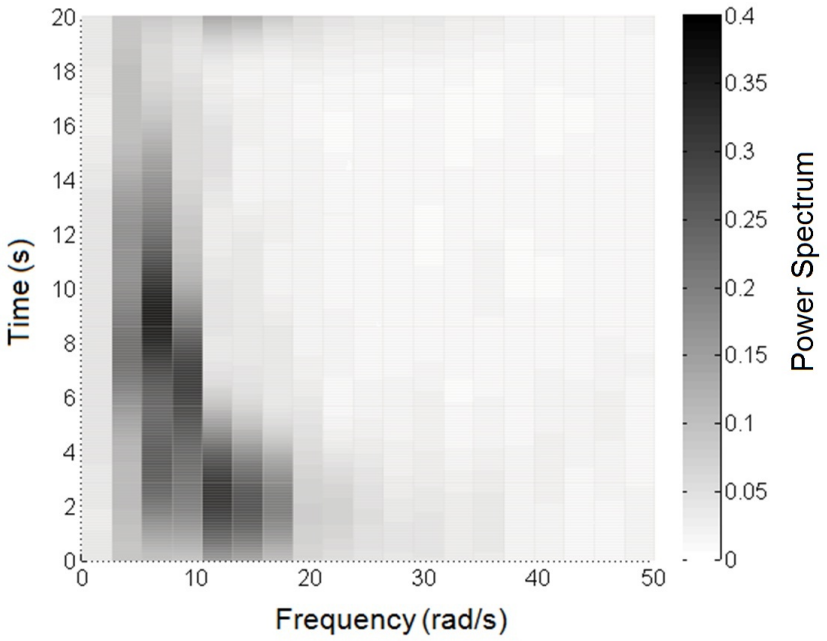


Figure 3.23: GHWT non-separable spectrum with 50% missing data at random locations for 25 averaged time-histories using ANN

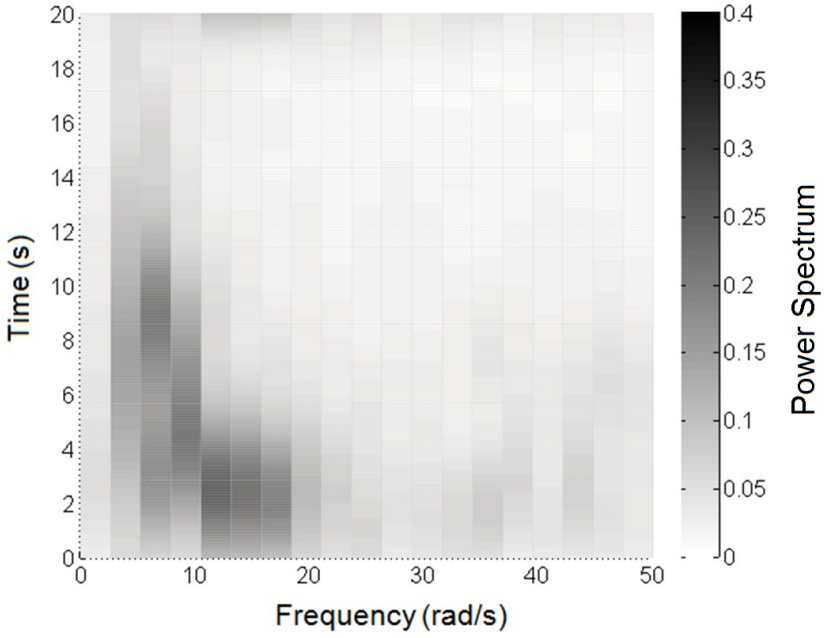


Figure 3.24: GHWT non-separable spectrum with 50% missing data at random locations for 25 averaged time-histories using zero-filled gaps

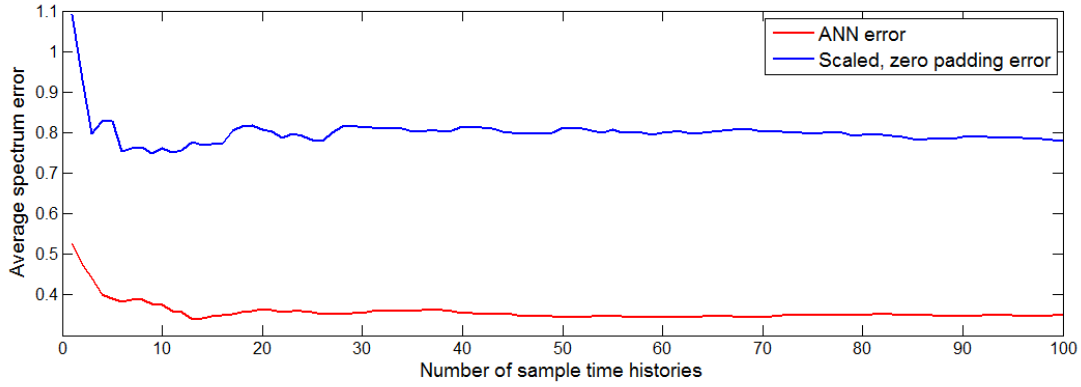


Figure 3.25: Average difference between estimated non-separable power spectrum from 100 time histories compatible with Eq.2.69 and ANN and zero padded reconstructions. Samples suffer from 50% missing data in uniformly distributed random locations.

3.5.4 Results overview

The results highlight the consistency in performance of the ANN based reconstruction across a range of spectra. It has shown to produce reliable estimates in the non-stationary case for spectra with multiple peaks, and of the non-separable variety with 50% missing data. Further, the technique remains robust when faced with varying arrangements of missing data. As previously stated, 25 samples are used in numerical examples as to show that the technique is applicable with realistic data set sizes. Further, it is shown in Figure 3.25 that significantly more samples has little effect on the average error in the estimated spectrum.

3.6 Chapter Summary

The novel aspects of the research herein described in Chapter 3 have also been communicated through international conferences / publications [57, 58, 59]. The work has been well received by the peer-review community, winning the "Best Student Paper Award" at the IEEE Symposium Series on Computational Intelligence (2013) from among approximately 300 papers. The ideas and outputs have been refined as a consequence of such input, leading to a collaborative paper with researchers at Federico Santa María Technical University, Chile, to test the procedure in the context of a structural reliability analysis [60].

Chapter 4

Compressive sensing based stochastic process power spectrum estimation subject to missing data

4.1 Introduction

In this chapter, another approach to solving the problem of missing data when estimating stochastic process power spectra is investigated. Compressive sensing (CS) is a signal reconstruction method that is commonly used in image processing and becoming a widely used tool in civil and mechanical engineering. CS, when applied to missing data problems requires more assumptions to be made concerning the nature of the process of interest than the previously discussed ANN method. However, it will be shown that in many problem cases, especially those related to environmental processes, these assumptions can be made with confidence. In the group of missing data problems for which CS is applicable, significant gains in spectrum estimation accuracy and computational efficiency can be achieved over the previously described ANN method in Chapter 3, and indeed other methods discussed in Chapter 2. Here, two approaches to applying CS techniques to missing data problems are introduced, both with numerical examples for stationary and non-stationary cases. The first approach (section 4.3) is a basic application of CS to reconstruct environmental process time histories suffering from missing data. The second approach (section 4.4.1) builds on the first, but makes the further assumption that an ensemble of process records are available (some or all of which have missing data). With this, an adaptive basis re-weighting procedure is proposed, which significantly enhances the accuracy of the estimated spectrum. Before applications to missing data are discussed, a brief introduction to CS and related theory is given.

4.2 Compressive sensing

The Shannon-Nyquist theorem states that a time-dependent signal with maximum frequency f can be completely determined when sampled at time intervals of $\frac{f}{2}$ or smaller. This maximum sampling frequency is commonly known as the Shannon-Nyquist rate. Compressive sensing is a recently developed signal processing technique that allows for signal reconstruction even if the maximum frequency f present in the signal is greater than half the signal's sampling rate [61]. Note that the idea shares many features with existing "lossy" compression algorithms (e.g. JPEG image compression) that take advantage of a signal's relative sparsity in some basis or frame [62].

When a data set is captured, it is often convenient to expand it into a new basis. In the case of "lossy" compression techniques, bases or frames (a basis with redundancy) are chosen, so that the vast majority of coefficients of the transformed signal will be close or equal to zero. If these coefficients are simply removed, the amount of space required to store the signal is reduced significantly (possibly by several orders of magnitude).

Next, when the signal is finally reconstructed back into its original form, for instance in the case of digital images, music or videos, it is often indiscernible from the true signal. Compressive sensing explores the possibility of recording data directly in its compressed state, allowing not only the space-saving advantages of compressed data, but also saving on recording time, complexity and compression processing [61].

Applications of CS techniques are numerous and has gained widespread interest, particularly in the field of image processing. Significant efficiency gains can be made when using medical scanning equipment such as in Magnetic Resonance Imaging (MRI) by making use of CS techniques [63]. Recently, applications of CS have shown to be effective in the operation of wireless sensor networks [64]. In the field of Civil Engineering, wireless sensor networks play an important role in structural health monitoring where CS techniques are seeing increased attention e.g., [65, 66, 75, 76]. This is because a reduction processing requirements at each sensor and bandwidth for data transfer as a result of CS lead to lower power requirements and increased efficiency whilst utilizing relatively simple, inexpensive sensors e.g., [67].

4.2.1 Signal sparsity

For robust compressive sensing there are several important properties to be considered, and one restriction is that of sparsity. The signal being sampled must be sparse in some known basis, i.e. it must be possible to represent the full signal with far fewer coefficients than the number determined by the Shannon-Nyquist rate. A discrete time signal, x may be viewed as an N *by* 1 column vector. Given an orthogonal N *by* N basis matrix A in which the columns A_i are the basis functions, x may be represented in terms of this basis via a set of N *by* 1 basis coefficients y, i.e.,

$$x = \sum_{i=1}^{N} A_i y_i, \tag{4.1}$$

Or more compactly, as in section 2.25,

$$x = Ay. \tag{4.2}$$

The vector x is said to be K-sparse in the basis A if y has K non-zero entries and $K < N$, i.e.,

$$x = \sum_{i=1}^{K} A_{n_i} y_{n_i}, \tag{4.3}$$

where n_i are the integer locations of the K non-zero entries in y. Hence y is an N *by* 1 column vector with only K non-zero elements. Therefore,

$$|y|_{L0} = K, \tag{4.4}$$

where $|.|_{Lp}$ denotes the Lp norm defined as

$$|y|_{Lp} = \left(\sum_i y_i^p \right)^{\frac{1}{p}}. \tag{4.5}$$

The L0 norm used in Eq.4.4 is defined as the limit of the Lp norm as $p \to 0$. In general the L0 norm is the total number of non zero elements in a vector,

$$|y|_{L0} = \sum_i \begin{cases} 1 & y_i \neq 0 \\ 0 & otherwise \end{cases} \tag{4.6}$$

It is important to note that for real signals, it is highly unlikely that they are exactly sparse in any orthogonal basis. Even a minimal amount of random noise on top of an otherwise K-sparse signal will produce non zero coefficients for all N. However, a large number of coefficients may be very small and in this case the signal is considered to be compressible.

4.2.2 Incoherence property

Another requirement for robust compressive sensing is that the sampling domain and the relatively sparse transformation domain must have high incoherence. This implies that a sparse signal in the transform domain must have a non-sparse representation in the sampling domain (i.e. a single Fourier coefficient in the transform domain would form a harmonic signal in the sampling domain spanning the entire sample length). Specifically, the coherence between two matrices A and Φ can be measured as the maximum absolute value of correlation between their elements [68],

$$\mu(A, \Phi) = \sqrt{N} \max_{k \geq 1, j \geq N} \langle A_k | \Phi_j \rangle \,, \tag{4.7}$$

where A_k and Φ_j are the rows of A and columns of Φ respectively. (For compression from the time domain to the Fourier or wavelet domains, a time domain sampling matrix would take the form of the identity matrix). The coherence lies within the range,

$$1 \leq \mu \leq \sqrt{N}. \tag{4.8}$$

Hence a coherence of 1 represents maximum incoherence and a coherence of $\sqrt{N}$ represents minimum incoherence. Further, if the number of measurements, M (rows in A) selected uniformly at random satisfies Eq.4.9, then the sparsest solution is said to be exact with high probability [68].

$$M \geq CK\mu(A, \Phi) \log N, \tag{4.9}$$

where C is a constant. Hence, the greater the incoherence between A and Φ, the fewer measurements required for signal reconstruction.

As an example, Figure 4.1 (dotted line) shows the following simple discrete time waveform sampled 256 times,

$$x\,(t) = 2\sin\,(12t + 1) + \sin\,(35t + 2) + 1.5\sin\,(120t + 3)\,, \tag{4.10}$$

where $0 \leq t < 2\pi$. The function of Eq.4.10 can be represented in the frequency domain by 6 peaks, 2 for each of the real and imaginary components of the harmonics at 12, 35 and 120 rad/s. The absolute amplitudes for each harmonic are shown in Figure 4.3 (resulting in 3 peaks on the plot). In the frequency domain, Eq.4.10 is clearly sparse, as the majority of the data is equal to zero. Assuming now that Eq.4.10 was a real signal being captured on a digital recorder, given the knowledge that the signal is sparse in the frequency domain, it is no longer necessary to capture 100% of the data in the time domain at the Shannon-Nyquist rate. Figure 4.2 shows the same signal as in Figure 4.1 sampled only 32 times ($1/8^{th}$ the full signal) at uniformly distributed random points. White noise following a normal distribution with zero mean and standard deviation of 0.6 is added to the signal to simulate measurement error. By applying CS, with only this limited amount of data, the sparse solution in the frequency domain is identified (Figure 4.3), and the signal may be reconstructed in the time domain as shown in Figure 4.4.

Despite the relatively small number of samples and added noise, CS has perfectly identified the positions of the original basis coefficients with a good approximation of their magnitude, and thus, the reconstructed signal is very similar to the original.

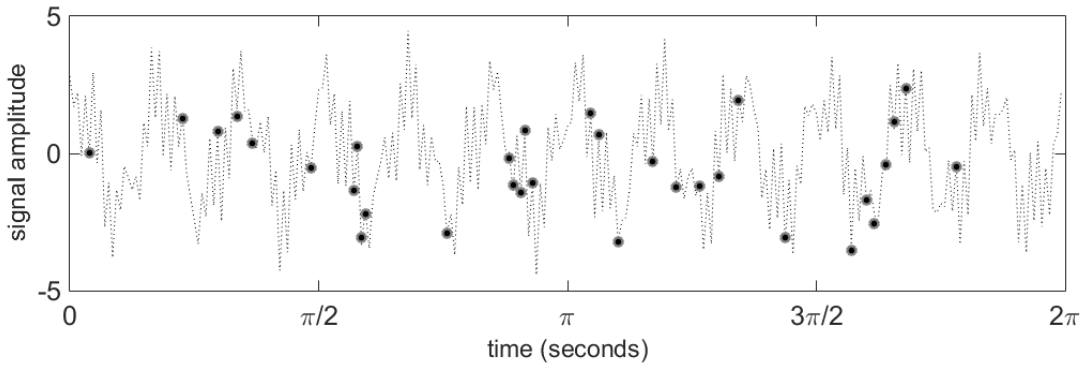


Figure 4.1: Graphical output of Eq.4.10 with randomly selected points to sample

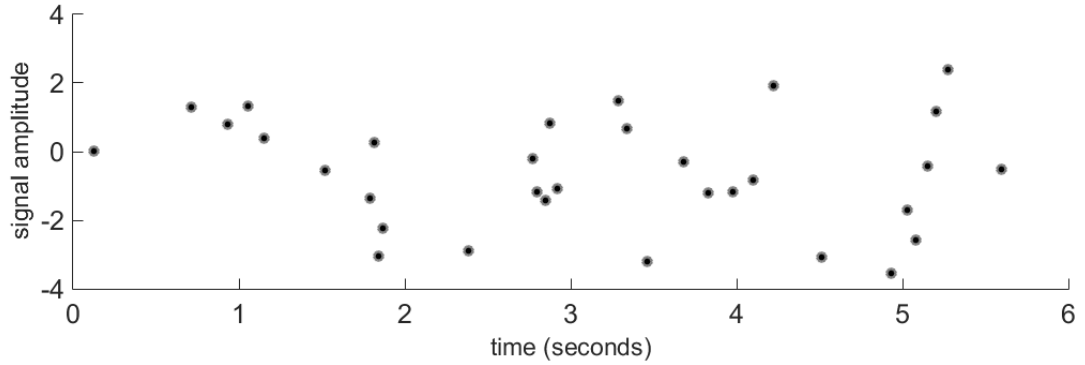


Figure 4.2: Randomly sampled points without original signal

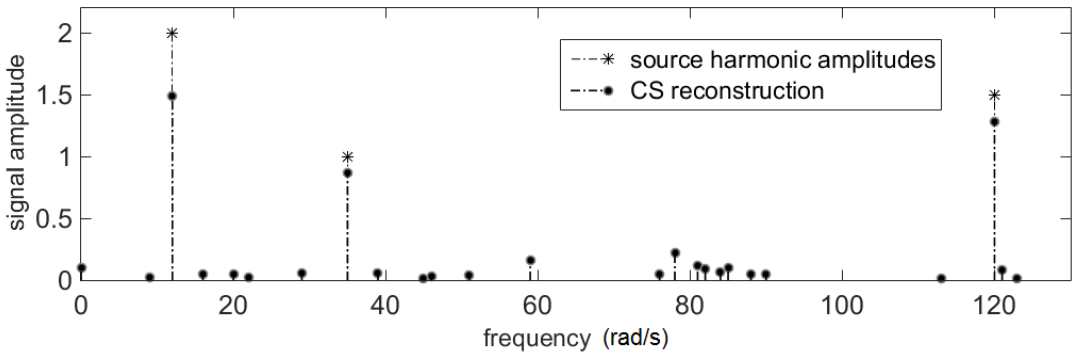


Figure 4.3: Eq.4.10 represented in the frequency domain as a sparse signal with CS estimation of frequency domain coefficients from Figure 4.1

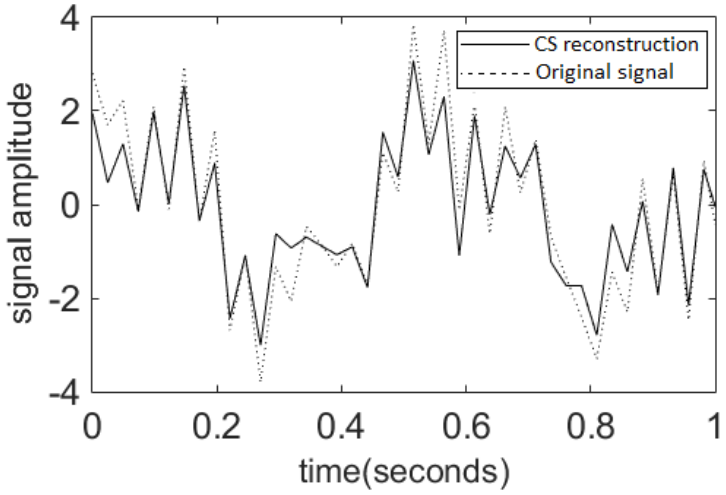


Figure 4.4: Comparison of first second of Figure 4.1 against the CS estimation in the time domain from Figure 4.2

4.2.3 Restricted isometry property

Another important condition for reliable CS, in addition to signal sparsity and incoherence between bases, is the Restricted Isometry Property (RIP). A 'fat' sampling matrix (M *by* N where $N > M$), A, satisfies the RIP with sparsity K if there exists a constant, δ_K such that,

$$(1 - \delta_K) \left|\tilde{y}\right|_{L2}^2 \leq \left|A\tilde{y}\right|_{L2}^2 \leq (1 + \delta_K) \left|\tilde{y}\right|_{L2}^2 , \tag{4.11}$$

for every vector $\tilde{y}$ with at least K non-zero entries [69]; L2 denotes Euclidian norm, i.e., $p = 2$ in Eq.4.5. Equivalently, this means that if a signal has sparsity K (i.e. it can be represented by K coefficients in a chosen basis, A), any matrix comprised of K randomly selected columns of A should have full rank and be nearly orthonormal. Unfortunately, checking the RIP for any given matrix is NP-hard [70]. *A problem is NP-hard if an algorithm for solving it can be translated into one for solving any NP-problem (nondeterministic polynomial time). NP-hard therefore means "at least as hard as any NP-problem,"*. However, there are several matrices for which the RIP is known to hold with high probability. For instance, for a Gaussian random matrix, the RIP holds with high probability if

$$m \geq CK \log\left(\frac{N}{M}\right), \tag{4.12}$$

where m is the height of the measurement matrix (i.e. the number of measurements), N is the width, and C is a constant which tends towards 1 as N tends to infinity (e.g. [71]).

4.2.4 Sparse solution via L1 minimization

If it is known that a signal is sparse in a particular basis, then the aim of CS is to attempt to find the sparsest representation in that basis for the given data; this may be achieved by L1 norm minimization. Given a sample record y, of length $N_0 - N_m$, where N_0 is the original sample length and N_m is the number of missing data, assuming the locations of the missing data are known, a corresponding ($N_0 - N_m$ by N_0) sensing matrix, A can be drawn,

$$x = Ay \tag{4.13}$$

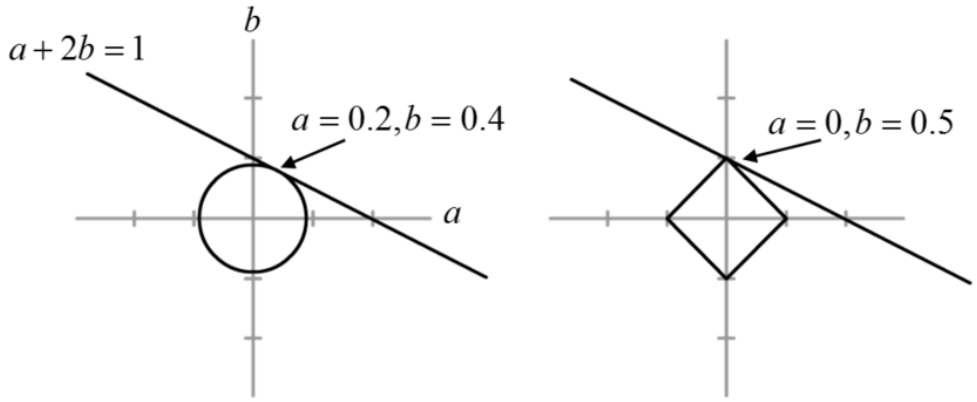


Figure 4.5: Minimum L2 and L1 solutions to the equation, $a + 2b = 1$

where y is the measurement vector assumed to be sparse. Eq.4.13 represents an underdetermined system with infinite solutions. This problem may be solved easily under the constraint that y must be minimized in the least-squares sense, i.e.

$$\min |y|_{L2} = A^T \left(AA^T\right)^{-1} x. \tag{4.14}$$

Considering a Fourier basis, this solution is similar to replacing the missing data with zeros and applying the Fourier transform in the standard way. Therefore, in the majority of cases, applying the least-squares solution (Eq.4.14) does not lead to a sparse solution. The sparsest solution of Eq.4.13 occurs when the $L0$ norm is minimized, often referred to as a pseudo-norm [72], and defined as

$$|y|_{L0} = \begin{cases} 1, & y > 0 \; or \; y < 0 \\ 2, & y = 0 \end{cases} \tag{4.15}$$

This optimization problem is non-convex with no known exact solution [73]. However, a viable alternative exists in minimizing the L1 norm instead i.e., $p = 1$ in Eq.4.5. L1 norm minimization promotes sparsity and will often yield the same result as L0 norm minimization in many cases [72]. Further, the problem becomes convex, and may be set in a convenient linear programming form, i.e.

$$\min |y|_{L1} \;\; subject \; to \; x = Ay \tag{4.16}$$

Eq.4.16 describes a basis pursuit optimization problem and can be easily solved via a gradient-based optimization method, e.g. [68]. Figure 4.5 shows how L1 minimization gives sparse solutions by comparing both L1 and L2 (least-squares) norm minimization for the simple 2-dimensional problem, $a+2b = 1$, (for which there are infinite solutions). Note, as the L2 ball is stretched (Figure 4.5, left), unless the equation for y is parallel to one of the axes, the solution will incorporate components of both a and b. However, unless the equation for y is parallel to the edge of the L1 ball (Figure 4.5, right), as it is stretched, the minimum solution will lie on one of the two axes. Unfortunately, real signals are rarely ever truly sparse; even low levels of noise will produce small coefficients across most bases. With a small modification to Eq.4.16 to account for noise, basis pursuit is still able to recover a good approximation to the original signal; this element of robustness was demonstrated in Figure 4.4. For a noisy signal,

$$x = Ay + z, \tag{4.17}$$

where z is some noise vector. Given a tolerance, e, relative to the variance of the noise, Eq.4.16 may be re-cast in the form,

$$\min |y|_{L1} \;\; subject \; to \;\; |Ay - x|_{L2} \leq e. \tag{4.18}$$

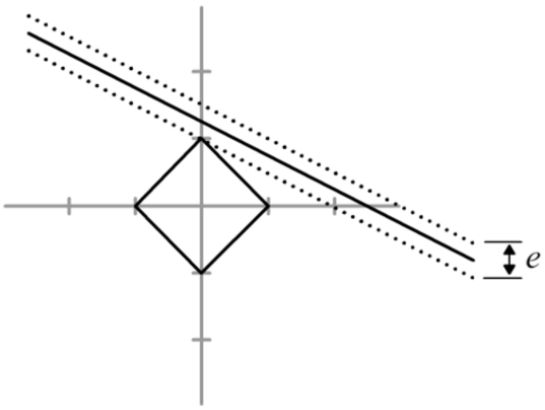


Figure 4.6: Minimum L1 solution to the equation, $a + 2b = 1$ with tolerance, e for noise vector, z

This modification has the effect of applying intervals to the solutions (Figure 4.6), further promoting sparsity. However, as the tolerance increases, the resulting basis coefficients tend to reduce not only in number, but also in magnitude. For the cases where either the signal is not sparse enough or the missing data are too extensive for L1 minimization to exactly reconstruct the original signal, it is important to note that there may still be significant advantages over a minimum L2 solution. In spectral estimation, minimizing the L2 norm (similar to zero-padding) is likely to spread the solution over many frequencies; this is because individually, large coefficients are heavily penalized. Minimizing the L1 norm however is far more likely to yield larger individual coefficients, having the effect of producing sharp, well-defined peaks at the key frequencies. The difference between L1 and L2 minimization for a real sparse signal in the time domain is clearly shown in Figure 4.7. Figure 4.7a shows radiosonde wind speed data recorded by a weather balloon launched from Halley Antarctic Research Station in January 2014 [74]. This data has a relatively sparse representation in the harmonic wavelet domain, and is therefore ideal for CS. With 50% of the data removed at random locations, Figures 4.7c and 4.7d show L1 and L2 wavelet reconstructions respectively.

4.3 Compressive sensing with missing data

CS is mostly applied in situations where some saving in data capture time or data size is useful. For example, if a series of sensors capture data for real-time structural health monitoring, data may need to be compressed to adhere to bandwidth limitations, after which most of the captured data is lost. Instead, the sensors could be designed to only capture a fraction of the data, reducing manufacturing cost. By utilizing CS with the compression basis (in which the signal has a sparse representation), data series with far higher resolution than those originally captured could be reconstructed (e.g. [75, 76, 64, 65, 66]). Not only would the sensors not need to capture as much data, but also the stored data would have a small file size, negating the requirement for compression processing at the sensor.

It is clear at this point that CS techniques have a wide range of potential applications in the area of environmental load modelling when considering many associated process records, such as those produced by earthquakes, sea waves, winds, and tidal patterns (including structural responses to these effects), can be characterized by a relatively small number of dominant frequencies in the frequency domain. Further, it is shown that under an appropriate basis selection, CS reconstruction may perform equally well

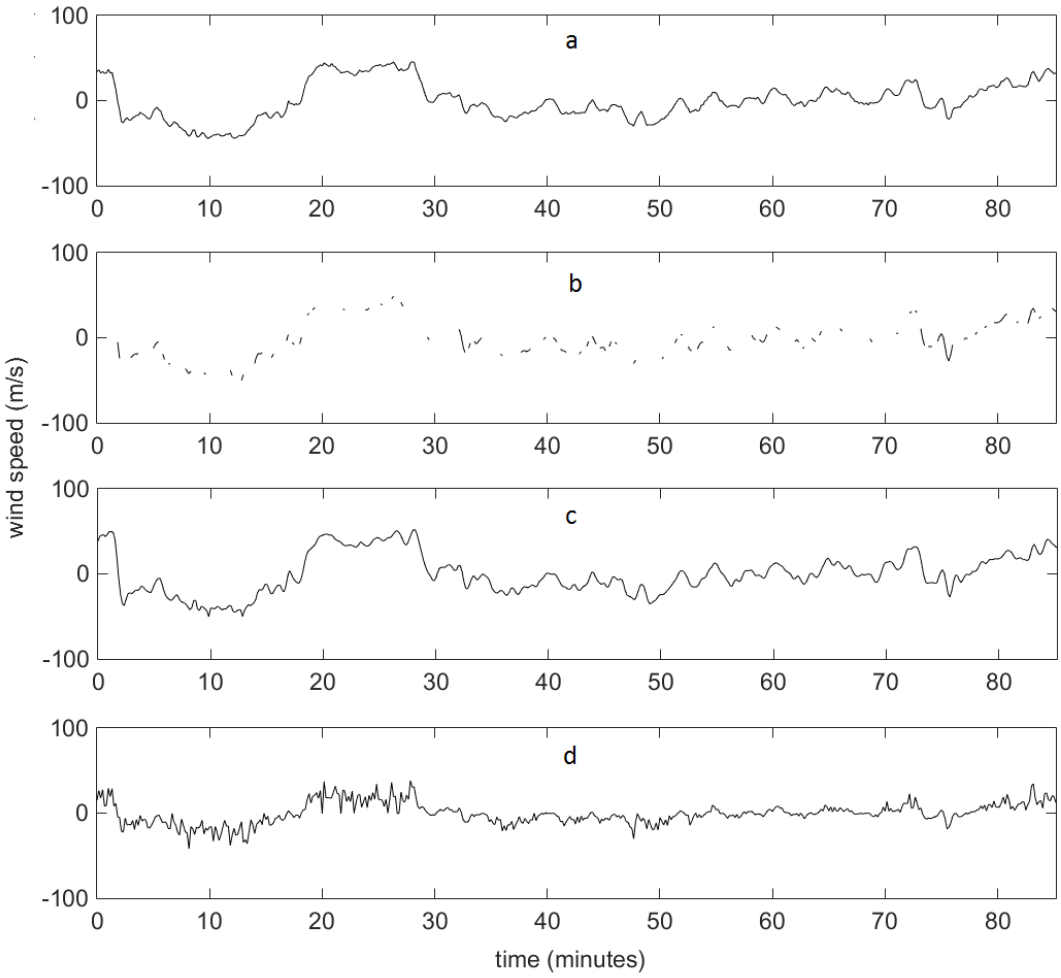


Figure 4.7: (a) Full windspeed record. (b) Full record with missing data. (c,d) Records with 50% missing data in random locations, down-sampled and reconstructed via L1 and L2 minimization of harmonic wavelets respectively. Data provided by [74]

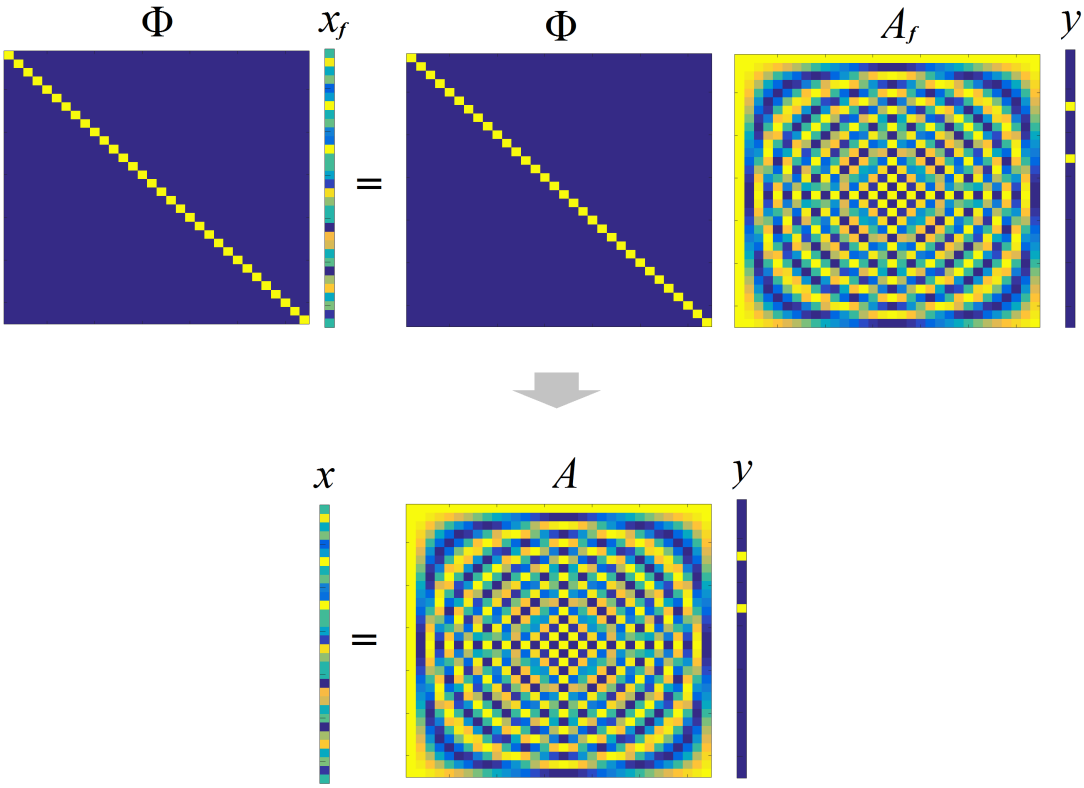


Figure 4.8: Signal acquisition of Eq.4.19 at every point in N note that x_f is sparse in the Fourier matrix A_f, hence y has only two entries (one representing $\cos(6t)$ and one representing $\cos(12t)$)

for both stationary, non-stationary and even for highly non-stationary processes under varying arrangements of missing data.

Nevertheless, applying CS theory to the problem of missing data differs primarily in one respect to standard CS applications; i.e., missing data are not commonly intentional. Unfortunately this removes control over one important step of CS: the arrangement of the sampling matrix. Consider a situation in which a harmonic signal (Eq.4.19) needs to be captured in the time domain and compressed.

$$x_f(t) = \cos(6t) + \cos(12t) \tag{4.19}$$

For compression, the DFT would be highly applicable in this case, producing only two coefficients. To apply compressive sensing to this problem, one must choose sample points in the time domain to store, assume sparsity in the Fourier domain, and reconstruct the signal afterwards via L1 minimization. Figure 4.8 depicts the sensing process if the entire signal of length $N = 32$ were captured. Here, the sensing matrix Φ is the square identity matrix, representing a series of Dirac pulses in the time domain, x_f is the full signal shown in Eq.4.19 and A_f is a real Fourier matrix (composed of cosines). To take advantage of the signal sparsity in A_f, the sensing matrix Φ may be arranged to capture only part of the signal x_f. A Fourier ensemble based on M uniformly random selected rows of the full Fourier matrix, the RIP is satisfied for a K sparse signal [77, 73] provided that

$$K \leq C \frac{M}{(\log(N))^6}. \tag{4.20}$$

The CS procedure is depicted in Figure 4.9 for the same signal Eq.4.19 with a compression ratio of 50% i.e., $M = \frac{N}{2}$. The sensing matrix is generating via random permutation of the identity matrix, then discarding 50% of the rows. As previously stated,

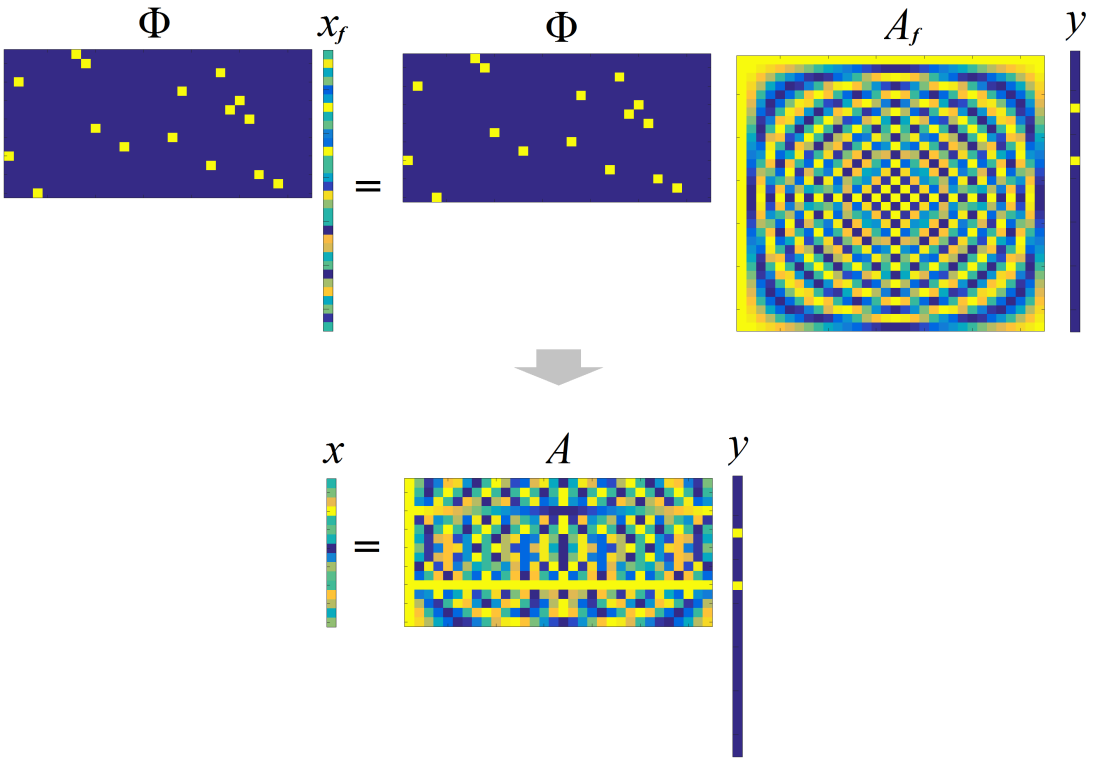


Figure 4.9: Signal acquisition of Eq.4.19 at $\frac{N}{2}$ uniformly distributed random locations over x_f for CS with a Fourier basis

CS relies on the choice of an appropriate sampling matrix. Uniform random Fourier matrices obey the RIP with high probability when data are sparse [77, 73]; similarly, random harmonic wavelet matrices may reconstruct sparse non-stationary signals exactly (however, there is lower incoherence between the wavelet and time domains which decreases with frequency resolution). Unfortunately, the missing data may not be uniformly distributed over the record; when using Fourier or harmonic wavelet matrices, regular or large gaps of missing data leads to 'lower' orthogonality between random columns of the sampling matrix. If we were to arrange a sampling matrix to simulate missing data in a recorded process record, it would appear similarly to Φ in Figure 4.9 with ordered rows as in Figure 4.10. Depending on the arrangement of the missing data, the result may be that greater numbers of measurements are required for reliable reconstruction. Despite these problems, CS reconstruction based on the assumption of sparsity is often still advantageous over more common least-squares/zero-padding approaches. This is because, despite massive data loss (in some cases $> 90\%$), CS can still identify sharp spectral peaks at dominant frequencies.

4.3.1 Basis matrix construction

For stationary and non-stationary processes, the basis matrices must be constructed after the process has been recorded and the missing data identified. Fourier and harmonic wavelet bases are chosen to represent the spectral content of process records in the numerical examples section, and so these will now be considered. However, in practise a wide range of suitable bases could be considered depending on the nature of the problem; in any case the basis matrix construction would follow the same fundamental steps as highlighted in this section.

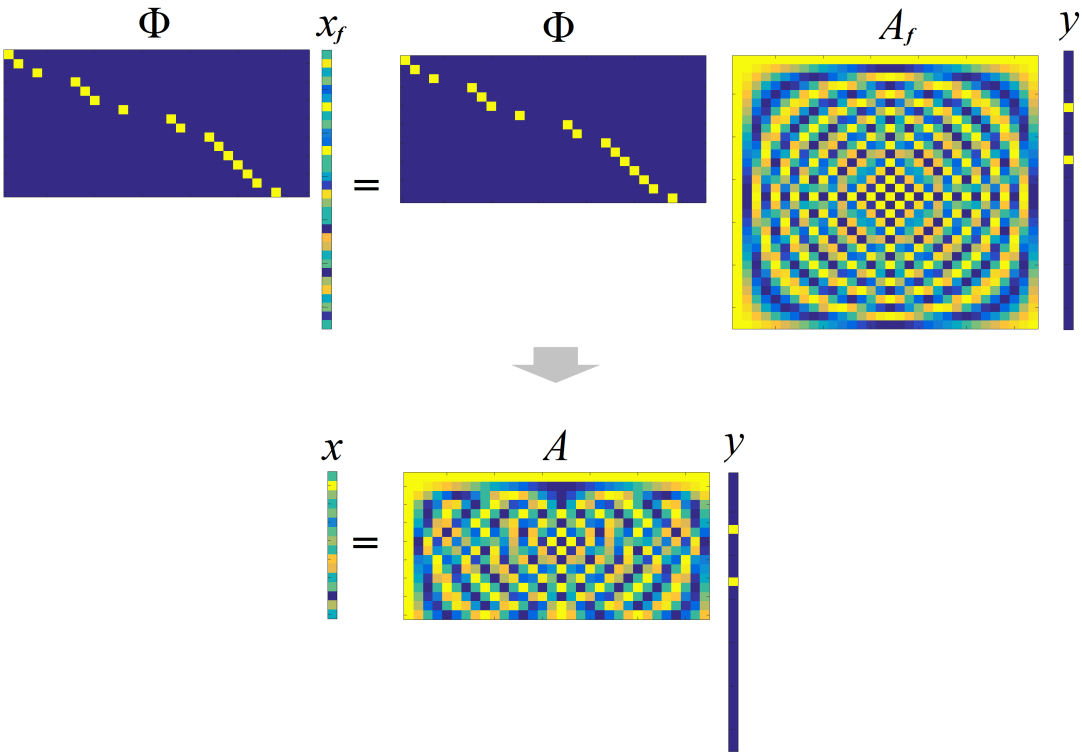


Figure 4.10: Eq.4.19 with $\frac{N}{2}$ uniformly distributed missing data over x_f set in CS framework with a Fourier basis

Stationary case

For stationary stochastic processes, represented by Eq.2.50, the power spectrum is estimated based on the mean square value of the Fourier transform over an ensemble of time-histories, e.g. [17]. As such, a 'partial' Fourier basis is required for the CS approach. The partial Fourier basis is formed first by generating a full, square (N_0 by N_0) Fourier basis via the Inverse Fast Fourier Transform (IFFT). N_m rows are then removed from the matrix corresponding to the positions of the missing data, as shown in Figure 4.11.

Non-stationary case

The non-stationary case is slightly more involved, as specific properties of the harmonic wavelets are needed. In particular, wavelet scales must first be defined; that is, a set of non-overlapping frequency intervals corresponding to $(n-m)$ in Eq.2.44. In most cases these are chosen to be equally spaced with an interval size that gives the desired trade-off between time and frequency resolutions. However, should finer frequency or time resolutions over specific frequency bands be required, the sampling matrix can be altered accordingly. As in the stationary case, the harmonic wavelet basis components may be generated efficiently via IFFT. However, a single harmonic wavelet must be shifted $(n-m)$ times in the time domain to form an orthogonal basis (i.e., the non-redundant harmonic wavelet transform). The process used to build this harmonic wavelet basis matrix is depicted in Figure 4.12. In a similar manner as in the stationary case, rows must be removed corresponding to the missing data, yielding a sampling matrix with more columns than rows (Figure 4.13).

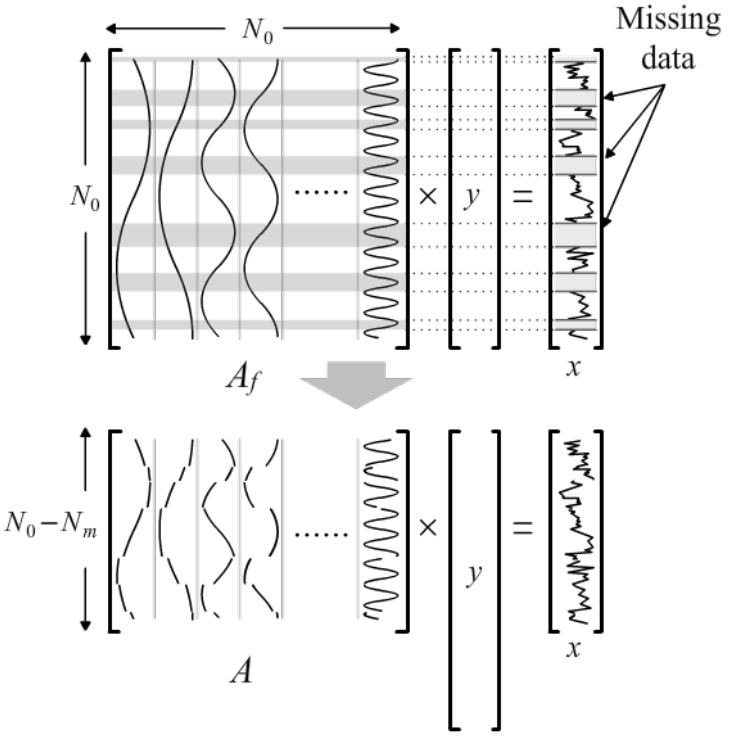


Figure 4.11: Fourier sampling matrix construction with missing data

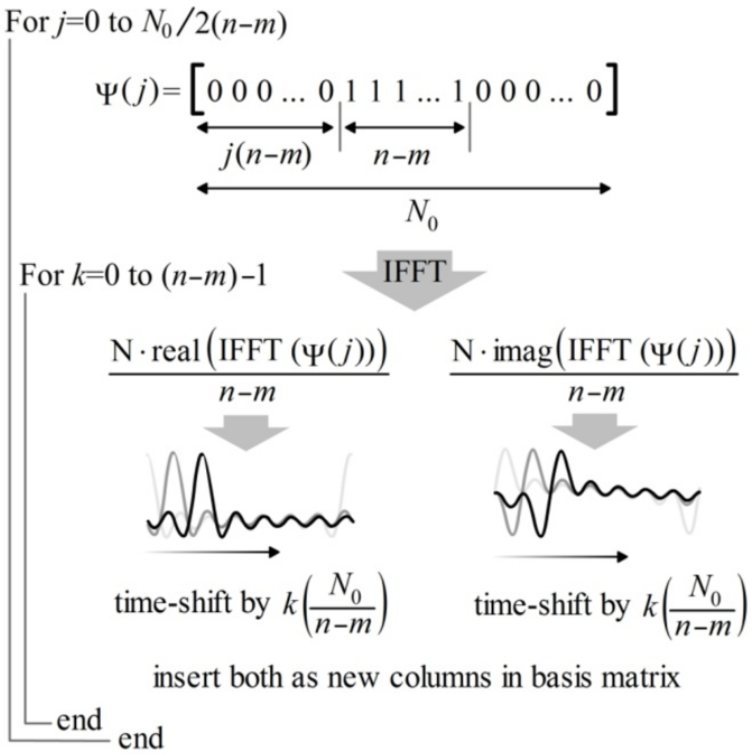


Figure 4.12: Non-redundant orthogonal Harmonic wavelet basis construction using IFFT and nested for-loops

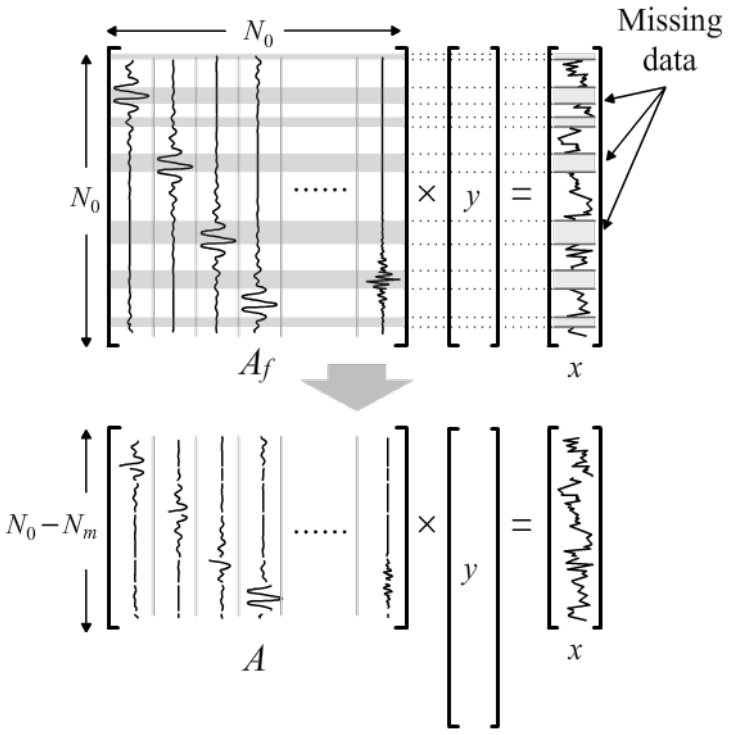


Figure 4.13: Harmonic wavelet sampling matrix construction with missing data

4.3.2 Uneven sampling

The previously described methods for forming stationary and non-stationary basis matrices assume that the original data has been regularly sampled. This means that the data always occur at multiples of the smallest sampling time increment. If the data has been irregularly sampled, the basis matrices can no longer be generated via the standard IFFT. However, this is not a major problem; it merely reduces the efficiency of the sampling matrix construction, as the basis functions must be evaluated for each sample point. This overhead is more apparent in the non-stationary case, as each instance of every wavelet must be individually calculated from Eq.2.45.

With the basis formed, the CS reconstruction may be solved via an appropriate minimization algorithm. Examples include linear programming basis pursuit [78, 73] and greedy algorithms [79].

4.4 An adaptive basis re-weighting procedure for ensemble process records

Now, CS techniques are applied in conjunction with an adaptive iterative algorithm for basis re-weighting. This method imposes similar basic restrictions on the nature of the data as standard CS reconstruction in the previous section. Primarily, the process is assumed to be sparse in the basis chosen to represent the power spectrum. However, a significant further requirement of this modified approach is that several process realizations must be available. This is because the method relies on the ability to apply CS to multiple process records iteratively, utilizing the cumulative information from all records for the purpose of seeking a sparse representation in the average sense over an ensemble. By introducing this iterative process to alter basis coefficients, significant gains in spectral estimation accuracy over standard CS can be achieved. When multiple

process records are not available, but the single time-history is large with respect to the total bandwidth of its frequency content, records may be down-sampled into several shorter records to enable the use of this method. For non-stationary processes, this down-sampling also requires that the maximum frequency of the important spectral content is far from the Nyquist frequency defined by the original sampling rate.

4.4.1 Re-weighting the basis matrix

It is shown in section 4.5.1 that CS with an appropriate basis alone can be applied to the problem of missing data when estimating power spectra, which can deliver significant improvements over least-squares and other more complex methods. However, when the target spectrum is estimated from a process record ensemble, there are further improvements to be gained. The improvements to the estimated spectrum are achieved by re-weighting columns of the basis matrix. The effect of re-weighting the basis can be demonstrated by examining a CS problem in which the L1 minimization does not produce the sparsest solution. Consider the following under-determined system of equations,

$$\begin{bmatrix}1\\1\end{bmatrix} = \begin{bmatrix}1 & 2 & 1\\-2 & 1 & 1\end{bmatrix}\begin{bmatrix}a\\b\\c\end{bmatrix}. \tag{4.21}$$

Indeed, the solution may lie anywhere along the intersection of the two planes $a+2b+c = 1$ and $-2a+b+c = 1$. Figure 4.14 and Figure 4.15 show the solutions to the problem via least-squares and L1 minimization respectively. Note that by expanding the L2-ball (Figure 4.14) from the origin, a minimum solution is reached at $\begin{bmatrix}-0.14 & 0.43 & 0.29\end{bmatrix}^T$, where all coefficients are non-zero. Expanding the L1 ball however (Figure 4.15), gives a minimum solution at $\begin{bmatrix}-0.2 & 0.6 & 0\end{bmatrix}^T$, having a single non-zero coefficient. L1 minimization has given a sparser solution in that only two of the basis vectors are required, however there exists a solution requiring only a single coefficient, $\begin{bmatrix}0 & 0 & 1\end{bmatrix}^T$. This ideal solution that would have been reached by L0 minimization was not realized because, although to a lesser extent than least-squares, L1 minimization still penalizes large coefficients more than small ones. The L0 minimization can be found in an L1 sense by doubling the magnitude of the c column in the basis matrix, making it less expensive in the minimization procedure. This can be done by multiplying by a diagonal basis re-weighting matrix, W, and solving the minimization problem,

$$\min |y|_{L1} \;\; subject\ to\ x = AWy;$$

$$W = \begin{bmatrix}1 & 0 & 0\\0 & 1 & 0\\0 & 0 & 2\end{bmatrix} \tag{4.22}$$

Equation (4.22) has the effect of altering the angle between the two planes in Figure 4.15 leading to a new line of intersection. Figure 4.16 shows that by changing the angle of this line, expanding the L1 ball again gives a minimum solution at $\begin{bmatrix}0 & 0 & 0.5\end{bmatrix}^T$. Pre-multiplying this solution by W, the L0 solution, $\begin{bmatrix}0 & 0 & 1\end{bmatrix}^T$ is reached.

4.4.2 Utilizing the ensemble

Although this shows that it is possible to improve the L1 minimization solution by suitably altering the magnitudes of the basis vectors, a decision must still be made on the coefficients of the re-weighting matrix, W. [80] suggests that sparsity can be increased

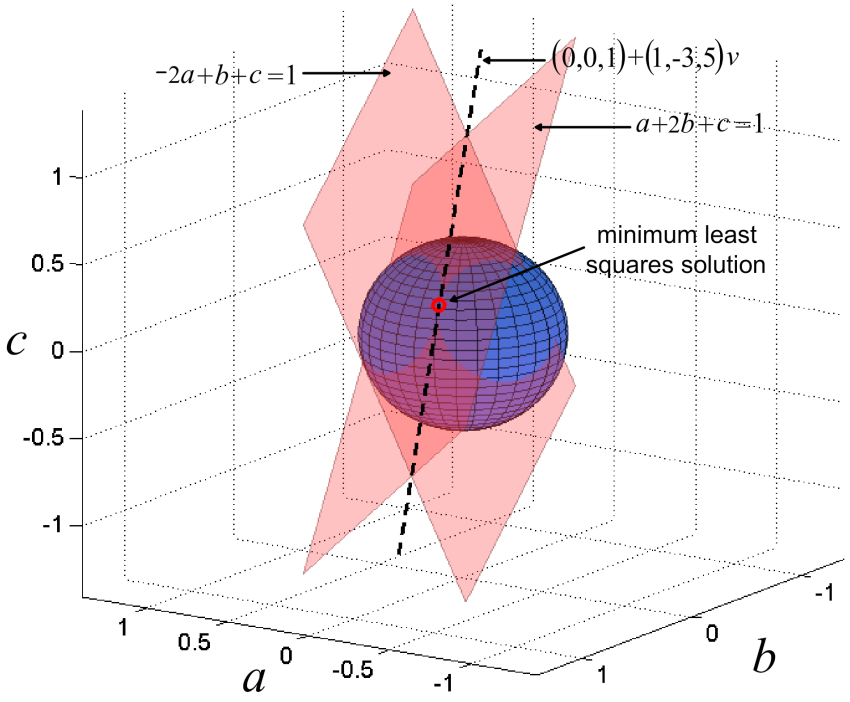


Figure 4.14: Least squares minimization solution of Eq.4.21

by iteratively re-weighting y based on its previous output from the optimization procedure. This only seems to work well under fairly strict conditions, usually when the first iteration of the L1 minimization is already close to the sparsest solution. In the case of Eq.4.21, in which the L1 minimization gives $\begin{bmatrix}-0.2 & 0.6 & 0\end{bmatrix}^T$, this procedure would not work; and the second iteration would further encourage a solution in a & b, penalizing c. When estimating the power spectrum from an ensemble of process records however, the objective may not be to find the sparsest spectral representation for each individual record, but instead to find the sparsest representation of the ensemble spectrum average.

Consider four systems of equations with the same dimensionality as Eq.4.21 (Figure 4.17), all have the same basis matrix, A. For each, L1 minimization has found the sparsest solution. For the first three, the solution lies only in a and c; for the fourth, the solution lies only in b. The bar chart in Figure 4.17 shows the sum of basis coefficients a, b and c for all solutions. This is akin to estimating the power spectrum from a process record ensemble, the only difference being that in the power spectrum, each coefficient likely comes from two or more basis coefficients (in the case of the Fourier power spectrum, one sine wave and one cosine wave). Note that the sums of the a & c coefficients are larger than b. If these coefficients are used to build the re-weighting matrix W, and the L1 minimization is repeated, the b vector will have a greater cost. The result is that all four solutions now use only components of a and c. Although the final system has a less sparse solution than it did originally, the non-zero expected basis coefficients across the ensemble have decreased from three to two. In higher dimensional problems, repeating this process (i.e. continuously re-using the basis coefficients to build the re-weighting matrix) will further modify the solution until some threshold limit is reached. In the case of power spectrum estimation, the procedure effectively calculates the mean ensemble spectrum shape, stores this shape within the re-weighting matrix, and then uses this information to influence individual power spectrum estimations.

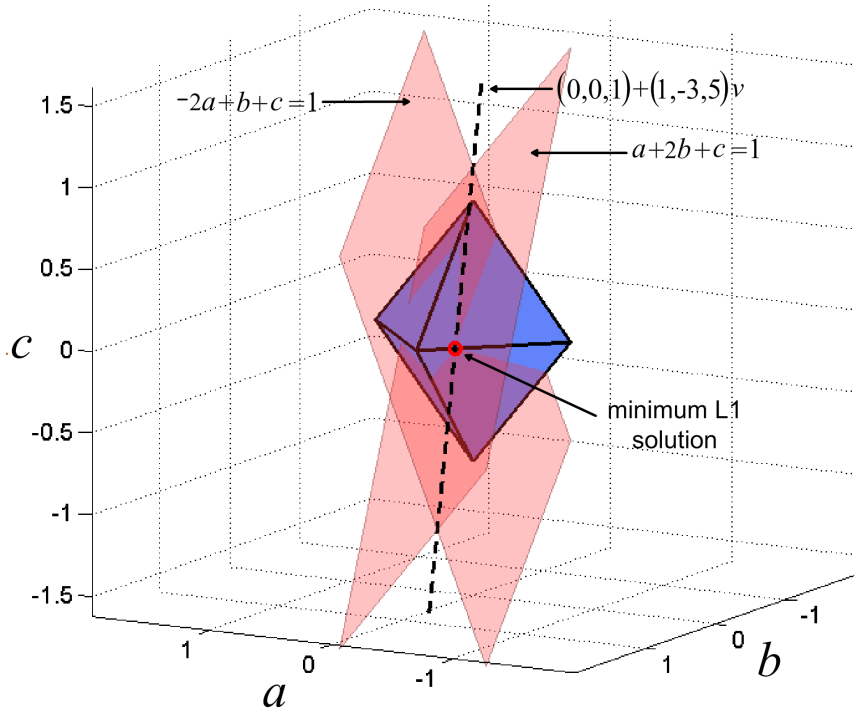


Figure 4.15: L1 minimization solution of Eq.4.21

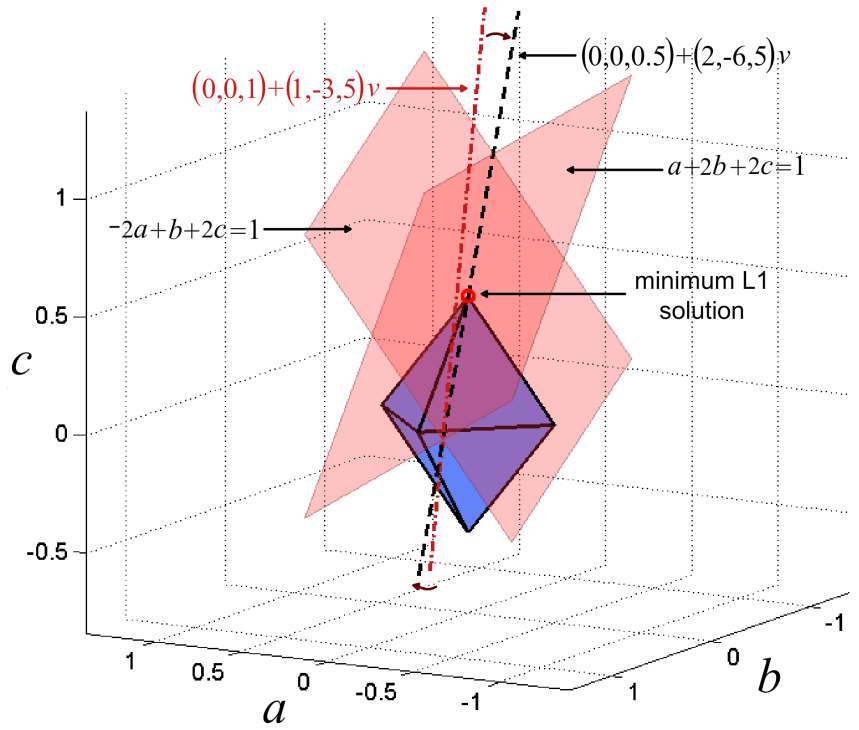


Figure 4.16: L1 minimization solution of Eq.4.21 after applying the re-weighting matrix in Eq.4.22

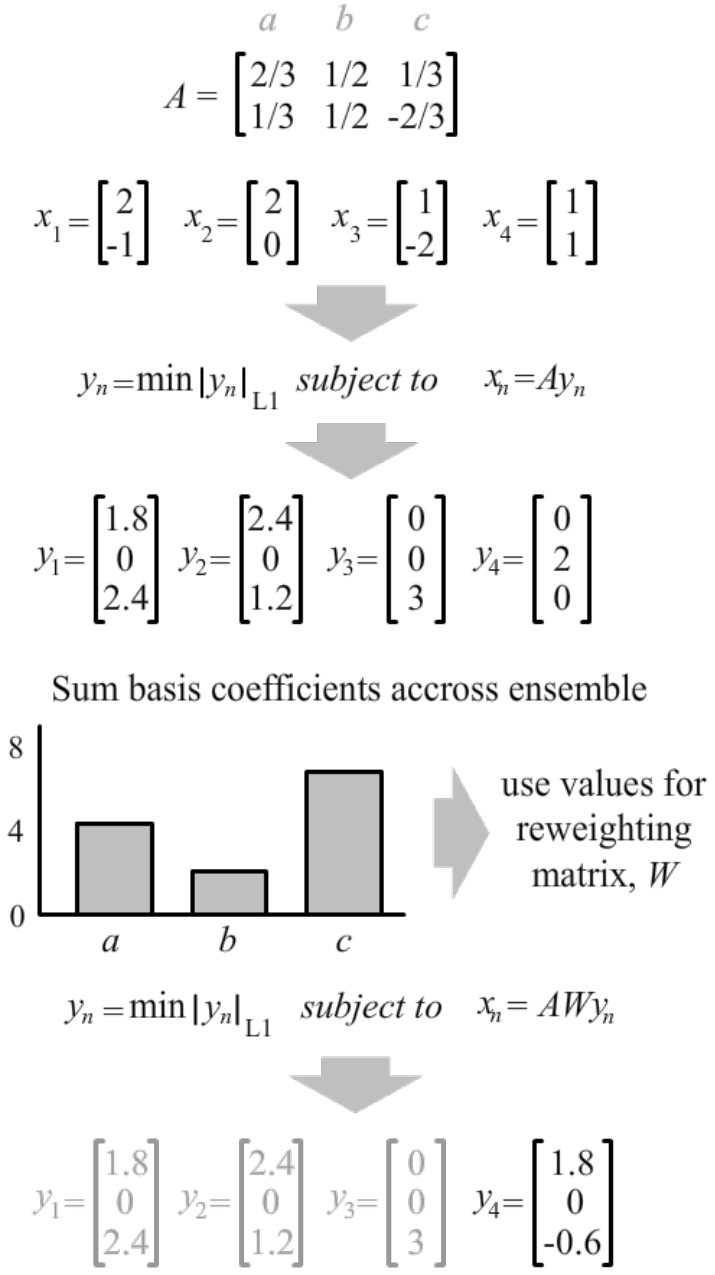


Figure 4.17: Example of basis re-weighting procedure promoting sparsity using only average basis coefficients to update the weights (notice y_1 through y_3 do not change as they were originally estimated in a and c)

Efficient least squares re-weighting

Two problems were identified with the aforementioned iterative L1 minimisation re-weighting scheme. Firstly, although each minimization problem is convex and may be solved via a number of efficient basis pursuit methods, for large process records and large ensembles, the processing time can be significant. Further, when dealing with only a small number of records in the ensemble with significant missing data, L1 minimization was found to occasionally produce large false peaks that could not be completely negated by the averaging procedure. To address both of these problems, least-squares minimization is considered for the re-weighting scheme (Figure 4.18). The numbers spanning the left column of Figure 4.18 indicate important steps that are elaborated on in detail below.

1. First, the sensing matrix is constructed as previously, based on the missing data locations and choice of basis functions (a Fourier matrix is shown in the figure).

2. Two temporary re-weighting matrices are required for the iterative procedure that follows. These are both initialized as square identity matrices.

3. The re-weighting matrix its generated iteratively. The termination criteria for the re-weighting iterations is based here on a minimum sum of the difference between previous re-weighting matrix and current one. Essentially, when the re-weighting matrix is only experiencing minor changes in each iteration, the loop is terminated. Other termination criteria could be used and this is only given as an example.

4. This first step in the while-loop sets the 'active' re-weighting matrix W equal to the last one generated in the loop W_2. W_2 is then set equal to a matrix of zeros, ready to be populated with new re-weighting coefficients.

5. During the least-squares re-weighting phase, the re-weighting matrix is based on a least-squares spectrum estimation. For multiple records, the least squares spectrum estimation for the entire process involves taking the expectation of all individual least squares periodograms across the ensemble. Hence, the next iteration of the re-weighting matrix is built up over a loop of all process records in the ensemble.

6. Each individual record must undergo least squares spectrum estimation, weighted by the previous re-weighting matrix. This is shown in the diagram using the Moore–Penrose pseudoinverse of AW.

7. Here, the least squares periodogram produces pairs of coefficients akin to the complex coefficients of the DFT. As the power spectrum model is not dependant on phase, the weighting of odd and even (in this case Sin and Cosine) components separately is not desired. Hence, the individual coefficients are split into pairs.

8. The Euclidean norm for these pairs is taken to convert them into frequency-only dependant coefficients.

9. The coefficients are then duplicated to make identical weighting pairs and fed into a diagonal matrix (i.e., entries outside the $i = j$ counting from top left to bottom right are zero). The effect of using identical pairs will weight Sine and Cosine components of the same frequency equally on the next iteration. These diagonal coefficients are added for each record in the ensemble. Once the loop at point 5. has ended, the function jumps back to point 3., the next re-weighting

matrix is set equal to the filled W_2 and then W_2 is once again set to zero. Hence, a new re-weighting matrix is used each time the entire process record ensemble is iterated through.

10. Finally L1-minimization is used upon the weighted basis to compute the final basis coefficients.

Not only does this decrease the computational effort of the re-weighting to a large extent, but also negates the problem of very large false peaks. A least-squares solution is more likely to underestimate the power of the key frequencies and also create significant noise elsewhere. These features are not desirable for the final power spectrum estimation, especially when the spectrum is assumed to be relatively sparse (i.e., not spread out over the entire domain). However, as long as least-squares minimization is able to identify key frequencies as being higher than other unwanted frequencies, it proves to be a reliable source for updating the re-weighting matrix. Therefore the proposed approach is to use the ensemble power spectrum estimated via least-squares to iteratively update the re-weighting matrix, before finally ceasing the procedure once the weights have stabilized. The final re-weighting matrix can next be utilized to appropriately modify the basis. L1 minimization is then used to estimate the final power spectrum.

4.5 Numerical examples

In this section, selected numerical examples for CS reconstruction of missing data in environmental stochastic process time-histories are presented for both standard / single pass CS and adaptive basis CS procedures. In both cases, examples include stationary and non-stationary process problems, utilizing Fourier and harmonic wavelet bases, respectively.

4.5.1 Single pass compressive sensing

CS reconstruction is first applied to simulated stationary sea-wave and earthquake ground acceleration processes. Next, simulated separable and non-separable earthquake processes are generated compatible with a prescribed EPS to demonstrate non-stationary process reconstruction.

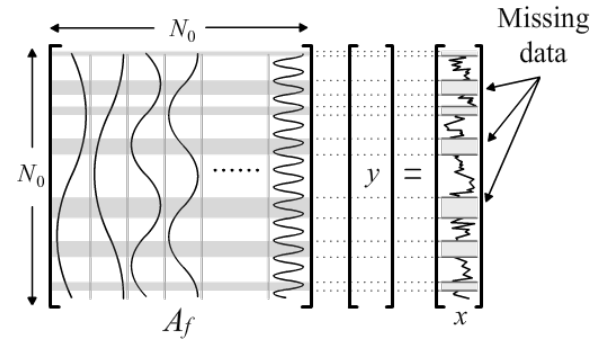
All results are compared to identical problems solved via zero-padding followed by the corresponding scaled transform (Fourier or wavelet).

Stationary simulated processes

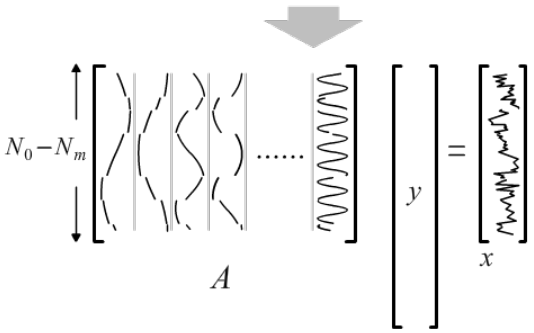
Stationary process records are generated compatible with a JONSWAP sea-wave spectrum (described in section 2.5.2) to simulate wave height over time (Eq.2.66). This spectrum is arranged with a very sharp, strong peak, making it relatively sparse (narrow-band) in the frequency domain. Compatible time-histories of length $N_0 = 256$ are generated by utilizing Eq.2.53. The following spectrum parameters are used in Eq.2.66: $\alpha = 0.03$, $\omega_p = 0.5$, $\gamma = 3.3$, and $\sigma = \begin{cases} 0.07 & \omega < \omega_p \\ 0.09 & \omega > \omega_p \end{cases}$.
For comparison purposes, the procedure is demonstrated on simulated data compatible with the stationary spectrum model, Eq.2.65) as well. This spectrum is purposely arranged with wider peak and longer tail, and as a result is less suitable, in theory, for CS reconstruction. Again, stationary processes of length $N_0 = 256$ are generated utilizing Eq.2.53. The following spectrum parameters are used in Eq.2.65: natural frequency, $\omega_g = 12 rad/s$, damping ratio $\zeta = 0.4$ and $\alpha = 20$. Furthermore, Gaussian white noise

1. • Construct Fourier sensing matrix, A_f

   Separate real and imaginary components into separate columns



   • Reduce A_f based on locations of missing data



2. • Initialize reweighting matrix W, $W_2 = \mathrm{I}_{N_0}$

   where I *is the identity matrix*

3. while $|W_2 - W| >$ *threshold* *(skip "while" condition on 1st iteration)*

4. $W = W_2$

   $W_2 = \text{zeros}\left(N_0 \text{ by } N_0\right)$

   • Generate reweighting matrix via least-squares

5. for j=0 to m *where* m = *number of records in ensemble*

6. $y = (AW)^{\mathrm{T}}\left((AW)(AW)^{\mathrm{T}}\right) x_m$

7. $y = \left[y_1, y_2, y_3, y_4 \ldots, y_{N_0-1}, y_{N_0} \right]$

8. $\left\| y_1, y_2 \right\| \left\| y_3, y_4 \right\| \quad \left\| y_{N_0-1}, y_{N_0} \right\|$

9. $W_2 = W_2 + \text{diag}\left(\left[w_1, w_1, w_2, w_2 \ldots, w_{N_0/2}, w_{N_0/2} \right]\right)$

   end

   end

10. • Use L1-minimization to compute final Fourier coefficients

    $y = \min |y|_{\mathrm{L1}}$ *subject to* $x = Ay W_2$

Figure 4.18: CS with adaptive basis method using a Fourier basis)

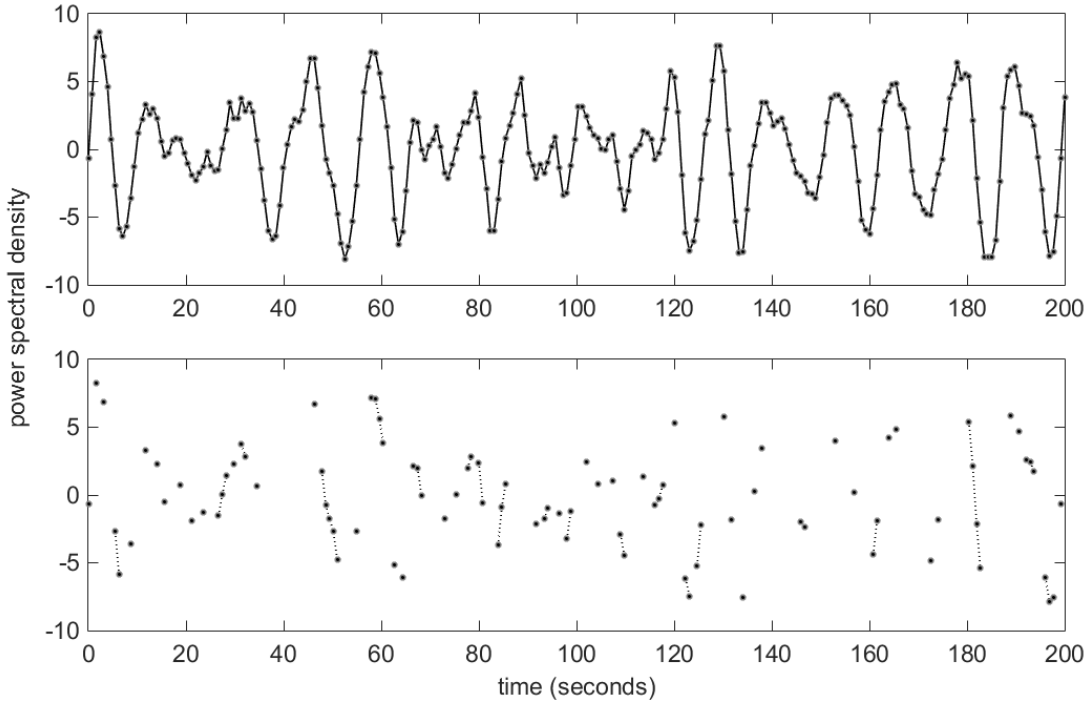


Figure 4.19: Example JONSWAP process (top) and the same process with 65% missing data (bottom)

is added to the process realizations of magnitude $1/5^{\text{th}}$ the standard deviation of the record in both cases. (White noise is a random signal with constant power spectral density). Three examples are shown, two for the JONSWAP spectrum and one for that based on Eq.2.65. For each example, ten process realizations are generated and data is removed via the methods described in section 2.6.2. The spectrum is then estimated via scaled zero-padding and L1 minimization in the Fourier basis. The first example is of the JONSWAP process with 65% missing data at random locations. Figure 4.19 shows a full example process realization alongside the same with 65% missing data. Figure 4.20 shows that L1 minimization in a Fourier basis (CS reconstruction) has correctly identified the key peak frequency of the process, and only lost roughly 25% of its original power. Further, comparing with the scaled zer- padding reconstruction which has a very low spectral peak and a long tail of noise across all other frequencies, it is clear that the CS approach is far superior.

The second example (Figure 4.21) uses the same spectrum, but this time with 50% missing data at intervals of length $N_0/32$ (Eq.2.75). The arrangement of missing data is less similar to an ideal CS sampling matrix (e.g. uniform random Fourier); therefore more data is needed for reliable results. Under these conditions, the CS reconstructed spectrum still out-performs scaled zero-padding significantly, showing a much more defined spectral peak.

The final stationary example shows the limitations of CS for spectral reconstruction when the signal is not as sparse in the Fourier domain. For the process defined by Eq.2.65, the number of missing data was reduced to 50% to account for the broader spectrum, by which point a zero-padding solution became a viable option (although it still retains greater noise at higher frequencies). The differences between the two approaches shown in Figure 4.22 are far less significant than in previous examples. In this case, either reconstruction method could be considered.

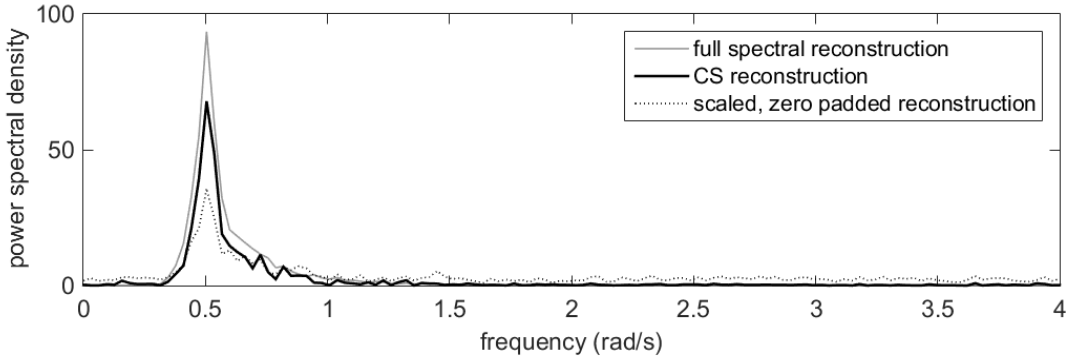


Figure 4.20: JONSWAP power spectrum reconstruction from 10 stationary process records via L1 minimization and scaled, zero-padding for 65% missing data at uniform random locations

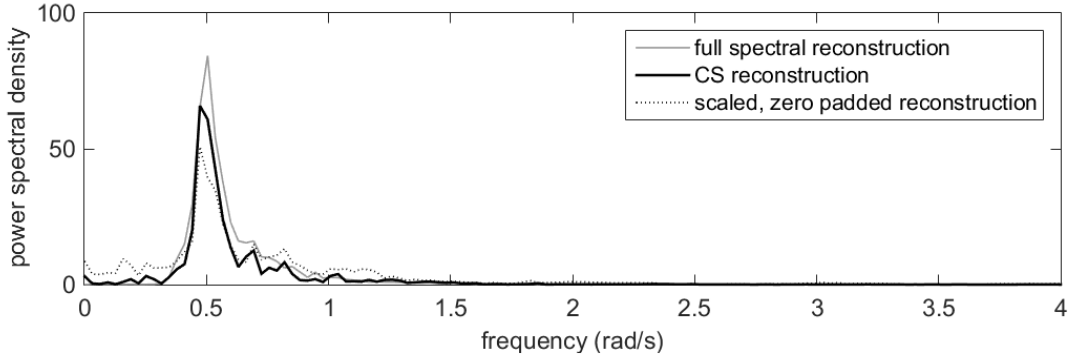


Figure 4.21: JONSWAP power spectrum reconstruction from 10 stationary process records via L1 minimization and scaled, zero-padding for 50% missing data over randomly located fixed intervals of length $N_0/32$

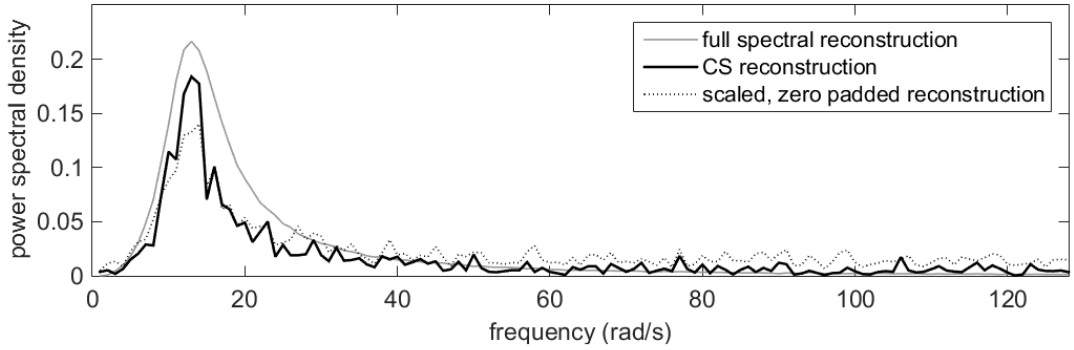


Figure 4.22: Eq.2.65 power spectrum reconstruction from 10 stationary process records via L1 minimization and scaled, zero-padding for 50% missing data at uniform random locations

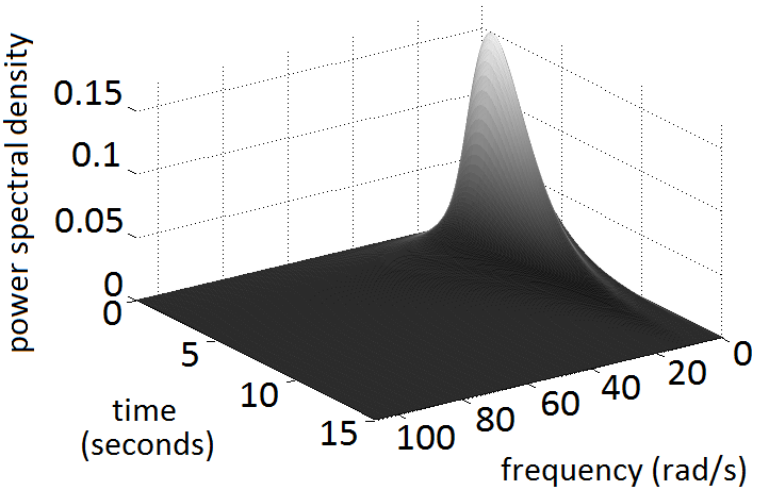


Figure 4.23: Separable target spectrum drawn from Eq.4.23

Non-stationary simulated processes

In this section results are shown for both separable and non-separable examples. For the separable case, process records are generated compatible with a time modulated Clough-Penzien earthquake spectrum,

$$S(\omega, t) = g(t)^2 S(\omega), \tag{4.23}$$

where $S(\omega)$ is defined by Eq.2.67 with parameters: $S_0 = 0.06$, $\omega_f = 1$, $\zeta_f = 0.6$, $\omega_g = 10$, $\zeta_g = 0.4$ and $g(t)$ is the envelope function defined by Eq.2.68. The envelope function is given the parameters $k = 4$, $a = 0.3$ and $b = 0.6$. Finally, for the non-separable case, the method is tested on the non-separable earthquake process defined by Eq.2.69, and shown again in Figure 4.25.

Naturally, for the non-stationary reconstructions, a harmonic wavelet basis is used. For the separable spectrum, the bandwidth of the wavelets is set at $1/32^{\text{nd}}$ the length of the corresponding Fourier power spectrum (or 8 rad/s for $N_0 = 512$). This resolution is more relaxed in the time domain than in the frequency domain; however, there is still a significant loss of resolution with such a rapidly changing signal (both in frequency and time). For the non-separable spectrum (Eq.2.69), a larger bandwidth of $1/16^{\text{th}}$ the length of the power spectrum is used ($16 rad/s$). This allows the GHWT to better capture the more rapid changes in power over time. Figures 4.23 and 4.25 show the target spectra drawn directly from Eq.4.23 and Eq.2.69. Comparatively, Figures 4.24 and 4.26 show the average power of the GHWT of 25 time-histories of the separable and non-separable processes (generated using Eq.2.62). There are power losses at the peak frequencies as they are spread over a number of frequencies and larger time intervals. Despite this loss, the spectra in Figures 4.24 and 4.26 still present a useful result, showing the location of the peak frequencies and the trends over time. Because of these limitations of the GHWT with no missing data, Figures 4.24 and 4.26 are considered to be the target spectra for reconstruction in the separable and non-separable examples respectively. In the following cases, 50% of the data are removed, and the power spectra are estimated based on an ensemble of 25 realizations.

For the separable process with uniformly distributed missing data, both the scaled zero-padding (Figure 4.27) and CS (Figure 4.28) reconstructed spectra identify the frequency location of the power peak and the decaying trend over time. However, CS has produced a much sharper peak and significantly less noise at higher frequencies than

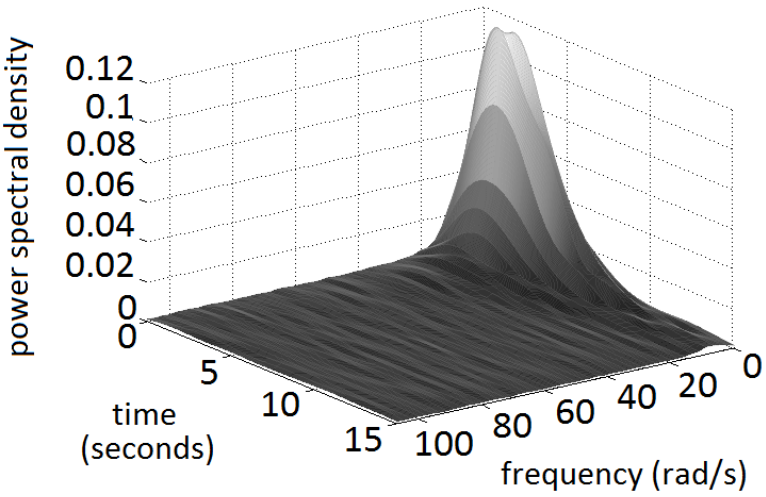


Figure 4.24: Averaged GHWT spectrum estimation with no missing data using 25 time-histories compatible with Eq.4.23

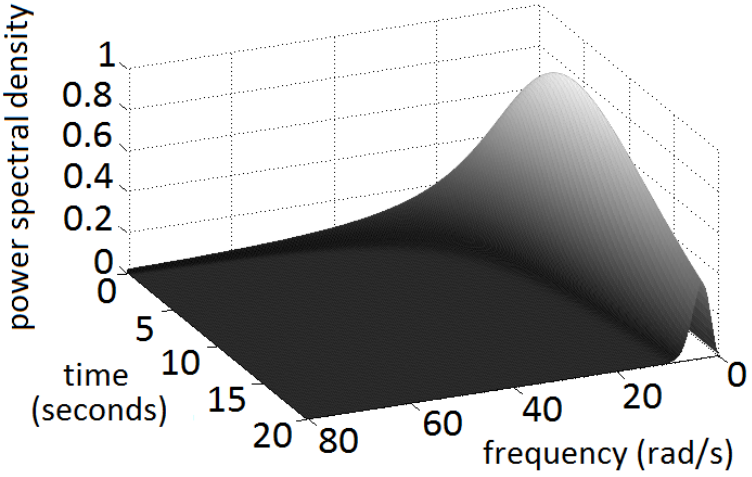


Figure 4.25: Non-separable target spectrum drawn from Eq.2.69

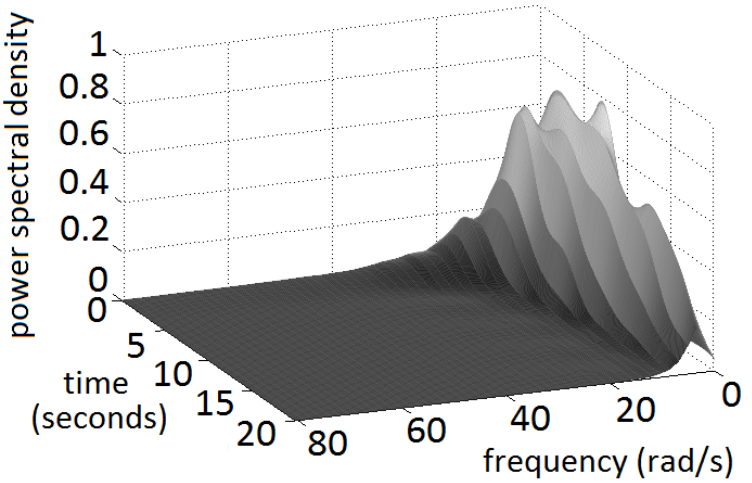


Figure 4.26: Averaged GHWT spectrum estimation with no missing data using 25 time-histories compatible with Eq.2.69

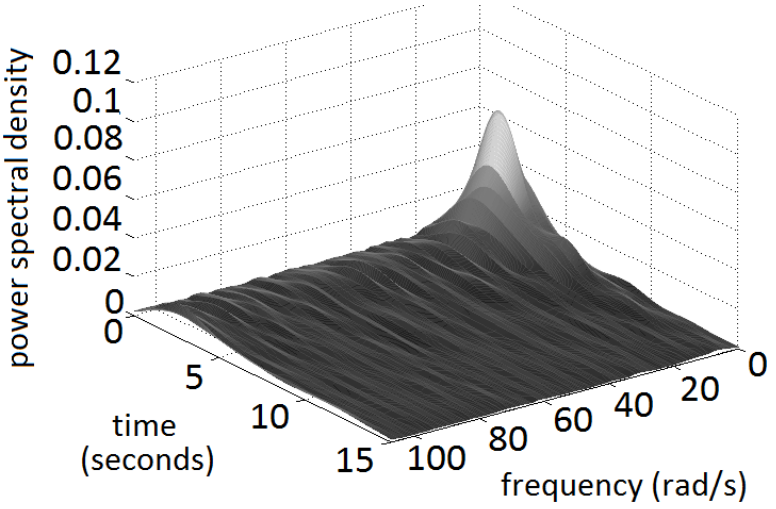


Figure 4.27: Separable earthquake power spectrum reconstruction from 25 stationary process records via zero-padding for 50%/ missing data at uniform random locations

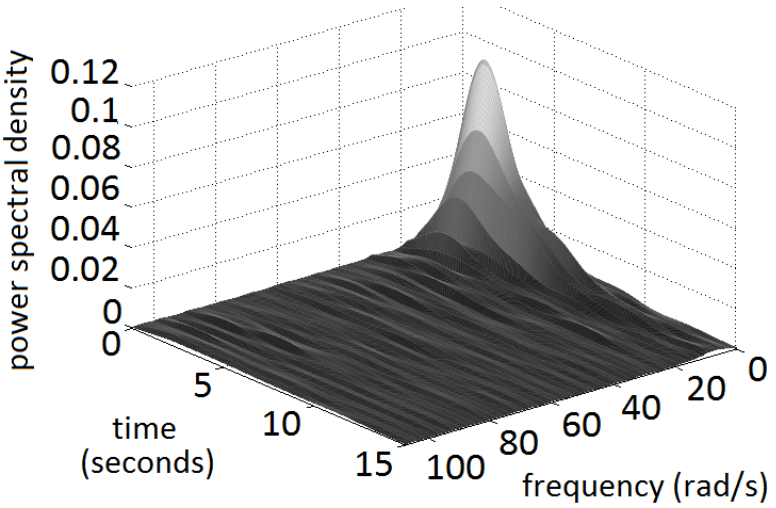


Figure 4.28: Separable earthquake power spectrum reconstruction from 25 stationary process records via L1 minimization for 50% missing data at uniform random locations

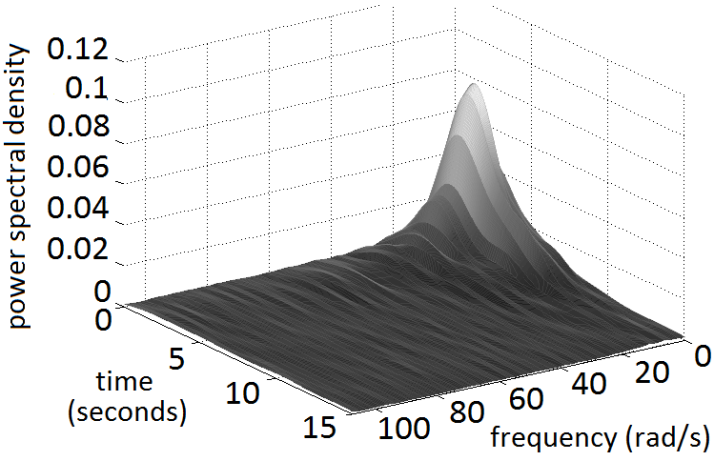


Figure 4.29: Separable earthquake power spectrum reconstruction from 25 stationary process records via zero-padding for 50% missing data over randomly located fixed intervals

zero-padding. Figures 4.29 and 4.30 show the same reconstructions, but with uniformly distributed intervals of missing data of length $N_0/32$. Here the differences are not so apparent, since although the CS reconstruction remains relatively unchanged, the scaled zero-padding solution has much lower noise than in Figure 4.27. This is because introducing large interval gaps of missing data rather than many individual points of missing data is less likely to generate false powers at higher frequencies. Although not easily identified in Figure 4.29, the zero-padding reconstruction does show significant powers right down to the minimum frequency (behind the peak), whereas it should drop down to zero (this is an important feature of Eq.2.67). The CS reconstruction much more accurately reproduces this drop at the lowest frequencies.

For the non-separable process, with uniformly distributed missing data there is a clear advantage of using CS (Figure 4.32) over scaled zero-padding (Figure 4.31). The low-power, high frequency spike that occurs at the very beginning of the process is lost in a sea of noise in Figure 4.31. The significance of this difference between the two reconstructions is made even clearer when using them to generate new process realizations. Figure 4.35.a shows a non-stationary process realization compatible with Eq.2.69; note the high frequency oscillations at the start, slowing over time. Figure 4.35.b shows a process realization generated from Figure 4.32 (the CS reconstruction). The high frequency content at the start trending into a lower frequency signal has clearly been captured in this case. Finally Figure 4.35.c shows a process realization generated from Figure 4.31 (the scaled zero-padding reconstruction). Here it is rather difficult to identify visually any frequency dependent change over time, and further high frequency noise has a significant presence throughout. For the interval-based missing data examples (Figures 4.33 & 4.34), as in the separable case, the differences are less apparent. However, Figure 4.34 does still present less noise at higher frequencies and has a more defined shape than Figure 4.33, more closely matching that of Figure 4.26.

4.5.2 Adaptive basis compressive sensing

Now with the adaptive CS basis, examples are given for stationary and non-stationary processes, again utilizing Fourier and generalized harmonic wavelet bases respectively. Spectra estimated via adaptive basis CS are compared against standard CS, zero-padding and full (no missing data) reconstructions. For all of the examples, 25 512-

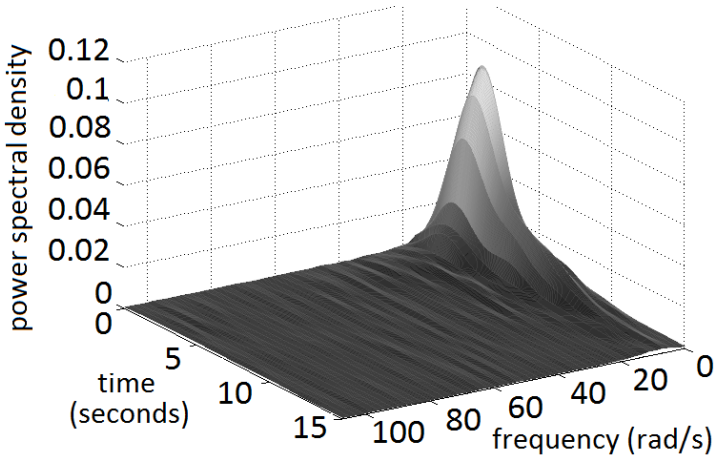


Figure 4.30: Separable earthquake power spectrum reconstruction from 25 stationary process records via L1 minimization for 50% missing data over randomly located fixed intervals

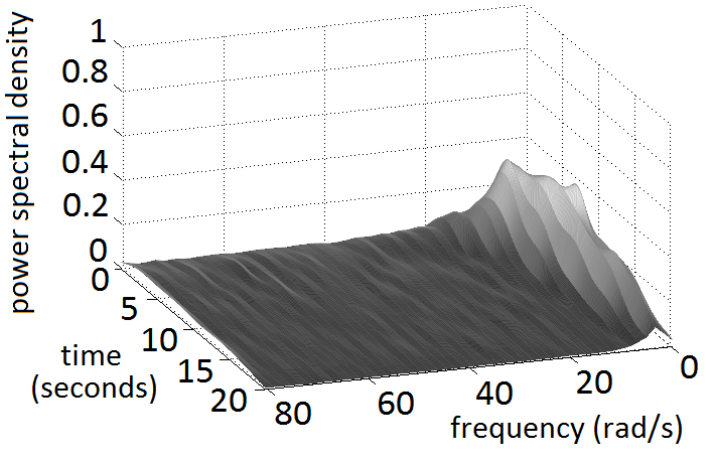


Figure 4.31: Non-separable earthquake power spectrum reconstruction from 25 stationary process records via zero-padding for 50% missing data at uniform random locations

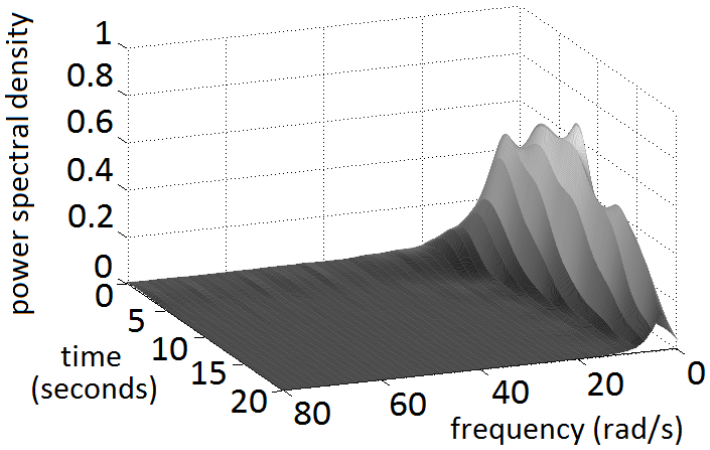


Figure 4.32: Non-separable earthquake power spectrum reconstruction from 25 stationary process records via L1 minimization for 50% missing data at uniform random locations

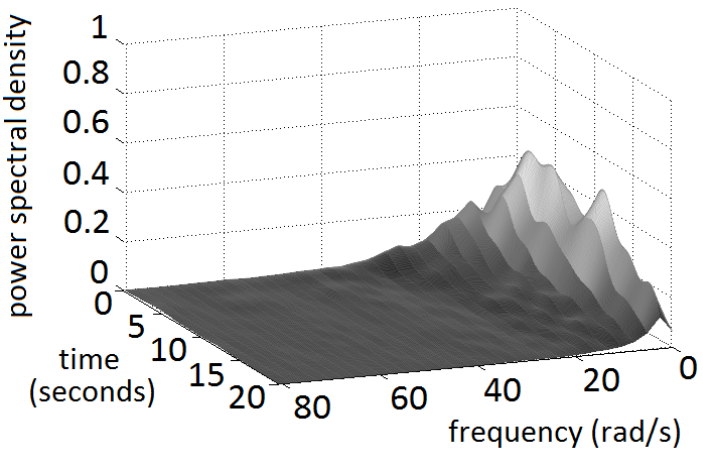


Figure 4.33: Non-separable earthquake power spectrum reconstruction from 25 stationary process records via zero-padding for 50% missing data over randomly located fixed intervals

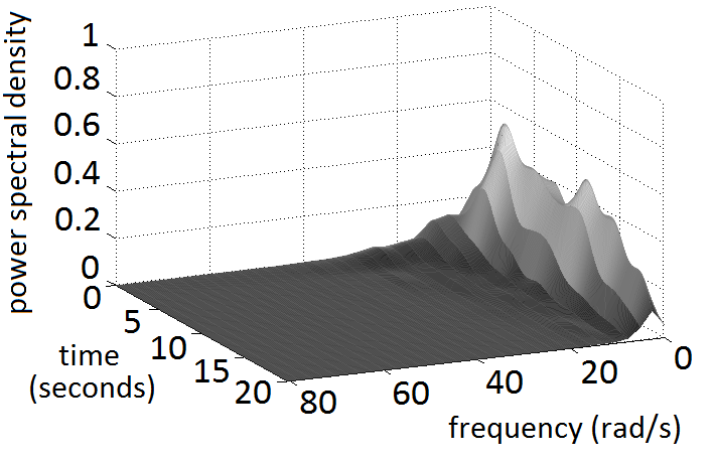


Figure 4.34: Non-separable earthquake power spectrum reconstruction from 25 stationary process records via L1 minimization for 50% missing data over randomly located fixed intervals

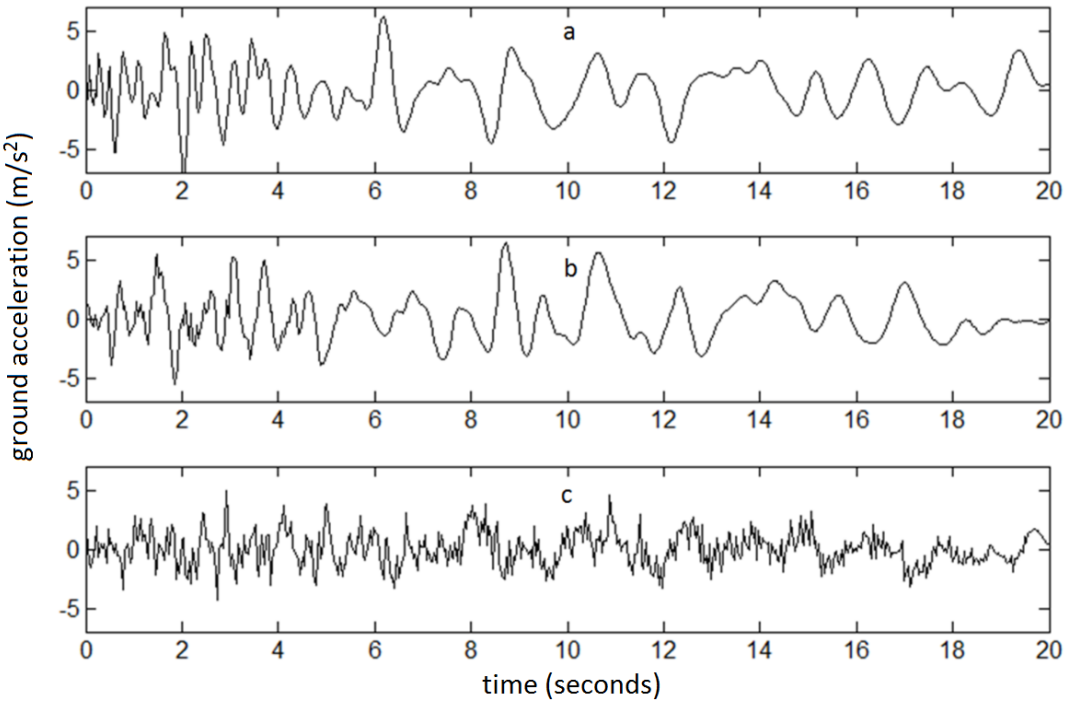


Figure 4.35: (From top to bottom) single realization of Eq.2.69, single realization from Figure 4.34 and a single realization of Figure 4.33

point process records are generated, with normally distributed white noise within the standard deviation of the process, and 75% of the data is removed. In the stationary case, process records are generated via Eq.2.53 compatible with the stationary power spectrum, Eq.2.65, where the natural frequency, $\omega_g = 10$ rad/s, damping ratio $\zeta = 0.25$ and $\alpha = 5$. For the non-stationary case, Eq.2.65 is modified by the envelope function,

$$S(\omega, t) = S(\omega)g(t); \quad g(t) = b(e^{-ct} - e^{-2ct}), \tag{4.24}$$

where $b = 4$ and $c = 8$. Process records are generated via Eq.2.62. Both spectra are sharp and decay quickly, making them relatively sparse in their respective domains.

Stationary simulated processes

For the stationary power spectrum estimation, the advantage of coupling CS with the adaptive basis is striking. Figure 4.36 shows the source spectrum, matched closely by its estimation with no missing data (the estimation with no missing data rather than the source is considered to be the target spectrum). In the same figure, zero-padding, single pass CS, and CS with adaptive basis are also shown. The zero-padding gives a very poor spectrum estimation despite significant re-scaling (with 75% missing data, the zero-padding estimation has been multiplied by 4). Showing a much more defined peak at the key frequency of 10 rad/s, the CS solution gives a significantly higher estimate of the maximum power; however it is still only half of that of the target. In contrast to zero-padding and standard CS, by introducing the adaptive basis, the CS reconstruction has produced an estimation that is very close to the target spectrum. Figure 4.37 and Figure 4.38 show how the CS reconstruction and the difference between the estimate and the target changes each time the sensing matrix is re-weighted.

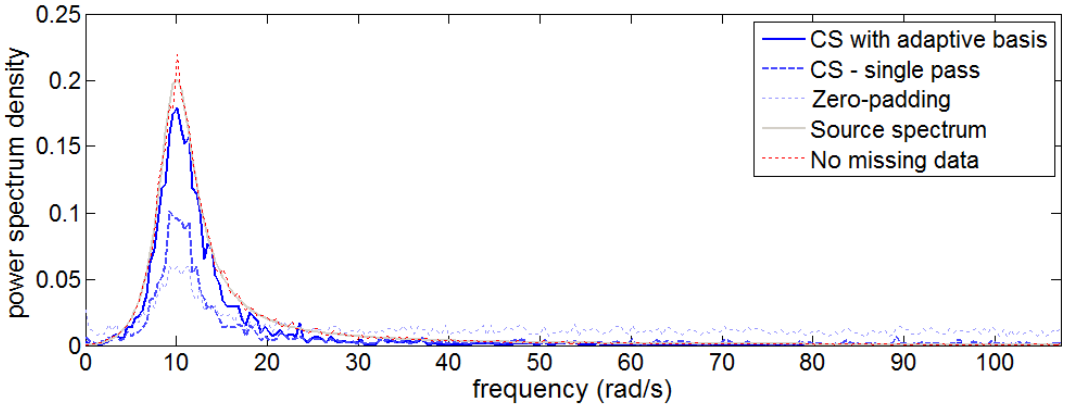


Figure 4.36: Comparison of stationary spectrum reconstructions

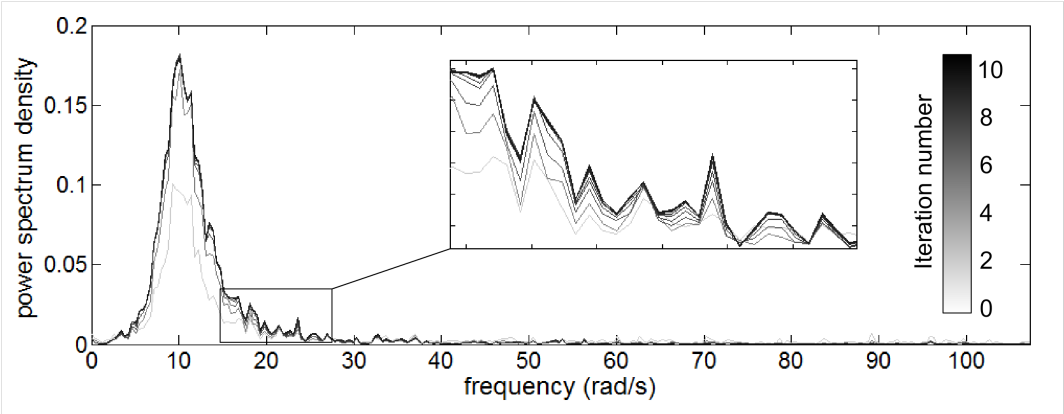


Figure 4.37: Convergence of stationary spectrum reconstructions to a single estimate

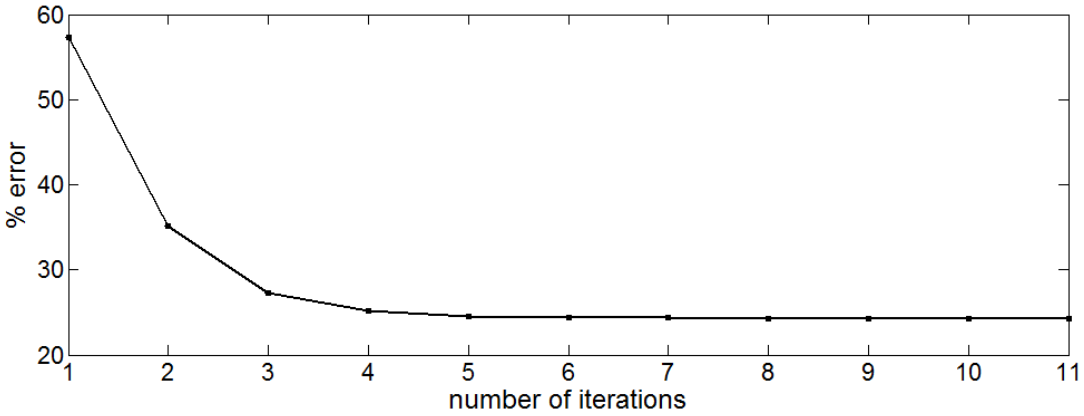


Figure 4.38: Convergence of % error between estimated spectrum and target spectrum

Non-stationary simulated processes

In the first example shown in Figure 4.39, missing data were located randomly as in the stationary case. Figure 4.39.a shows the target spectrum. This target is based on the average harmonic wavelet transform power spectrum for all 25 records with no missing data. Figure 4.39.b shows the CS estimated power spectrum with 75% missing data using the adaptive basis. Figure 4.39.c and Figure 4.39.d show the standard CS and zero-padding estimations respectively. Zero-padding gives a very poor approximation of the spectrum, as in the stationary case. Further it has a very long tail of noise of significant magnitude compared to the size of the peak. Standard CS improves greatly on the zero-padding result, decreasing noise and better estimating the magnitude of the peak. However it is clear that Figure 4.39.b gives the best approximation to the target with a significant improvement over Figure 4.39.c.

In the second example shown in Figure 4.40, missing data were grouped in clusters of length $N/10$ (Eq.2.75). When missing data occurs in groups, the zero-padding spectrum estimation becomes more effective (Figure 4.40.d). Because the missing data occurs less often and for longer periods of time, zero-padding no longer favours high frequency noise tails, and instead noise is grouped around the lower frequencies; this can also be seen in Figure 4.40.c for the standard CS reconstruction. In this case, standard CS gives a lower estimation of the peak spectral power than zero-padding, and although the peak is slightly more defined it is not necessarily a superior estimation. However, when CS is combined with the adaptive basis, the power spectrum estimation is clearly a better match, with the highest peak and closest shape to that of the target (Figure 4.40.b).

4.5.3 Results overview

The results of the CS based reconstructions demonstrate the extent to which the reliable spectral estimates may be produced for up to 75% missing data. For processes that conform to the requirement of sparsity, both single pass and adaptive basis CS show increased performance over the previous ANN method. 25 samples are used in non-stationary numerical examples and 10 for stationary examples to show that the technique is applicable with realistic data set sizes. Further, it is shown in Figure 4.41 that significantly more samples has little effect on the average error in the estimated spectrum.

4.6 Chapter summary

As with the previous chapter, the research herein provided in Chapter 4 has also benefited in its development and outputs from the author presenting emerging findings at international conferences [81, 82, 83]. Further, this work has lead to an ongoing collaboration with the oceanography department at the Indian Institute of Technology Madras (Through a European Commission funded project, PLENOSE), to investigate thresholding difficulties when collecting and processing oscillating wave column data.

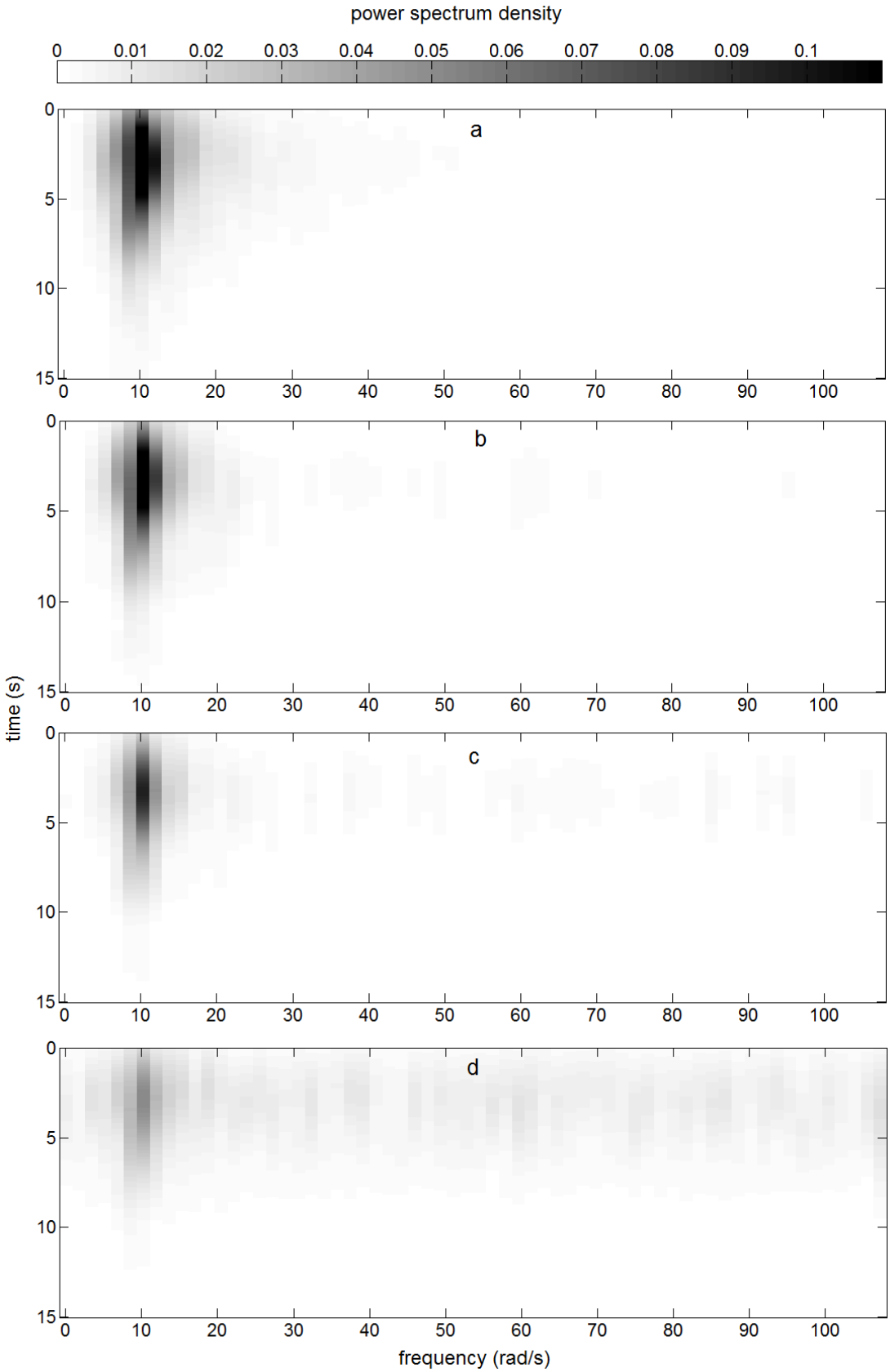


Figure 4.39: a. Target spectrum for Eq.4.24 with no missing data, b, c & d. Spectrum reconstructions with 75% missing data at random locations for CS with adaptive basis, CS without adaptive basis and zero-padding respectively

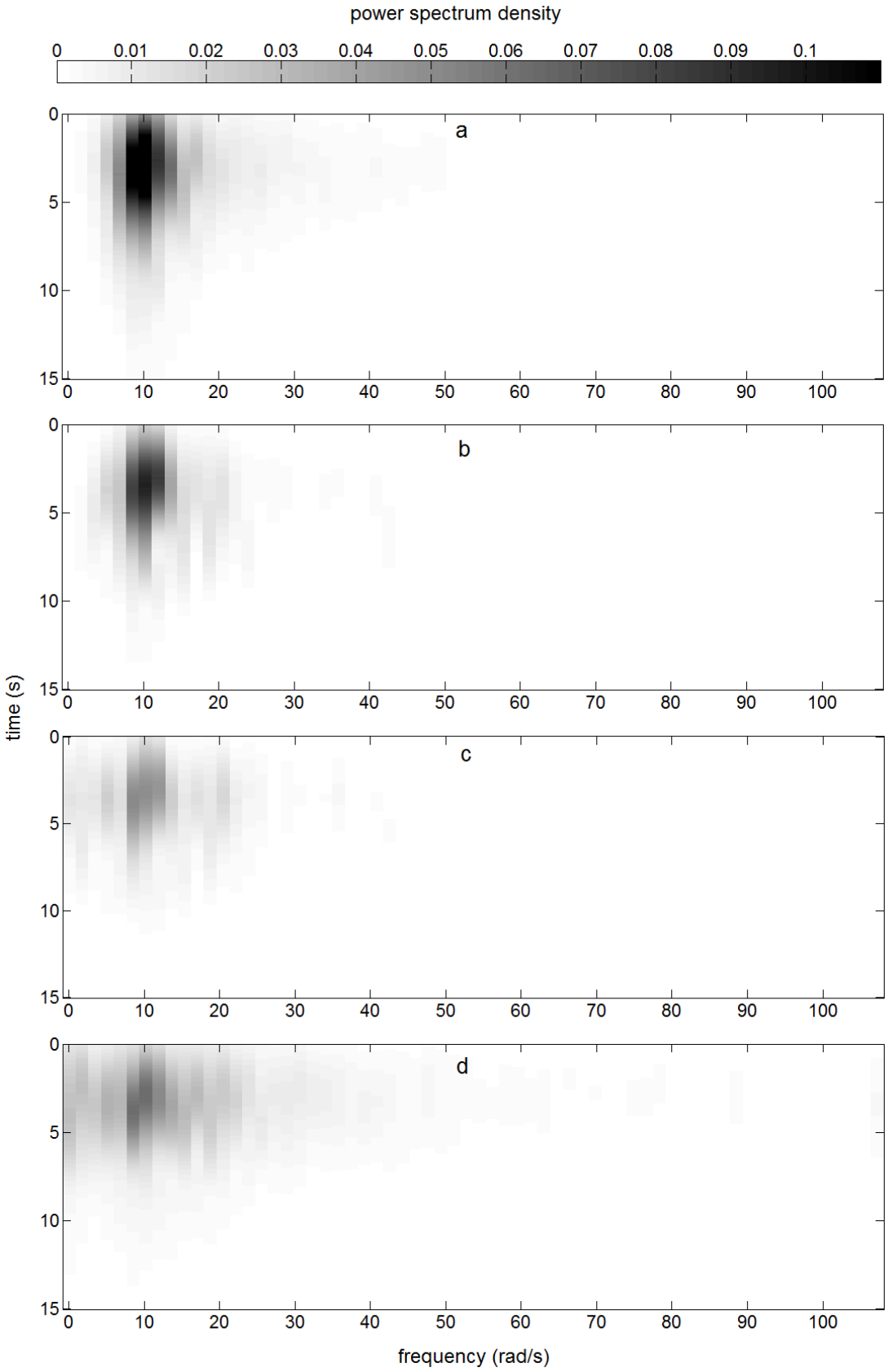


Figure 4.40: a. Target spectrum for Eq.4.24 with no missing data, b, c & d. Spectrum reconstructions with 75% missing data over random intervals for CS with adaptive basis, CS without adaptive basis and zero-padding respectively

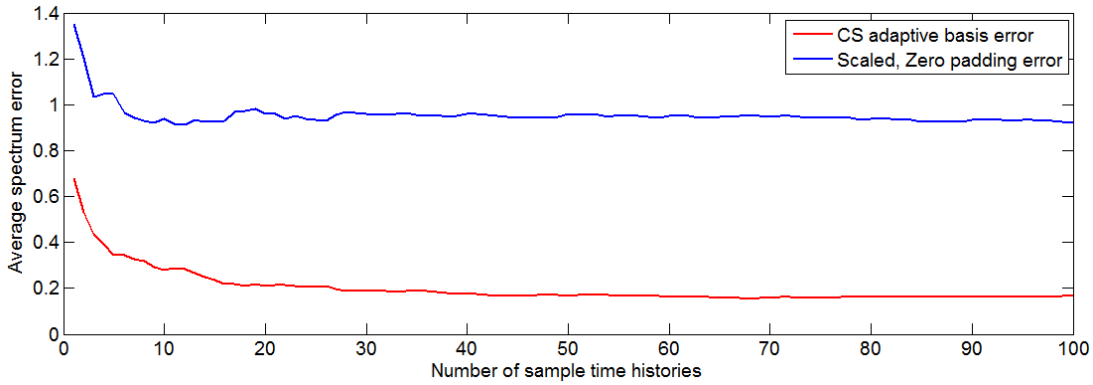


Figure 4.41: Average difference between estimated non-separable power spectrum from 100 time histories compatible with Eq.2.69 and adaptive basis CS and zero padded reconstructions. Samples suffer from 50% missing data in uniformly distributed random locations.

Chapter 5

Quantifying the uncertainty of stochastic process power spectrum estimates subject to missing data

5.1 Introduction

The focus of Chapters 3 and 4 was to produce a deterministic estimate of the power spectrum via reconstructing missing data in the time domain. Both of these approaches, and those introduced in section 2.6.1 will propagate inaccuracies from missing data predictions in the time domain through to the final spectral estimate. These inaccuracies may in some cases be particularly small; a specific example would be the case where only a small percentage of data is missing in a stationary signal with a highly narrow band spectrum. However, in many cases uncertainty due to missing data can be of great significance, particularly as all of the models listed previously for reconstructing the spectrum in the presence of missing data are used to produce a deterministic estimate for the data missing in the time domain, and ultimately, for the power spectrum coefficients. Clearly, these power spectrum estimates could be misleading, as they provide no information concerning the degree of uncertainty related to the original incomplete data; thus, potentially leading to poor decision-making if the magnitude of potential error is not properly quantified. Indeed, even without the presence of missing data, there are many cases where simply building a Fourier power spectrum without any understanding of the underlying process can lead to erroneous results. For example, [22] states that in a well cited data set describing periodic variations in the Earth's orbit [84], falsely detected peaks from the periodogram are given unwarranted significance, and when compared with a maximum likelihood analysis (e.g.[85]), all but disappear. Therefore when combining spectral estimation with missing data, we must be all the more prudent to avoid over confidence when making predictions from our analyses. The importance of understanding uncertainty in spectral estimates is made even clearer when considering applications where design decisions that affect system reliability can impact human lives. An example could be the case of structural system design where "resonance" is the key mechanism. In this regard, miscalculated resonant frequencies in the analysis phase of structural system design could lead to catastrophic failures. While missing data methods listed in section 2.6.1, as well as those developed in Chapters 3 and 4, rely solely on deterministic estimates of the power spectrum, as an alternative and based on the aforementioned arguments, it can be argued that there is merit in relaxing the assumptions on the missing data by modelling them as random variables. In this regard, by assuming a probability density function (PDF) to represent the missing data and ultimately the nature of the process as a whole, the power spectrum may be characterised by a range of possible values rather than a single value at a given frequency or time-frequency band. This "probabilistic" power spectrum provides a useful comparison tool for assessing the reliability of alternative reconstruction methods. In this chapter, a technique is demonstrated for determining a closed-form expression for the "probabilistic" power spectrum in cases where the missing data are

assumed to follow a Gaussian distribution. However, the method is flexible to non-Gaussian random variables in many cases, by virtue of the central limit theorem (e.g. [86]) as is highlighted in the following sections. Further, the approach to uncertainty quantification in power spectrum estimation presented in this chapter yields results that can, alternatively, be generated via Monte-Carlo simulation. In this regard, for a single time history with missing data, random data compatible with a given distribution are generated. For each generated set of data, the Fourier or Harmonic Wavelet transform is performed, yielding an estimate for the power spectrum. The process is repeated many times, eventually resulting in a histogram for the possible power spectrum values corresponding at a specific frequency. However, to achieve an accurate distribution estimate, this approach can be computationally highly demanding.

5.2 Power spectrum PDF methodology

In the following subsections, closed-form expressions for power spectrum estimate PDFs for both stationary and non-stationary stochastic process records are derived. First, the stationary case is studied, and PDFs for the power spectrum estimates are determined corresponding to each and every frequency value. This is done by resorting to the definition of the DFT in conjunction with an ergodicity assumption for the underlying stationary stochastic process (see section 2.2.4). Following this, harmonic wavelets are utilized as a basis for representing non-stationary stochastic processes and for estimating evolutionary power spectra. In both cases, the PDF of the power spectrum is derived under the assumption that missing data are standard, normal uncorrelated random variables. This restriction serves to simplify the derivation of the spectrum and provide a basic basis for further study. In this regard, example time histories are normalized to have standard deviation 1 and zero mean. Further, this research has already been extended to include correlated missing data based on a Kriging model in the spectrum PDF estimation [87].

5.2.1 Stationary case

To produce power spectrum estimates for real, discrete data, an appropriate transformation methodology to project from the time domain to the frequency or the joint time-frequency domain is required. For discrete time data where only a single process realization is available, this was defined in section 2.5. For the single process record case, the time averaged stationary state of the process record is assumed to be the same as the average over all states, hence the process must be ergodic (section 2.2.4). I.e., The power spectrum $S_X(\omega)$ can be estimated by computing the temporal mean value of the square of the DFT of the available record,

$$S_X(\omega_k) = \lim_{T\to\infty} \frac{2\Delta T}{T} \left| \sum_{t=0}^{T-1} x_t e^{-2\pi i k t/T} \right|^2 \tag{5.1}$$

where T is the number of data points, t is the data point index in the record, ΔT is the sampling time increment, and k is the integer frequency for ω_k (i.e. $\omega_k = \frac{2\pi k}{T_0}$ where T_0 is the total length in time of the record). In the ensuing analysis, the symbol $\lim_{T\to\infty}$ in Eq.5.1 is omitted for convenience, assuming that the number T of available data is large enough for practical purposes. Further, in the case where there are missing data, clearly, one or more of x_t are unknown. Then, Eq.5.1 can be written in the form

$$S_X(\omega_k) = \frac{2\Delta T}{T} \left| \sum_{t_\alpha} x_{t_\alpha} e^{-2\pi i k t_\alpha/T} + \sum_{t_\beta} x_{t_\beta} e^{-2\pi i k t_\beta/T} \right|^2 , \tag{5.2}$$

where t_α and t_β represent the index positions of known and unknown data, respectively.

Next, instead of attempting to impose certain rather strict assumptions on x_t with the aim of obtaining deterministic estimates for x_t by utilizing one the many available methodologies for addressing missing data, a rather relaxed assumption on the nature of x_t is adopted, i.e. the unknown x_t are assumed to be independent Gaussian random variables. Of course, different PDFs can be assumed based on the *a priori* available information (if any) about the missing data. Note, however, that in the context of the maximum entropy principle, the Gaussian PDF has maximum entropy among all real-valued distributions with specified mean and standard deviation (e.g.[88, 89]). Therefore, the assumption of normality imposes the minimal prior constraint beyond these moments; and thus, the Gaussian PDF can be a good candidate for cases where minimal information is known about the missing data. In any case, it can be argued that a probabilistic description of the unknowns imposes fewer restrictions and assumptions on the nature of the missing data, as compared to a purely deterministic description.

Further, by taking advantage of the fact that the sum of a series of such random variables is also Gaussian with variance equal to the sum of the series variances, it is possible to group random variables to simplify the function. Also, by virtue of the Central Limit Theorem (CLT), (e.g. [90, 91]), we may relax the assumption of modelling the missing data as Gaussian random variables, and include other sets of non-identical distributions. Hence, for large numbers of independent random variables, their summations will tend towards the same Gaussian distributions as those defined in this section. Note also that for small numbers of non-Gaussian missing data where the CLT may not be assumed to govern their summation, the following general procedure is still valid, and may be followed while specifically taking account of the alternative (non-Gaussian) distributions.

Before we can sum the random variables in Eq.5.2, we must account for the fact that the standard deviations in this case are complex, i.e. $x_{t_\beta} e^{-2\pi i k t_\beta / T}$ has a real and imaginary component. Simply taking the sum of the squares of these complex deviations in an attempt to define a single complex random variable would not make sense. This is because the PDF of a single complex Gaussian random variable could be plotted in two dimensions along a vector in the complex plane (this may also be visualized as a joint density function between real and imaginary parts with unit correlation). However, if we plotted the PDF of a sum of complex random variables, assuming that the complex deviations were not multiples of each other, they would yield a three-dimensional joint PDF in the complex plane with correlation < 1. In order to achieve the latter, the complex coefficients in Eq.5.2 must be split into their sine and cosine counterparts, i.e.

$$\begin{aligned} S_X(\omega_k) = \frac{2\Delta T}{T} \Bigg| \sum_{t_\alpha} x_{t_\alpha} \cos\left(\frac{2\pi k t_\alpha}{T}\right) + \sum_{t_\beta} x_{t_\beta} \cos\left(\frac{2\pi k t_\beta}{T}\right) ... \\ -i \left(\sum_{t_\alpha} x_{t_\alpha} \sin\left(\frac{2\pi k t_\alpha}{T}\right) + \sum_{t_\beta} x_{t_\beta} \sin\left(\frac{2\pi k t_\beta}{T}\right) \right) \Bigg|^2 , \end{aligned} \tag{5.3}$$

where X_{n_β} are independent Gaussian random variables representing the missing data, x_{n_β}. In the following, random variables allocated to all of the missing points in time are assumed to be identically distributed as well. Note that as previously stated, these assumptions serve to simplify notation and are not a requirement of the analysis. Available data are pre-scaled so that their standard deviation (σ) is equal to one, and shifted so that their mean (μ) is equal to zerol; therefore all random variables are also standard normal, i.e. $X_{n_\beta} = \text{Normal}(0, 1)$. Eq.5.3 can now be simplified by replacing the real and imaginary parts with single Gaussian random variables, by summing the

variances

$$S_X(\omega_k) = |X_1 - iX_2|^2 \tag{5.4}$$

or equivalently,

$$S_X(\omega_k) = X_1^2 + X_2^2, \tag{5.5}$$

where

$$X_1 = \sqrt{\frac{2\Delta T}{T}} \left(\sum_{t_\alpha} x_{t_\alpha} \cos\left(\frac{2\pi k t_\alpha}{T}\right) + \sum_{t_\beta} x_{t_\beta} \cos\left(\frac{2\pi k t_\beta}{T}\right) \right) \tag{5.6}$$

and

$$X_2 = \sqrt{\frac{2\Delta T}{T}} \left(\sum_{t_\alpha} x_{t_\alpha} \sin\left(\frac{2\pi k t_\alpha}{T}\right) + \sum_{t_\beta} x_{t_\beta} \sin\left(\frac{2\pi k t_\beta}{T}\right) \right) \tag{5.7}$$

Treating X_1 and X_2 as Gaussian by summing means and variances yields

$$X_j = \text{Normal}(\mu_{X_j}, \sigma_{X_j}^2); \quad j = 1, 2 \tag{5.8}$$

where

$$\mu_{X_j} = \sqrt{\frac{2\Delta T}{T}} \left(\sum_{t_\alpha} x_{t_\alpha} \cos\left(\frac{2\pi k t_\alpha}{T} - \varphi_j\right) \right); \quad j = 1, 2; \quad \varphi_j = 0, \pi/2 \tag{5.9}$$

and

$$\sigma_{X_j}^2 = \frac{2\Delta T}{T} \sum_{t_\beta} \cos^2\left(\frac{2\pi k t_\beta}{T} - \varphi_j\right); \quad j = 1, 2; \quad \varphi_j = 0, \pi/2 \tag{5.10}$$

If X_1 and X_2 were independent, $S_X(\omega_k)$ would take a form similar to a non-central chi-squared distribution (e.g.[92]). However, because both X_1 and X_2 are functions of the same pool of random variables (i.e. each missing point contributes the same underlying random variable to both the real and the imaginary parts of the transform), they exhibit some degree of correlation. In this regard, the correlation coefficient for X_1 and X_2 can be written as

$$\rho_{X_1,X_2} = \frac{\frac{2\Delta T}{T} E\left[\sum_{t_\beta} X_{t_\beta} \cos\left(\frac{2\pi k t_\beta}{T}\right) \sum_{t_\beta} X_{t_\beta} \sin\left(\frac{2\pi k t_\beta}{T}\right)\right]}{\sigma_{X_1}\sigma_{X_2}} \tag{5.11}$$

Because the expectation of X_{t_β} with itself for the same t_β is equal to 1, and for different t_β is equal to zero, Eq.5.11 can be equivalently written as

$$\rho_{X_1,X_2} = \frac{\frac{\Delta T}{T} \sum_{t_\beta} \sin\left(\frac{4\pi k t_\beta}{T_\beta}\right)}{\sigma_{X_1}\sigma_{X_2}} \tag{5.12}$$

Further, to simplify the determination of the PDF of $S_X(\omega_k)$, $S_X(\omega_k)$ is arranged in terms of uncorrelated, independent standard normal random variables X and Y in the form

$$S_X(\omega_k) = (aX + \mu_{X_1})^2 + (bX + cY + \mu_{X_2})^2 \tag{5.13}$$

where, $a = \sigma_{X_1}$, $b = \rho_{X_1,X_2}\sigma_{X_2}$, $c = \sqrt{\sigma_{X_2}^2 + b^2}$ and $X = Y = \text{Normal}(0, 1)$. Hence, Eq.5.13 is treated as a function of random variables [93], and we apply the celebrated input-output PDF relationship, i.e., if Z is a function of random variables X and Y, the CDF of Z, $F_Z(z)$ is defined as

$$F_Z(z) = P(Z \le z) = P(g(X,Y) \le z) = P[(X,Y) \in D_z] = \iint_{x,y \in D_z} f_{XY}(x,y)\, dxdy, \tag{5.14}$$

and the PDF of Z is

$$f_Z(z) = \frac{dF_Z(z)}{dz}. \tag{5.15}$$

Considering Eqs.5.13, 5.14 and 5.15, an expression can be derived for the PDF of the power spectrum estimate $S_X(\omega_k)$ at a given frequency ω_k of the form

$$P_{S_X(\omega_k)}(z) = \frac{d}{dz}\left[\frac{1}{2\pi}\iint_{\left(ax+\mu_{X_1}\right)^2+\left(bx+cy+\mu_{X_2}\right)^2\leq z} e^{\frac{1}{2}\left(y^2+x^2\right)}dydx\right]. \tag{5.16}$$

I.e., Eq.5.16 is the PDF of the power spectrum. Finally, resorting to numerical integration of the double integral in Eq.5.16, and setting the integration limits equal to

$$x = \frac{-\mu_{X_1} \pm \sqrt{z}}{a}, \quad y = \frac{-\left(bx+\mu_{X_2}\right) \pm \sqrt{z-\left(a^2x^2+2a\mu_{X_1}x+\mu_{X_1}^2\right)}}{c}, \tag{5.17}$$

the PDF $P_{S_X(\omega_k)}(z)$ can be readily determined.

Overall, it has been shown that in the case where missing data in a realization are modelled as independent identically distributed Gaussian random variables, a simple expression (Eq.5.16) can be derived for the PDF of the underlying process power spectrum value at a given frequency. In this regard, the uncertainty propagation from the measured incomplete signal in the time/space-domain to its power spectrum estimate in the frequency domain is efficiently quantified without resorting to computationally demanding Monte-Carlo simulations. Note again that for non-identically distributed Gaussian random variables, individual variable means and variances are simply added when calculating the statistics of X_1 and X_2. This is also the case when dealing with non-identically distributed non-Gaussian random variables if the central limit theorem holds. To elaborate further on the computational cost aspect, it becomes clear that for long time-histories with a significant amount of missing data (and thus, with a large number of random variables), attempting to generate these "probabilistic" power spectra via a Monte Carlo simulation treatment could become potentially prohibitive. At the same time, the complexity of solving Eq.5.16 would remain unchanged. This is demonstrated in the examples section, where Figure 5.7 shows the required number of FFT simulations to approximate the power spectrum PDF. Further, Eq.5.16 can be potentially used for assessing the performance of alternative power spectrum estimation techniques subject to missing data which provide with a deterministic estimate of the power spectrum value at a specific frequency.

Note that in the case where ergodicity is not assumed and the expectation operator is understood in an ensemble average sense, the stationary power spectrum is estimated by

$$S_X(\omega_k) = E\left(\frac{2\Delta T}{T}\left|\sum_{t=0}^{T-1} x_t e^{-2\pi ikt/N}\right|^2\right). \tag{5.18}$$

Here, for a specific frequency ω_k, the power spectrum $S_X(\omega_k)$ constitutes a scaled sum of random variables, which gives the distribution of the mean outcome of Eq.5.16 over the ensemble set. For standard normal missing data, with each additional sample, the variance of $S_X(\omega_k)$ decreases. Clearly this is to be expected by virtue of the CLT (see also [91, 90]) when using independent random variables, since the standard deviation of $S_X(\omega_k)$ tends towards zero as the ensemble size tends toward infinity i.e.,

$$\sigma_{S_X(\omega_k)} = \frac{1}{N_{rec}}\sqrt{\sum_{i=1}^{N_{rec}}\sigma_{S_{X_i}(\omega_k)}}; \quad \lim_{N_{rec}\to\infty}\left(\sigma_{S_X(\omega_k)}\right) = 0, \tag{5.19}$$

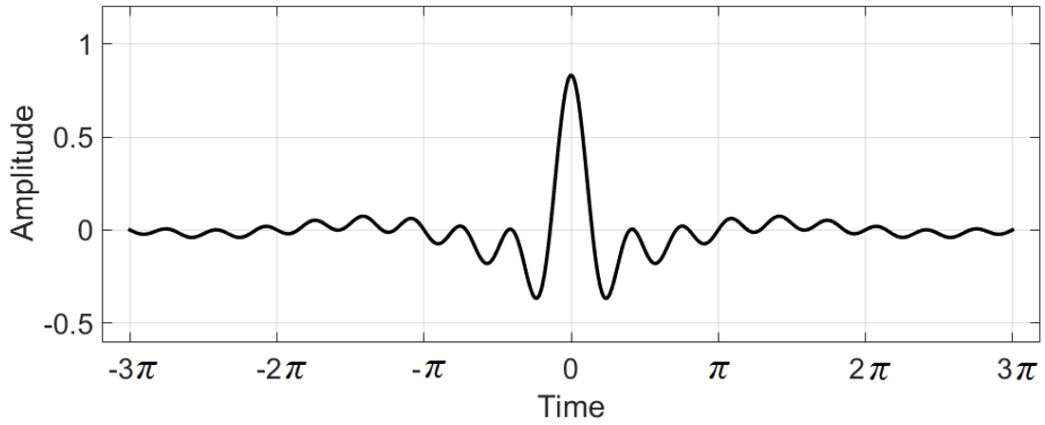


Figure 5.1: Continuous harmonic wavelet defined by Eq.2.44 with for $k = 0$, $m = 1$ and $n = 6$

where N_{rec} is the number of records in the ensemble. It can be readily seen that if an 'infinite' number of realizations with missing data was available, the aforementioned approach/modelling would provide a deterministic estimate for $S_X(\omega_k)$.

5.2.2 Non-stationary case

To build a non-stationary power spectrum PDF, as with the stationary case an appropriate basis is required. In this regard, the derivation of an EPS PDF follows the rigorous harmonic wavelets based representation of the power spectrum in section 2.5.1.

When working with discrete-time signals of finite length, a time-limited form of Eq.2.44 is required to compute the wavelet transform. This is realized by taking the inverse discrete Fourier transform of Eq.2.43, having the effect of wrapping the tails of the wavelet outside the recorded time history back in on themselves [30, 31], yielding a periodic function over T_0, i.e.,

$$\Psi^{G_D}_{(m,n),k}(t) = \frac{1}{\Delta\omega\,(n-m)} \sum_{f=m}^{n-1} e^{2\pi i f \Delta\omega \left(\frac{t}{T_0} - \frac{kT_0}{n-m}\right)}. \tag{5.20}$$

A comparison of the original finite energy and periodic harmonic wavelets are shown in Figures 5.1 and 5.2 respectively.

Notice that by comparing Figures 5.1 and 5.2, the source of possible end-effects discussed in section 2.5.1 when using the discrete GHWT can be clearly seen (i.e. wavelet power is likely to 'leak' around to the other side). Next, utilizing Eq.2.45, and assuming that $T_0 = 2\pi$, T is the number of sample points and t are the integer point indices, we can now write the discrete harmonic wavelet transform of x as

$$W^{G_D}_{(m,n),k} = \frac{1}{T} \sum_{t=0}^{T-1} x_t \sum_{f=m}^{n-1} e^{-2\pi i f \left(\frac{t}{T} - \frac{2\pi k}{n-m}\right)}. \tag{5.21}$$

Combining Eq.2.63 with Eq.5.20, the evolutionary power spectrum for a discrete time recorded process is estimated by

$$S_X(\omega, t) = S^X_{(m,n),k} = E\left[F_x(\omega_{m,n}, k)\right] \tag{5.22}$$

where

$$F_x(\omega_{m,n}, k) = \frac{2\Delta T}{(n-m)\,T} \left| \sum_{t=0}^{T-1} x_t \sum_{f=m}^{n-1} e^{-2\pi i f \left(\frac{t}{T} - \frac{2\pi k}{n-m}\right)} \right|^2, \tag{5.23}$$

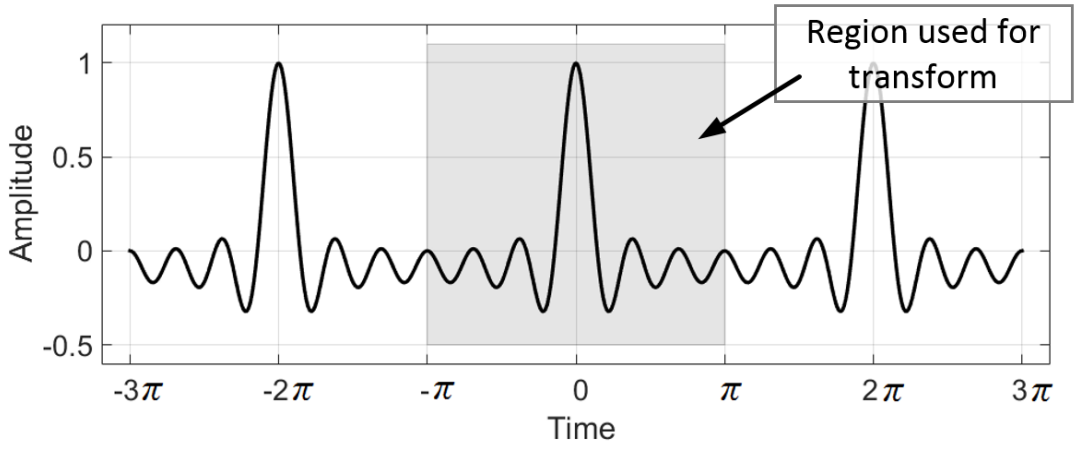


Figure 5.2: Discrete harmonic wavelet defined by Eq.5.20 for $k = 0$, $m = 1$ and $n = 6$

$\omega_{(m,n)}$ is the frequency band and k the time increments. Further, for the purpose of determining a 'probabilistic' power spectrum and following a similar approach as the one for the stationary process case, Eq.5.23 can be split into known and unknown cosine and sine components (see Eq.5.2 and Eq.5.3). As in the stationary case, t_α and t_β are used to represent the known and missing integer data point locations, respectively. Note that the assumption of ergodicity in the stationary case is no longer applicable in the non stationary case. So the usefulness of the following non-stationary result is limited in a practical sense, giving the PDF of the harmonic wavelet spectrum estimate for a single record only. Following the methodology for the stationary case, the result is that for the non-stationary case, Eq.5.23 takes a similar form to Eq.5.5, i.e.

$$S_X(\omega, t) = Y_1^2 + Y_2^2 \tag{5.24}$$

where

$$\begin{aligned} Y_1 = \sqrt{\frac{2\Delta T}{T(n-m)}} \Bigg(& \sum_{t_\alpha} x_{t_\alpha} \sum_{f=m}^{n-1} \cos\left(-2\pi f \left(\frac{t_\alpha}{T} - \frac{2\pi k}{n-m}\right)\right) \\ & + \sum_{t_\beta} x_{t_\beta} \sum_{f=m}^{n-1} \cos\left(-2\pi f \left(\frac{t_\beta}{T} - \frac{2\pi k}{n-m}\right)\right) \Bigg) \end{aligned} \tag{5.25}$$

and

$$\begin{aligned} Y_2 = \sqrt{\frac{2\Delta T}{T(n-m)}} \Bigg(& \sum_{t_\alpha} x_{t_\alpha} \sum_{f=m}^{n-1} \sin\left(-2\pi f \left(\frac{t_\alpha}{T} - \frac{2\pi k}{n-m}\right)\right) \\ & + \sum_{t_\beta} x_{t_\beta} \sum_{f=m}^{n-1} \sin\left(-2\pi f \left(\frac{t_\beta}{T} - \frac{2\pi k}{n-m}\right)\right) \Bigg) \end{aligned} \tag{5.26}$$

In Eqs.5.25,5.26, $Y_j = \text{Normal}\left(\mu_{Y_j}, \sigma^2_{Y,j}\right)$; $j = 1, 2$ where

$$\mu_{Y_j} \sqrt{\frac{2\Delta T}{T(n-m)}} \sum_{t_\alpha} x_{t_\alpha} \sum_{\omega=m}^{n-1} \cos\left(\omega \left(\frac{t_\alpha}{T} - \frac{k}{n-m}\right) - \varphi_j\right); \quad j = 1, 2; \quad \varphi_j = 0, \pi/2 \tag{5.27}$$

$$\sigma^2_{Y_j} = \frac{2\Delta T}{T\,(n-m)} \sum_{t_\beta} \left(\sum_{\omega=m}^{n-1} \cos\left(\omega \left(\frac{t_\beta}{T} - \frac{k}{n-m} \right) - \varphi_j \right) \right)^2 ; \quad j = 1, 2; \quad \varphi_j = 0, \pi/2 \tag{5.28}$$

$$\rho_{Y_1,Y_2} = \frac{2\Delta T}{T\,(n-m)} \cdot \frac{\sum_{t_\beta} \sum_{\omega=m}^{n-1} \cos\left(\omega \left(\frac{t_\beta}{T} - \frac{k}{n-m} \right) \right) \sum_{\omega=m}^{n-1} \sin\left(\omega \left(\frac{t_\beta}{T} - \frac{k}{n-m} \right) \right)}{\sigma_{Y_1} \sigma_{Y_2}} \tag{5.29}$$

The PDF for a non-stationary process power, from a single realization at a given time and frequency, may be drawn by evaluating Eq.5.16 with the above parameters. Similarly to the stationary case, if multiple process records are considered as part of an ensemble, the variance of the power spectrum estimate at each frequency decreases.

5.3 Numerical Examples

To demonstrate the applicability of the proposed "probabilistic" power spectrum estimation approach, stationary and non-stationary process time-histories are generated by Eq.2.62 and Eq.2.62 respectively. Both stationary and non-stationary examples feature earthquake processes with various arrangements of missing data.

5.3.1 Stationary power spectrum PDF

For the stationary case, sample realizations compatible with the stationary spectrum, Eq.2.65, are generated, where $\omega_g = 12$ rad/s, $\zeta = 0.6$ and $\alpha = 20$. Further, two samples are generated and analysed separately, with 10% and 20% missing data in uniformly distributed random locations. Figure 5.3 compares the original spectrum with the "probabilistic" spectrum estimate, based on Eq.5.16 after 10% of the data has been removed. Figure 5.4 shows distributions for three selected frequencies compared to estimates corresponding to no missing data (represented by vertical lines). It can be readily seen that the true power spectrum values for each frequency lie within the distributions' effective domains, and in this case within two standard deviations of the mean. Further, it is also noted that the spectral estimation uncertainty decreases as the true power spectrum value nears zero. This is partly due to the fact that power spectrum values cannot be negative, and so the PDFs become more skewed near zero. Higher true powers are also more drastically effected by missing data, due to the squared nature of the power spectrum, e.g. a fluctuation of ± 2 about zero when squared has an interval of width 4, but a fluctuation of ± 2 about 1 when squared has an interval of width 9.

Figures 5.5 and 5.4 show the same analysis for 20% missing data. As the number of missing data increases, the estimated power spectrum value exhibits larger uncertainty as anticipated; that is, the power spectrum PDF variance increases. However, the probability of the process having zero power at and around the predominant frequency, 12 rad/s, remains extremely low. This is because the known data still exhibit a strong correlation with this frequency.

To produce these PDF plots, Eq.5.16 is evaluated numerically using adaptive quadrature [94]. This procedure takes around 3 seconds on a standard modern mid-range computer (as of 2015). In comparison, applying a Monte-Carlo approach, 10^5, 10^6 and 10^7 samples took 2.5, 26 and 267 seconds, respectively. Clearly, a multiple of the above values would be required for the case where an ensemble of realizations were considered. The Monte-Carlo estimated PDFs are shown in Figure 5.7 (for an example estimation of the power at 18 rad/s with 10% missing data). Even with 10^7 samples, near the peak there is still a noticeable difference when compared to the target exact PDF.

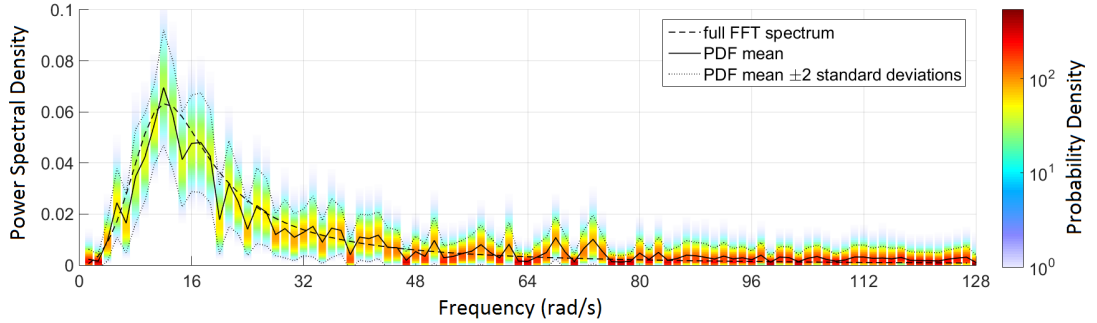


Figure 5.3: Power spectral probability densities with 10% missing data replaced by independent, identically distributed normal random variables

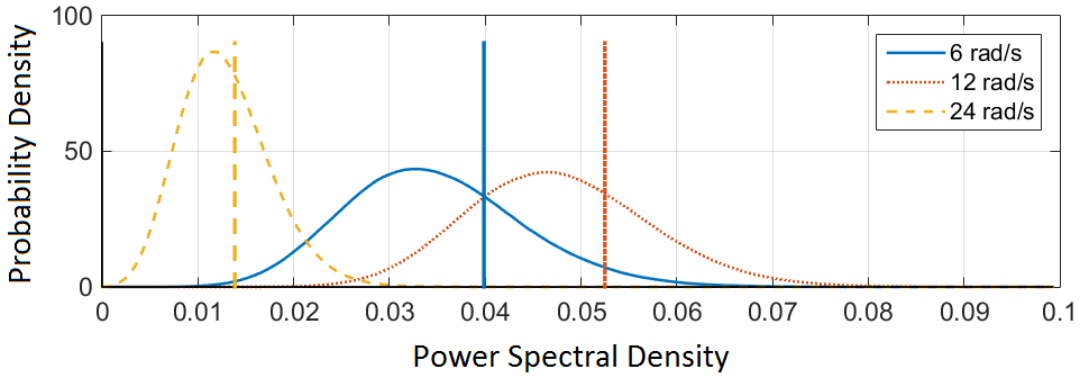


Figure 5.4: Selected PDFs from Figure 5.3 at 6 rad/s, 12 rad/s and 24 rad/s. The vertical lines show the estimated spectral power with no missing data for each PDF

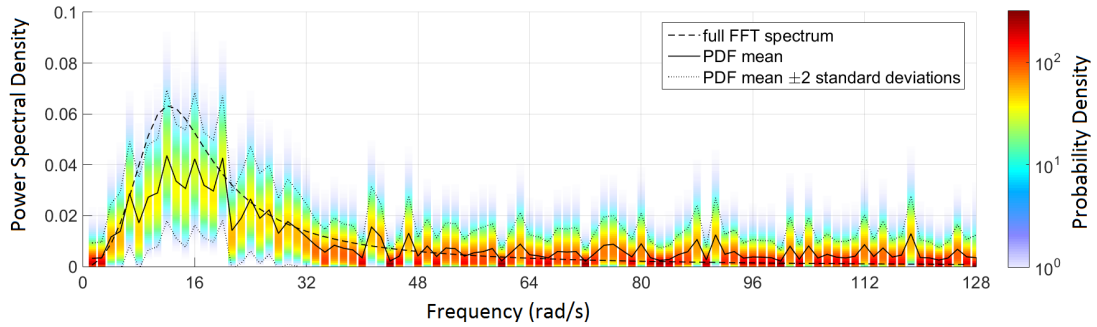


Figure 5.5: Power spectral probability densities with 20% missing data replaced by independent, identically distributed normal random variables

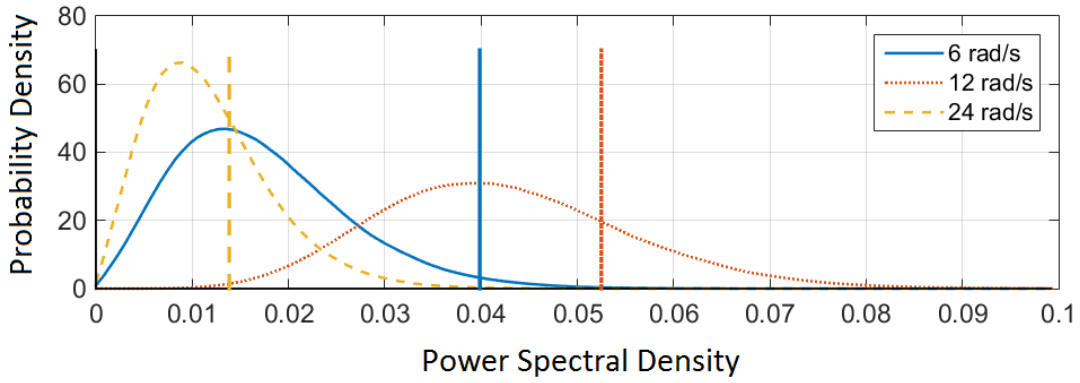


Figure 5.6: Selected PDFs from Figure 5.5 at 6 rad/s, 12 rad/s and 24 rad/s. The vertical lines show the estimated spectral power with no missing data for each PDF

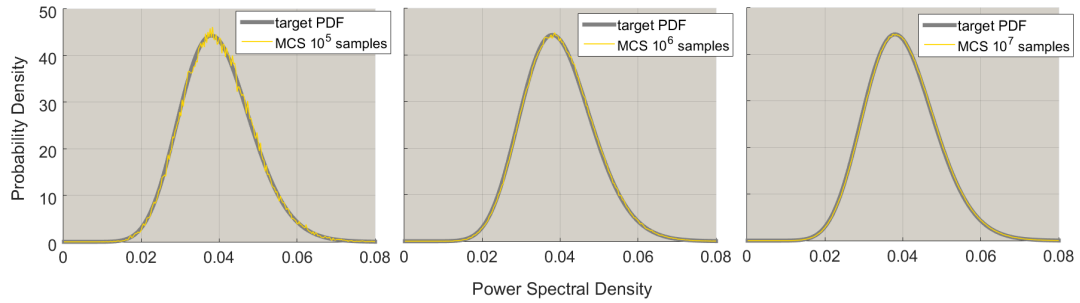


Figure 5.7: Comparison of Monte-Carlo simulated PDF and numerical integration solution for 10^5, 10^6 and 10^7 samples for 18 rad/s with 10% missing data

For the non-stationary case, the method is tested on the non-separable earthquake excitation process described in section 2.5.2 (Eq.2.69). The full spectrum is shown in Figure 5.8, and that for a single sample with no missing data is shown in Figure 5.9. A wavelet bandwidth of 6 rad/s is chosen to give an even trade-off between time and frequency resolution when using a 128 point sample. Two examples are shown initially, one with 10% missing data (Figure 5.10) and another with 20% missing data (Figure 5.11). An additional key is provided (Figure 5.13) to clarify the interpretation of these figures. Again, as with the stationary case, the probabilistic harmonic wavelet based power spectrum PDFs generally have higher mean values in the locations where the full estimated spectrum shows the greatest power spectral density. Similarly, the PDFs of the power spectral densities tend to have lower variance and greater skewness as their mean values approach zero. However, unlike in the stationary process examples, this tendency is not always the case. Because the wavelets are time-limited as well as frequency band-limited, power spectrum PDFs are significantly affected by the arrangement of the missing data in time. Notice in Figure 5.10 at $k = 1$ and $18 < \omega < 24 rad/s$, the estimated power spectrum value is comparatively high, yet the variance of the PDF is low. This is due to the fact that if we looked at the sample time history, there would be few missing data near the start of the record ($k = 1$). We can highlight this effect by purposely removing a continuous interval of data, rather than using uniformly distributed missing points. This is shown in Figure 5.12 where 10% of the record is missing near the start. The result is that bands of higher uncertainty occur, in this case for $k = 1, 2$ (i.e. the PDFs have higher variance). For higher k values, the PDFs have much lower variance, and the true estimated power spectrum is therefore less ambiguous. The effect is very clearly demonstrated by comparing selected

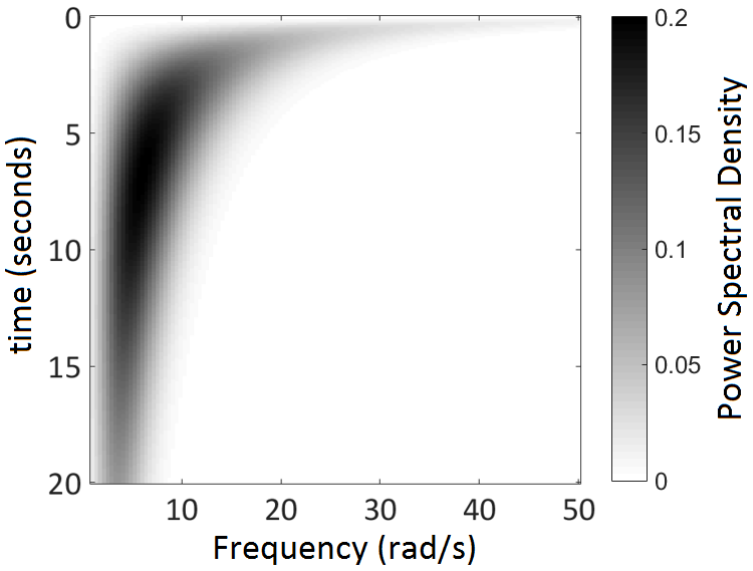


Figure 5.8: Non-separable earthquake power spectrum defined by Eq.2.69

PDFs at $k = 1$ where there are missing data, and at $k = 4$ where there are no missing data (Figures 5.14 and 5.15 respectively).

5.4 Chapter summary

The third novel approach, herein described in Chapter 5, was first presented during the 2nd International Conference on Vulnerability, Risk Analysis and Management (ICVRAM) in July 2014 [59]. A full journal paper featuring the evolutionary spectrum PDF submitted to the International Journal of Sustainable Materials and Structural Systems (IJSMSS) is currently under review. As a result of this work, the author is currently developing a non-probabilistic counterpart interval-analysis approach to power spectrum uncertainty quantification, which will serve in a similar manner as a complementary tool to deterministic estimation.

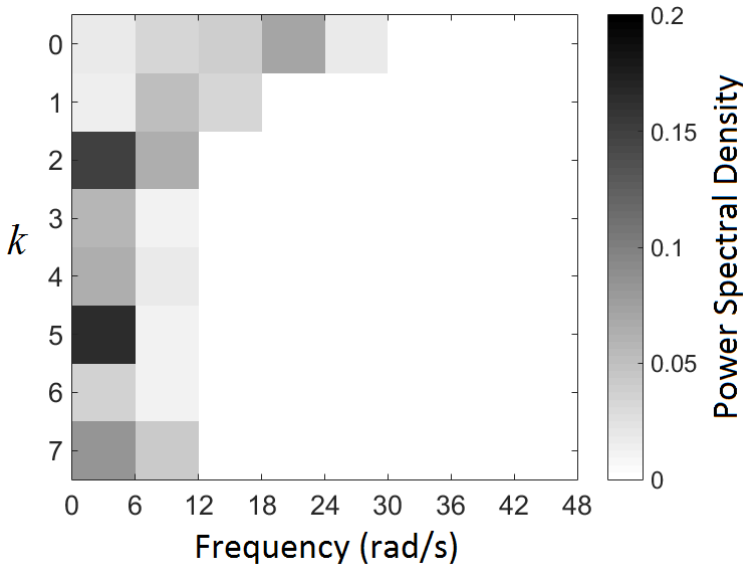


Figure 5.9: Harmonic wavelet power spectrum for single time history compatible with Eq.2.69

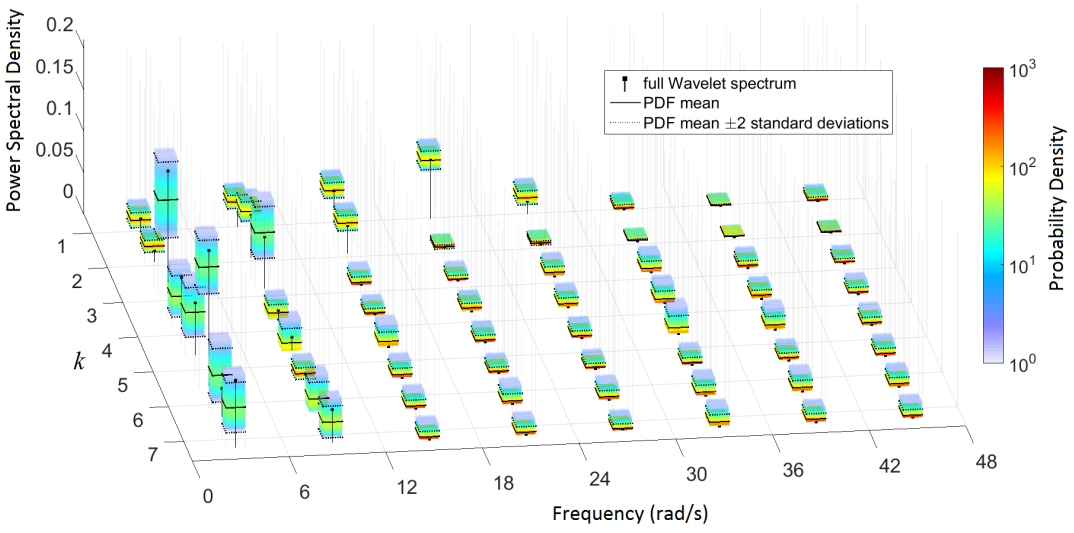


Figure 5.10: PDFs of power spectral density for the single realization of Eq.2.69 shown in Figure 5.9 with 10% missing data in uniformly distributed random locations

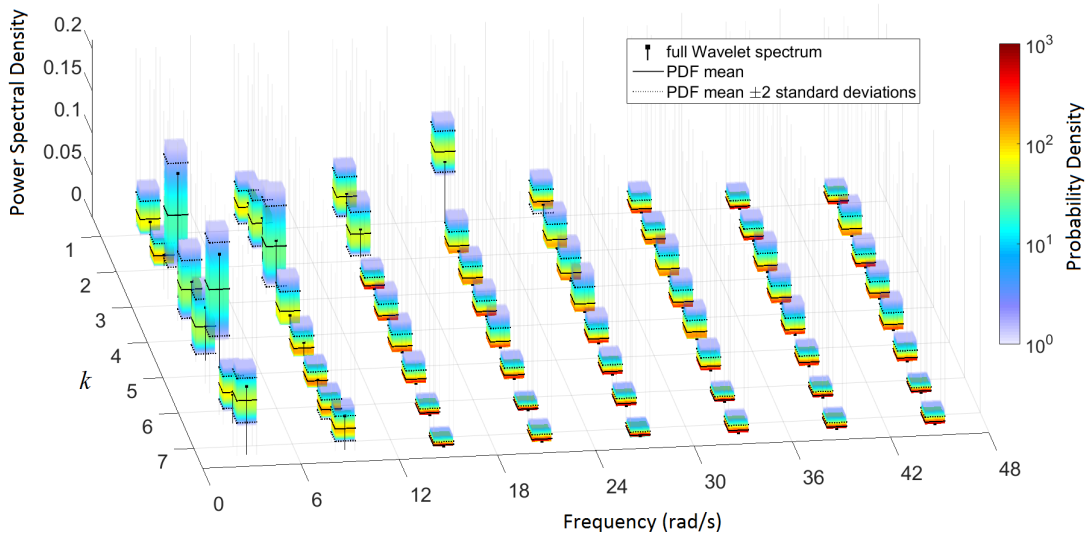


Figure 5.11: PDFs of power spectral density for the single realization of Eq.2.69 shown in Figure 5.9 with 20% missing data in uniformly distributed random locations

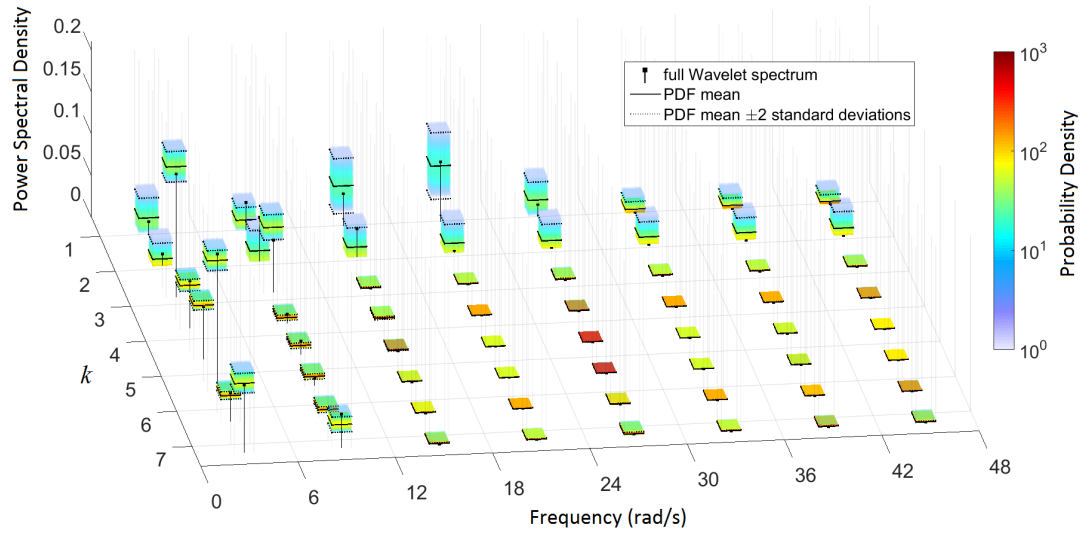


Figure 5.12: PDFs of power spectral density for a single realization of Eq.2.69 shown in Figure 5.9 with 10% missing data near the beginning of the record

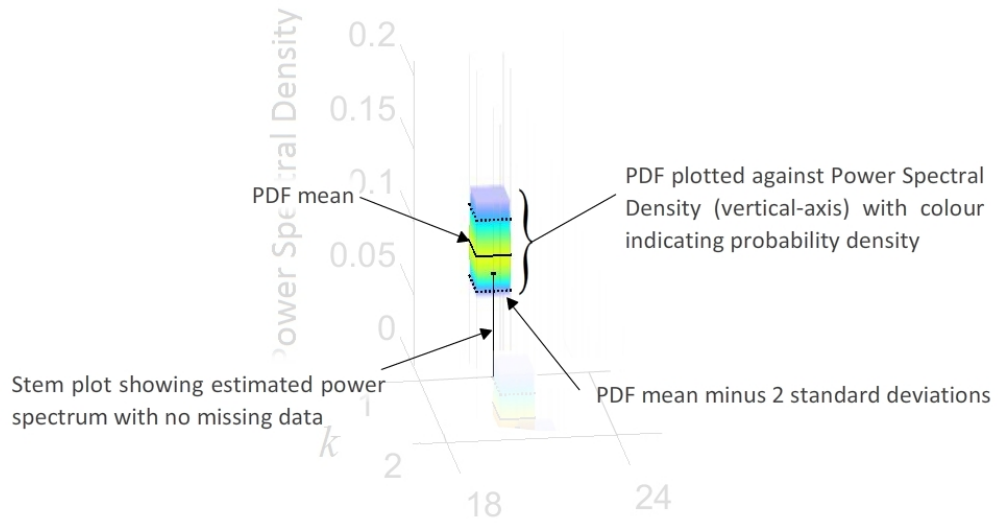


Figure 5.13: Additional explanation for interpreting Figures 5.10, 5.11 and 5.12

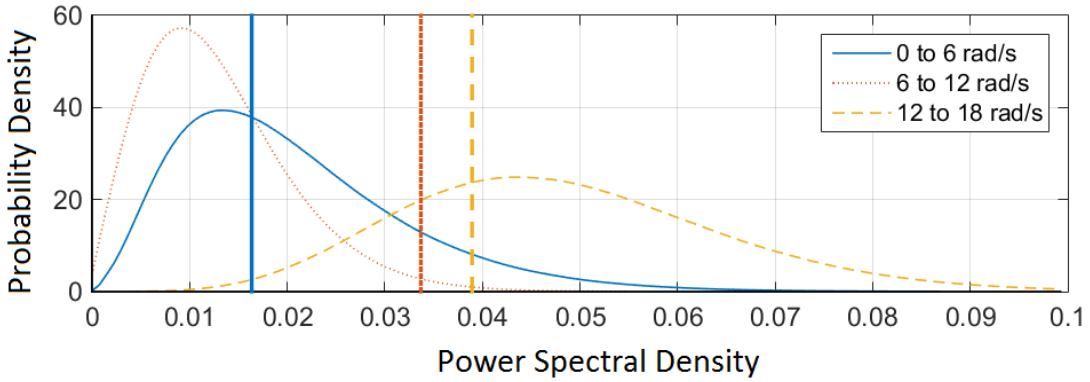


Figure 5.14: Selected PDFs from Figure 5.12 for $k = 1$ at 6 rad/s, 12 rad/s and 24 rad/s. The vertical lines show the estimated spectral power with no missing data for each PDF

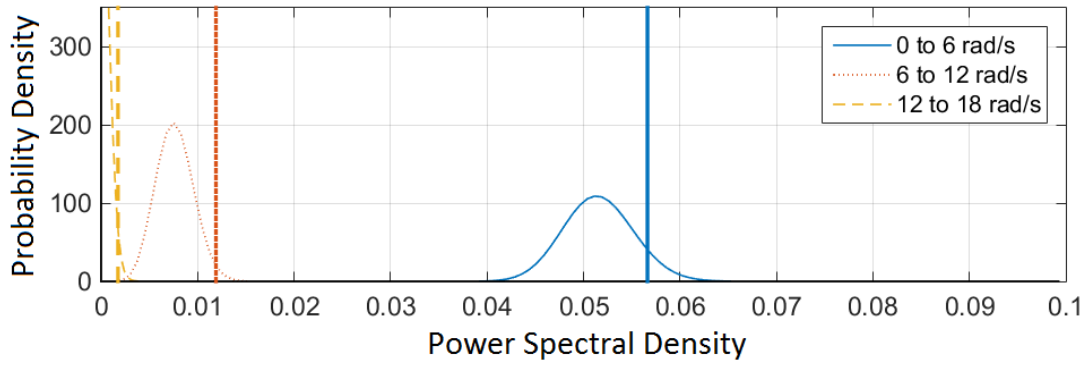


Figure 5.15: Selected PDFs from Figure 5.12 for $k = 4$ at 6 rad/s, 12 rad/s and 24 rad/s. The vertical lines show the estimated spectral power with no missing data for each PDF

Chapter 6

Conclusions and Recommendations

This work set out to investigate new approaches to problems associated with missing data in discrete analysis of stochastic processes. With a particular focus on evolutionary power spectrum estimation of non-stationary processes via harmonic wavelets, three novel methodologies, based on a critical appraisal of pertinent literature, were developed yielding very promising results. Concluding remarks pertaining to these three approaches, as well as the research investigation as a whole, are provided in this section, along with a number of suggestions for recommended areas of further research.

One of the primary features of the Artificial Neural Network (ANN) based approach outlined in Chapter 3 is that it can provide a relatively unrestrictive solution to the problem of missing data in spectral analysis and process simulation. ANNs are capable of adapting to suit the type of application without the need to define stringent model parameters, such as expected number of spectral peaks or total spectral power. In both stationary and non-stationary tests, an ANN was shown to be capable of learning a stochastic process and filling missing data gaps fitting with the original power spectrum. Not only was the network able to learn the processes, but a comparison of spectral analysis with the zero-filled, model-free approach showed the ANN to be superior, regardless of how the missing data was arranged. It should be noted that when applied to larger interval gaps than those demonstrated in this study, the network requires significant training to prevent the outputs from diverging, and this can be computationally demanding. Further, even in the examples shown here for which the ANN gave promising results, the training procedure was time consuming. This became more apparent when testing the proposed method in a structural reliability context [60] where large numbers of networks were trained on long processes. Fortunately the scheme is to a large extent parallelizable, but further research into alternative network architectures and learning algorithms could yield significant gains in efficiency.

Chapter 4 outlined an entirely different approach to stationary and non-stationary stochastic process power spectrum estimation subject to missing data, utilizing Compressive Sensing (CS) theory. Specifically, when applied to both stationary and non-stationary processes, CS theory, combined with an appropriate harmonic basis and applied in the presence of missing data, was shown to be highly effective in reconstructing the power spectrum in many cases. The approach appears efficient and superior to zero-padding based solutions, provided that the recorded time-domain data is relatively sparse in the frequency domain - an assumption that is often valid, especially when environmental process or structural response histories are considered. Although most effective when missing data are not grouped and randomly distributed, CS reconstruction was shown to be capable of providing accurate power spectrum estimates under a range of missing data configurations. The accuracy of non-separable estimated spectra is comparable to that of the separable spectrum results, suggesting that the developed approach can be efficacious in a range of problems in which little *a priori* information is known about the shape of the spectrum. Further, in cases where an ensemble of process records are used to form a single basis in which the process is sparse, applying an adaptive basis procedure in conjunction with CS was shown to

yield significant improvements over a standard CS approach. Of particular note is the extent to which the adaptive basis allows CS-reconstruction with harmonic wavelets, to reliably reproduce evolutionary power spectra under realistic conditions with up to 75% missing data. Recent work undertaken by the author to extend the adaptive basis CS procedure to higher dimensional data sets and cross-spectral density estimation has however found that in large problems, finding a solution to the L1 minimization problem can be computationally demanding. Optimizing this step should be a focus for further work in this area. Of additional interest is investigation into the addition of a probabilistic framework during the re-weighting procedure, where ensembles are large enough to consider higher order statistics than a simple average.

In Chapter 5, the issue of quantifying the uncertainty in stochastic process power spectrum estimates based on realizations with missing data was addressed. In this regard, relying on relatively relaxed assumptions for the missing data, probability density functions (PDFs) for power spectrum values corresponding to specific frequencies were produced. Specifically, modelling the missing data as Gaussian random variables, relying on fundamental concepts from probability theory, and resorting to Fourier and harmonic wavelets based representations for the stationary and non-stationary processes respectively, the uncertainty related to the power spectrum estimate was quantified by deriving a closed-form expression for the respective PDF. It is noted that the Gaussian assumption is not restrictive, and that the approach can be applied in a relatively straightforward manner for other alternative non-Gaussian PDFs chosen to model the missing data. Further, by virtue of the central limit theorem (CLT) the result is directly applicable for cases where the number of missing data is large enough to consider arbitrary PDFs as Gaussian. Also, the premise of performing a probabilistic spectral analysis is not limited to this approach, and there is significant scope for introducing additional assumptions to more accurately represent uncertainty under more specific conditions. Depending on the specific application, this work could be extended to include, for instance, time-varying distribution functions and correlated missing data. The results demonstrated the large extent to which any given single estimate, even for small amounts of missing data, may be unrepresentative of the target spectrum. Furthermore, considering the herein numerical examples with missing data $<10\%$ for both stationary and non-stationary cases, the true power spectrum values were all within two standard deviations of the PDF means. Therefore this approach could be used to bound deterministic estimates, providing specific validation criteria for missing data reconstruction. Finally, it is noted that the results were produced utilizing the exact closed-form expression at a significantly reduced computational cost, as opposed to an alternative Monte-Carlo simulation approach.

The methods outlined in Chapters 3, 4 and 5 are naturally not suitable in all scenarios involving spectral estimation under missing data, and are therefore meant as complementary methods to a much wider set of tools. Throughout this work, zero-padding in particular has been applied as a benchmark against which to demonstrate the superior performance of the proposed methods. Nevertheless, it is important to remember that zero-padding, along with other methods outlined in Chapter 2 can be highly applicable, producing reliable results with high efficiency in their own ideal scenarios. In this regard, a possible avenue of further work stemming from this research in its entirety could take the form of developing an unbiased, comprehensive list of tools for approaching missing data problems. By categorizing different methods based on qualitative and quantitative missing data criteria, a valuable resource for practical use in a wide range of applications could be produced, especially if combined with a user-friendly missing data 'toolbox'; having multiple uses in research, teaching and training.

To finalize, this doctoral work has generated a number of journal and conference papers, fostered collaboration with external research bodies, and will continue to be taken forward by the author and partners in multiple new directions in the future.

Bibliography

[1] M. Shinozuka, G. Deodatis, Simulation of stochastic processes by spectral representation, Applied Mechanics Reviews 44 (4) (1991) 191–204.

[2] M. Priestley, Evolutionary spectra and non-stationary processes, Journal of the Royal Statistical Society Series B-Statistical Methodology 27 (2) (1965) 204–237.

[3] J. Liang, S. R. Chaudhuri, M. Shinozuka, Simulation of nonstationary stochastic processes by spectral representation, Journal of Engineering Mechanics-Asce 133 (6) (2007) 616–627.

[4] B. Priestley, Spectral analysis and time series, no. v. 1-2, Academic Press, 1982, 80040235.

[5] P. Broersen, Automatic spectral analysis with missing data, Digital Signal Processing 16 (6) (2006) 754–766.

[6] P. Broersen, R. Bos, Time-series analysis if data are randomly missing, Ieee Transactions on Instrumentation and Measurement 55 (1) (2006) 79–84.

[7] N. Lomb, Least-squares frequency-analysis of unequally spaced data, Astrophysics and Space Science 39 (2) (1976) 447–462.

[8] J. Scargle, Studies in astronomical time-series analysis .2. statistical aspects of spectral-analysis of unevenly spaced data, Astrophysical Journal 263 (2) (1982) 835–853.

[9] P. Vanicek, Approximate spectral analysis by least-squares fit - successive spectral analysis, Astrophysics and Space Science 4 (4) (1969) 387–&.

[10] G. Fahlman, T. Ulrych, A new method for estimating the power spectrum of gapped data, Monthly Notices of the Royal Astronomical Society 199 (1) (1982) 53–65.

[11] J.-C. Hung, A genetic algorithm approach to the spectral estimation of time series with noise and missed observations, Information Sciences 178 (24) (2008) 4632–4643.

[12] S. Baisch, G. Bokelmann, Spectral analysis with incomplete time series: an example from seismology, Computers & Geosciences 25 (7) (1999) 739–750.

[13] D. Roberts, J. Lehar, J. Dreher, Time-series analysis with clean .1. derivation of a spectrum, Astronomical Journal 93 (4) (1987) 968–989.

[14] S. Qian, Introduction to Time-Frequency and Wavelet Transforms, 1st Edition, Prentice-Hall International, Upper Saddle River, N.J, 2002.

[15] S. Mallat, A Wavelet Tour of Signal Processing: The Sparse Way, Elsevier Science, 2008.

[16] F. D. Mix, K. J. Olejniczak, Elements of Wavelets for Engineers and Scientists, 1st Edition, John wiley & Sons, Hoboken, NJ, 2003.

[17] D. E. Newland, An introduction to random vibrations, spectral and wavelet analysis, 3rd Edition, Longmans Scientific & Technical, Harlow, 1993.

[18] D. E. Newland, Harmonic wavelet analysis, Proceedings of the Royal Society of London Series A-Mathematical Physical and Engineering Sciences 443 (1993) 203–225.

[19] P. D. Spanos, G. Failla, Wavelets: Theoretical concepts and vibrations related applications, The Shock and Vibration Digest 37 (2005) 359–375.

[20] P. D. Spanos, A. Giaralis, N. P. Politis, Time-frequency representation of earthquake accelerograms and inelastic structural response records using the adaptive chirplet decomposition and empirical mode decomposition, Soil Dynamics and Earthquake Engineering 27 (7) (2007) 675–689.

[21] L. Cohen, Time-Frequency Analysis, Prentice Hall, 1995, 94039843.

[22] R. A. Muller, G. J. MacDonald, Ice ages and astronomical causes : Data, Spectral Analysis and Mechanisms, 2nd Edition, Praxis Publishing, Chichester, UK, 2002.

[23] M. I. Knight, M. A. Nunes, G. P. Nason, Spectral estimation for locally stationary time series with missing observations, Statistics and Computing 22 (4) (2012) 877–895.

[24] H. Nyquist, Certain topics in telegraph transmission theory, American Institute of Electrical Engineers, Transactions of the 47 (2) (1928) 617–644.

[25] T. Kijewski, A. Kareem, On the presence of end effects and their melioration in wavelet-based analysis, Journal of Sound and Vibration 256 (5) (2002) 980–988.

[26] A. Haar, On the theory of orthogonal function systems, Math. Ann 69 (1910) 331–371.

[27] Z. Zhang, J. Moore, Intrinsic feature extraction in the coi of wavelet power spectra of climatic signals, in: Image and Signal Processing (CISP), 2011 4th International Congress on, Vol. 5, IEEE, 2011, pp. 2354–2356.

[28] I. Daubechies, Orthonormal bases of compactly supported wavelets, Communications on Pure and Applied Mathematics 41 (7) (1988) 909–996.

[29] R. R. Coifman, M. V. Wickerhauser, Entropy-based algorithms for best basis selection, Information Theory, IEEE Transactions on 38 (2) (1992) 713–718.

[30] D. E. Newland, Harmonic and musical wavelets, Proceedings of the Royal Society of London Series A-Mathematical Physical and Engineering Sciences 444 (1922) (1994) 605–620.

[31] D. E. Newland, Ridge and phase identification in the frequency analysis of transient signals by harmonic wavelets, Journal of Vibration and Acoustics-Transactions of the Asme 121 (2) (1999) 149–155.

[32] C. Torrence, G. P. Compo, A practical guide to wavelet analysis, Bulletin of the American Meteorological society 79 (1) (1998) 61–78.

[33] E. Wigner, On the quantum correction for thermodynamic equilibrium, Physical Review 40 (5) (1932) 0749–0759.

[34] J. d. Ville, et al., Théorie et applications de la notion de signal analytique, Cables et transmission 2 (1) (1948) 61–74.

[35] S. C. Bradford, Time-frequency analysis of systems with changing dynamic properties (2007).

[36] R. G. Stockwell, L. Mansinha, R. P. Lowe, Localization of the complex spectrum: The s transform, Ieee Transactions on Signal Processing 44 (4) (1996) 998–1001.

[37] J. B. Roberts, P. D. Spanos, Random vibration and statistical linearization, Courier Corporation, 2003.

[38] A. M. Yaglom, R. A. Silverman, An Introduction to the Theory of Stationary Random Functions, Dover Publications, 2004, 2003068803.

[39] H. Cramèr, R. Leadbetter, Stationary and related stochastic processes: sample function properties and their applications, Wiley, 1967.

[40] G. P. Nason, R. von Sachs, G. Kroisandt, Wavelet processes and adaptive estimation of the evolutionary wavelet spectrum, Journal of the Royal Statistical Society.Series B (Statistical Methodology) 62 (2) (2000) 271–292.

[41] R. Dahlhaus, Fitting time series models to nonstationary processes, The annals of Statistics 25 (1) (1997) 1–37.

[42] P. D. Spanos, I. A. Kougioumtzoglou, Harmonic wavelets based statistical linearization for response evolutionary power spectrum determination, Probabilistic Engineering Mechanics 27 (1) (2012) 57–68.

[43] P. Spanos, J. Tezcan, P. Tratskas, Stochastic processes evolutionary spectrum estimation via harmonic wavelets, Computer Methods in Applied Mechanics and Engineering 194 (12-16) (2005) 1367–1383.

[44] M. Beer, P. D. Spanos, A neural network approach for simulating stationary stochastic processes, Structural Engineering and Mechanics 32 (1) (2009) 71–94.

[45] K. Hasselmann, T. Barnett, E. Bouws, H. Carlson, D. Cartwright, K. Enke, J. Ewing, H. Gienapp, D. Hasselmann, P. Kruseman, et al., Measurements of wind-wave growth and swell decay during the joint north sea wave project (jonswap), Tech. rep., Deutches Hydrographisches Institut (1973).

[46] R. W. Clough, J. Penzien, Dynamics of structures, Tech. rep. (1975).

[47] K. Kanai, Semi-empirical formula for the seismic characteristics of the ground. university of tokyo. bulletin of earthquake research institute.

[48] H. Tajimi, A statistical method for determining the maximum response of a building structure during an earthquake. in: Proceedings of the 2nd world conference on earthquake engineering.

[49] F. Sabetta, A. Pugliese, Estimation of response spectra and simulation of nonstationary earthquake ground motions, Bulletin of the Seismological Society of America 86 (2) (1996) 337–352.

[50] D. Meisel, Fourier transforms of data sampled in unequally spaced segments, The Astronomical Journal 84 (1979) 116–126.

[51] R. Y. Rubinstein, Simulation and the Monte Carlo Method, John Wiley & Sons, 1981.

[52] H. J.A., Aperture synthesis with a non-regular distribution of interferometer baselines, Astronomy and Astrophysics, Supplement 15 (1974) 417, provided by the SAO/NASA Astrophysics Data System.

[53] S. Haykin, Neural networks : a comprehensive foundation, 2nd Edition, Prentice-Hall International, Upper Saddle River, N.J., 1999.

[54] E. Toth, A. Brath, A. Montanari, Comparison of short-term rainfall prediction models for real-time flood forecasting, Journal of Hydrology 239 (1) (2000) 132–147.

[55] J. C. Ochoa-Rivera, Prospecting droughts with stochastic artificial neural networks, Journal of hydrology 352 (1) (2008) 174–180.

[56] X. Yao, Evolving artificial neural networks, Proceedings of the IEEE 87 (9) (1999) 1423–1447.

[57] L. A. Comerford, I. A. Kougioumtzoglou, M. Beer, An artificial neural network based approach for power spectrum estimation and simulation of stochastic processes subject to missing data, in: Proceedings of the IEEE Symposium Series on Computational Intelligence, 2013.

[58] L. Comerford, I. Kougioumtzoglou, M. Beer, An artificial neural network based approach for power spectrum estimation subject to limited and/or missing data, in: Safety, Reliability, Risk and Life-Cycle Performance of Structures and Infrastructures, CRC Press, 2014, pp. 1083–1090–.

[59] L. Comerford, I. Kougioumtzoglou, M. Beer, An artificial neural network approach for stochastic process power spectrum estimation subject to missing data, Structural Safety 52 (2015) 150–160.

[60] L. Comerford, H. Jensen, M. Beer, C. Mayorga, I. Kougioumtzoglou, D. Kusanovic, Structural system response and reliability analysis under imcomplete earthquake records.

[61] M. A. Davenport, M. F. Duarte, Y. C. Eldar, G. Kutyniok, Introduction to Compressed Sensing, Compressed Sensing: Theory and Applications, Cambridge University Press, 2012.

[62] K. Sayood, Introduction to data compression, Newnes, 2012.

[63] M. Lustig, D. L. Donoho, J. M. Santos, J. M. Pauly, Compressed sensing mri, Signal Processing Magazine, IEEE 25 (2) (2008) 72–82.

[64] M. F. Duarte, G. Shen, A. Ortega, R. G. Baraniuk, Signal compression in wireless sensor networks, Philosophical Transactions of the Royal Society of London A: Mathematical, Physical and Engineering Sciences 370 (1958) (2012) 118–135.

[65] J. Y. Park, M. B. Wakin, A. C. Gilbert, Modal analysis with compressive measurements, Signal Processing, IEEE Transactions on 62 (7) (2014) 1655–1670.

[66] B. Tausiesakul, K. Gkoktsi, A. Giaralis, Compressive sensing spectral estimation for output-only structural system identification, in: Proc. 7th Int. Conf. Computational Stochastic Mechanics, 2014.

[67] S. OConnor, J. Lynch, A. Gilbert, Compressed sensing embedded in an operational wireless sensor network to achieve energy efficiency in long-term monitoring applications, Smart Materials and Structures 23 (8) (2014) 085014.

[68] I. Orovic, et al., Multimedia signals and systems, Springer Science & Business Media, 2012.

[69] E. J. Candès, J. K. Romberg, T. Tao, Stable signal recovery from incomplete and inaccurate measurements, Communications on pure and applied mathematics 59 (8) (2006) 1207–1223.

[70] A. Bandeira, E. Dobriban, D. Mixon, W. Sawin, Certifying the restricted isometry property is hard, Information Theory, IEEE Transactions on 59 (6) (2013) 3448–3450.

[71] M. Fornasier, H. Rauhut, Handbook of mathematical methods in imaging, chapter compressive sensing (2010).

[72] V. M. Patel, R. Chellappa, Sparse representations and compressive sensing for imaging and vision, Springer Science & Business Media, 2013.

[73] E. J. Candès, J. Romberg, T. Tao, Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information, Information Theory, IEEE Transactions on 52 (2) (2006) 489–509.

[74] British antarctic survey: high resolution radiosonde data from halley and rothera stations, [internet]. ncas british atmospheric data centre, 2008-2014.

[75] Y. Huang, J. L. Beck, S. Wu, H. Li, Robust bayesian compressive sensing for signals in structural health monitoring, Computer-Aided Civil and Infrastructure Engineering 29 (3) (2014) 160–179.

[76] Y. Yang, S. Nagarajaiah, Output-only modal identification by compressed sensing: Non-uniform low-rate random sampling, Mechanical Systems and Signal Processing 56 (2015) 15–34.

[77] D. L. Donoho, Compressed sensing, IEEE Transactions on Information Theory 52 (4) (2006) 1289–1306.

[78] S. S. Chen, D. L. Donoho, M. A. Saunders, Atomic decomposition by basis pursuit, SIAM journal on scientific computing 20 (1) (1998) 33–61.

[79] J. A. Tropp, A. C. Gilbert, Signal recovery from random measurements via orthogonal matching pursuit, Information Theory, IEEE Transactions on 53 (12) (2007) 4655–4666.

[80] E. J. Candes, M. B. Wakin, S. P. Boyd, Enhancing sparsity by reweighted ? 1 minimization, Journal of Fourier analysis and applications 14 (5-6) (2008) 877–905.

[81] L. Comerford, M. Beer, I. Kougioumtzoglou, et al., Compressive sensing based power spectrum estimation from incomplete records by utilizing an adaptive basis, in: Computational Intelligence for Engineering Solutions (CIES), 2014 IEEE Symposium on, IEEE, 2014, pp. 117–124.

[82] L. Comerford, M. Beer, I. Kougioumtzoglou, A compressive sensing based approach for evolutionary power spectrum estimation subject to missing data, in: Proceedings of the 7th International Conference on Stochastic Mechanics (CSM), 2014.

[83] L. Comerford, M. Beer, I. Kougioumtzoglou, A compressive sensing based approach for estimating stochastic process power spectra subject to missing data, in: Proceedings of the 9th International Conference on Structural Dynamics (EURODYN), Reliability and robustness of dynamic systems, 2014.

[84] J. D. Hays, J. Imbrie, N. J. Shackleton, et al., Variations in the earth's orbit: pacemaker of the ice ages, American Association for the Advancement of Science, 1976.

[85] G. J. MacDonald, Spectral analysis of time series generated by nonlinear processes, Reviews of Geophysics 27 (4) (1989) 449–469.

[86] P. Billingsley, Probability and measure, John Wiley & Sons, 2008.

[87] Y. Zhang, L. Comerford, I. Kougioumtzoglou, E. Patelli, M. Beer, Spectral moments estimates uncertainty quantification under incomplete data (currently under review), in: Proceedings of the 6th Asian-Pacific Symposium on Structural Reliability and its Applications, 2016.

[88] K. Sobczyk, Information dynamics: premises, challenges and results, Mechanical Systems and Signal Processing 15 (3) (2001) 475–498.

[89] K. Sobczyk, J. Tr?bicki, Maximum entropy principle and nonlinear stochastic oscillators, Physica A: Statistical Mechanics and its Applications 193 (3) (1993) 448–468.

[90] J. W. Lindeberg, Eine neue herleitung des exponentialgesetzes in der wahrscheinlichkeitsrechnung, Mathematische Zeitschrift 15 (1) (1922) 211–225.

[91] R. B. Ash, C. Doleans-Dade, Probability and measure theory, Academic Press, 2000.

[92] S. J. Press, Linear combinations of non-central chi-square variates, The Annals of Mathematical Statistics (1966) 480–487.

[93] H. Stark, J. W. Woods, Probability and Random Processes With Applications to Signal Processing, Prentice Hall PTR, 2002.

[94] J. N. Lyness, Notes on the adaptive simpson quadrature routine, Journal of the ACM (JACM) 16 (3) (1969) 483–495.