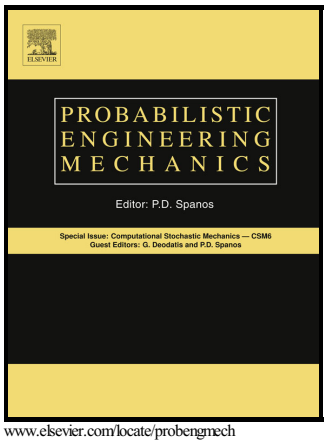
Compressive sensing based stochastic process power spectrum estimation subject to missing data

Liam Comerford, Ioannis A. Kougioumtzoglou, Michael Beer



www.elsevier.com/locate/probengmech

PII: S0266-8920(15)30043-6
DOI: http://dx.doi.org/10.1016/j.probengmech.2015.09.015
Reference: PREM2860

To appear in: *Probabilistic Engineering Mechanics*

Received date: 17 August 2015
Accepted date: 29 September 2015

Cite this article as: Liam Comerford, Ioannis A. Kougioumtzoglou and Michael Beer, Compressive sensing based stochastic process power spectrum estimation subject to missing data, *Probabilistic Engineering Mechanics,* http://dx.doi.org/10.1016/j.probengmech.2015.09.015

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

Liam Comerford[a], Ioannis A. Kougioumtzoglou[b], Michael Beer[a]

[a]*Institute for Risk and Uncertainty, University of Liverpool, Liverpool, L69 3GH, UK*
[b]*Department of Civil Engineering and Engineering Mechanics, Columbia University, New York, NY 10027, USA*

Abstract

A compressive sensing (CS) based approach for stationary and non-stationary stochastic process power spectrum estimation subject to missing data is developed. Stochastic process records such as wind and sea wave excitations can often be represented with relative sparsity in the frequency domain. Relying on this attribute, a CS framework can be applied to reconstruct a signal that contains sampling gaps in the time domain, possibly occurring for reasons such as sensor failures, data corruption, limited bandwidth/storage capacity, and power outages. Specifically, first an appropriate basis is selected for expanding the signal recorded in the time domain. In this regard, Fourier and harmonic wavelet bases are utilized herein. Next, an $L1$ norm minimization procedure is performed for obtaining the sparsest representation of the signal in the selected basis. Finally, the signal can either be reconstructed in the time domain if required or, alternatively, the underlying stochastic process power spectrum can be estimated in a direct manner by utilizing the determined expansion coefficients; thus, circumventing the computational cost related to reconstructing the signal in the time domain. The technique is shown to estimate successfully the essential features of the stochastic process power spectrum, while it appears to be efficient even in cases with 65% missing data demonstrating superior performance in comparison with alternative existing techniques. A significant advantage of the approach is that it performs satisfactorily even in the presence of noise. Several numerical

Email addresses: l.comerford@liv.ac.uk (Liam Comerford), iak2115@columbia.edu (Ioannis A. Kougioumtzoglou), m.beer@liv.ac.uk (Michael Beer)

examples demonstrate the versatility and reliability of the approach both for stationary and non-stationary cases.

Keywords: Compressive Sensing, Stochastic Process, Missing Data

1. Introduction

Acquired data corresponding, for instance, to environmental processes, are often pivotal for defining and calibrating probabilistic engineering load models to be used in subsequent analyses of critical engineering systems. In utilizing these data, power spectrum estimation can be an invaluable tool and an important building block in engineering systems analyses, especially within a Monte Carlo simulation framework, e.g.[1]. Nevertheless, in the presence of missing data, there are certain limitations to standard spectral analysis techniques such as those based on Fourier transform. Missing data in this context refer to a stochastic process time-history record, which for some reason has been sampled irregularly or lost some of its original content. There are numerous situations in which missing data may be unavoidable. These include sensor failures, data loss or corruption, as well as limited allocated time with shared equipment. In these situations it may be infeasibly expensive or logistically impossible to re-record the process in full, and therefore alternative analysis techniques are required to best analyse and process the available data.

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 [2, 3]. Further, least-squares sinusoid fitting and zero-padded gaps [4, 5, 6] offer efficient solutions for re-constructing 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 [7, 8] or assume that the underlying process comprises a highly limited number of significant harmonic components [9, 10].

Recently, an artificial neural network based approach was developed by the authors for power spectrum estimation and simulation of stochastic processes subject to missing data [11]. A significant advantage of the approach relates to the fact that no prior knowledge of statistics of the underlying process is required.

Note that for the non-stationary case, additional challenges arise when estimating the evolutionary power spectrum. In this regard, alternatives to stationary methods (e.g. the Fourier transform) can be utilized, such as wavelet [12, 13, 14, 15, 16] or chirplet (e.g. [17]) transforms, the short-time Fourier transform, Gabor transform (e.g. [12]) and Wigner-Ville distribution (e.g. [18, 19, 20]). However, the majority of approaches for addressing missing data in the stationary case are not directly applicable in the non-stationary case or impose the assumption that the process is locally stationary (e.g. [21]).

The approach to dealing with missing data in stationary and non-stationary processes developed in this paper relies on the fact that many environmental processes such as earthquakes, sea waves, winds and tidal patterns can be characterized by a relatively small number of dominant frequencies when expanded in the frequency domain. This feature leads to considering compressive sensing (CS) [22, 23] as a promising tool for signal reconstruction and spectral estimation both of stationary and non-stationary stochastic processes. In this regard, the capabilities of the recently developed CS framework are exploited herein for addressing the problem of estimating stationary and non-stationary stochastic process power spectra in cases where the available realizations exhibit missing data. It is shown that in conjunction with an appropriately selected basis, power spectra are satisfactorily estimated in the presence of large (up to 65%) amounts of missing data.

2. Stochastic process representation and spectral estimation

In this section, a concise review and related details on stationary and non-stationary stochastic process representation are included for the full, uniform time-history case. Specifically, Fourier and recently developed harmonic wavelet based power spectrum estimation approaches are delineated, providing a basis for the CS approach.

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

(e.g. [24, 25, 26])

$$X(t) = \int_0^{\infty} e^{i\omega t} dZ(\omega) \tag{1}$$

where $Z(\omega)$ has the properties

$$E\left(\left|dZ^2(\omega)\right|\right) = 4S_X(\omega)\, d\omega, \tag{2}$$

and

$$E(dZ(\omega)) = 0. \tag{3}$$

In Eq.(2), $S_X(\omega)$ 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.(1) [1] is given by

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

where Φ_k are uniformly distributed random phase angles in the range $0 \leq \Phi_k < 2\pi$. Furthermore, regarding estimation of the power spectrum of the process of Eq.(1) based on available realizations, this is given by the ensemble average of the square of the absolute Fourier transform amplitudes of the realizations. In this context, standard established Fast Fourier Transform algorithms can be utilized (e.g. [27]).

Next, for the case of non-stationary stochastic processes, a similar to Eq.(1) rigorous process representation of non-stationary stochastic processes needs to be employed (see also [28]). In this regard, a framework was developed in [29] for representing non-stationary stochastic processes by utilizing a time/frequency-localized wavelet basis as opposed to the Fourier decomposition of Eq.(1); the representation reads

$$X(t) = \sum_j \sum_k w_{j,k} \psi_{j,k}(t)\, \xi_{j,k}, \tag{5}$$

where $\xi_{j,k}$ is a stochastic orthonormal increment sequence; $\psi_{j,k}(t)$ is the chosen family of wavelets and j and k represent the different scales and translation levels, respectively. Further, the local contribution to the variance of the process of Eq.(5) is given by the term $|w_{j,k}|^2$. The wavelet-based model of Eq.(5) relies on the theory of locally stationary processes (see also [30]). 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.[30, 31, 32]).

Focusing next on generalized harmonic wavelets [33], they have a box-shaped frequency spectrum, whereas a wavelet of (m,n) scale and (k) position in time attains a representation in the frequency domain of the form

$$\Psi^{G}_{(m,n),k}(\omega)=\begin{cases}\frac{1}{(n-m)\Delta\omega}\exp\left(-i\omega\frac{kT_0}{n-m}\right), & m\Delta\omega\leq\omega\leq n\Delta\omega\\ 0 & otherwise\end{cases}, \tag{6}$$

where m, n and k are considered to be positive integers and $\Delta\omega=\frac{2\pi}{T_0}$; and T_0 is the total time duration of the signal under consideration. A collection of harmonic wavelets of the form of Eq.(6) spans adjacent non-overlapping intervals at different scales along the frequency axis. The inverse Fourier transform of Eq.(6) gives the time-domain representation of the wavelet which is equal to

$$\Psi^{G}_{(m-n),k}(t)=\frac{\exp\left(in\Delta\omega\left(t-\frac{kT_0}{n-m}\right)\right)-\exp\left(im\Delta\omega\left(t-\frac{kT_0}{n-m}\right)\right)}{i\left(n-m\right)\Delta\omega\left(t-\frac{kT_0}{n-m}\right)} \tag{7}$$

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

$$W^{G}_{(m,n),k}=\frac{n-m}{kT_0}\int_{-\infty}^{\infty}f(t)\,\overline{\psi^{G}_{(m,n),k}(t)}dt, \tag{8}$$

and projects the function $f(t)$ on this wavelet basis. Next, utilizing the generalized harmonic wavelets, Eq.(5) becomes (see [34])

$$X(t)=\sum_{(m,n)}\sum_{k}\left(X_{(m,n),k}(t)\right), \tag{9}$$

where

$$X_{(m,n),k}(t)=\sqrt{S_{(m,n),k}\left(n-m\right)\Delta\omega}\psi_{(m,n),k}(t)\,\xi_{(m,n),k}. \tag{10}$$

Eq.(10) 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 $S_{(m,n),k}$ represents the EPS $S_X(\omega,t)$ at scale (m,n) and translation (k). Further, Eq.(10) can be written in the form (see [34])

$$X_{(m,n),k}(t)=\int_{m\Delta\omega}^{n\Delta\omega}e^{i\omega\left(t-\frac{kT_0}{n-m}\right)}dZ_{(m,n),k}(\omega), \tag{11}$$

with the properties

$$E\left(dZ_{(m,n),k}\left(\omega\right)\right)=0, \tag{12}$$

and

$$E\left(\left|dZ_{(m,n),k}\left(\omega\right)\right|^{2}\right)=S_{(m,n),k}\left(n-m\right)\Delta\omega. \tag{13}$$

Furthermore, it has been shown that realizations compatible with $S_X\left(\omega,t\right)$ can be generated by utilizing a generalization of Eq.(4) of the form (see [35])

$$X\left(t\right)=\sum_{j=0}^{N-1}\sqrt{4S_X\left(\omega_j,t\right)\Delta\omega}\sin\left(\omega_j t+\Phi_j\right). \tag{14}$$

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 $S_X\left(\omega,t\right)$ of the process $X\left(t\right)$ is estimated by (e.g. [16, 34])

$$S_X\left(\omega,t\right)=S^{X}_{(m,n),k}=\frac{E\left(\left|W^{G}_{(m,n),k}\left[X\right]\right|^{2}\right)}{\left(n-m\right)\Delta\omega}, \tag{15}$$

where $m\Delta\omega\leq\omega\leq n\Delta\omega$, $\frac{kT_0}{n-m}\leq t\leq\frac{(k+1)T_0}{n-m}$, $S^{X}_{(m,n),k}$ represents the EPS of the process $X\left(t\right)$, 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.

In the ensuing analysis and specifically in the numerical examples section, stationary and non-stationary process realizations are generated by utilizing Eq.(4) and Eq.(14) respectively. Further, power spectrum estimates for stationary processes are determined by calculating the ensemble average of the square of the absolute Fourier transform amplitudes of the realizations. In the case of non-stationary processes Eq.(15) is employed.

3. 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 the half the signal's sampling rate [36]. 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 [37].

When data is captured, it is often convenient to expand it into a new basis. In the case of lossy compression techniques, bases or frames (e.g. redundant dictionaries) 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 [36].

3.1. Signal sparsity and incoherence

For robust compressive sensing there are several important properties to be considered, 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. Further, 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). As an example, Figure 1 (dotted line) shows the following simple discrete time waveform sampled 256 times,

$$y(t) = 2\sin(12t+1) + \sin(35t+2) + 1.5\sin(120t+3), \qquad (16)$$

where $0 \leq t < 2\pi$. The function of Eq.16 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 3 (resulting in 3 peaks on the plot). In the frequency domain, Eq.16 is clearly sparse as the majority of the data is equal to zero. Assuming now that Eq.16 was a real signal being captured

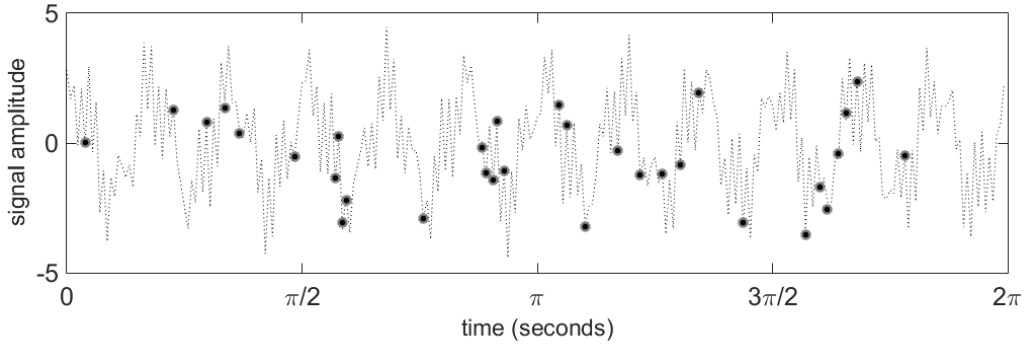


Figure 1: Graphical output of Eq.16 with randomly selected points to sample

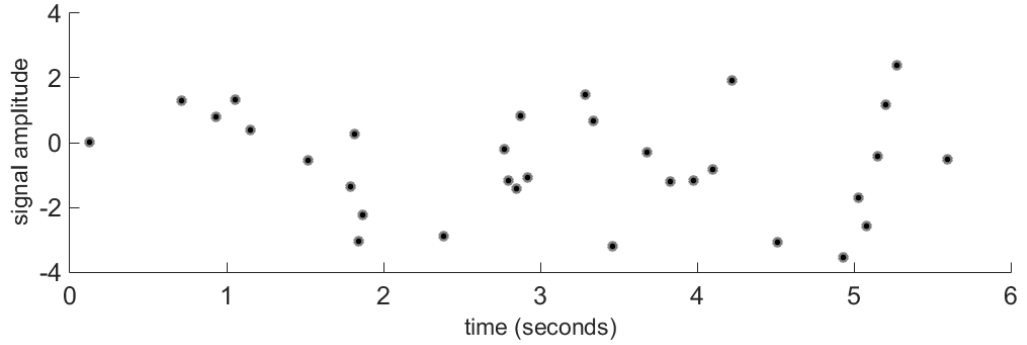


Figure 2: Randomly sampled points without original signal

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 2 shows the same signal as in Figure 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 3) and the signal may be reconstructed in the time domain as shown in Figure 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.

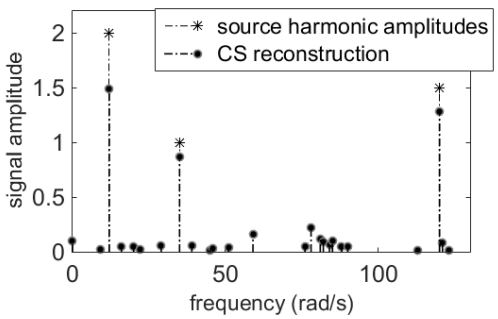


Figure 3: Eq.16 represented in the frequency domain as a sparse signal with CS estimation of frequency domain coefficients from Figure 1

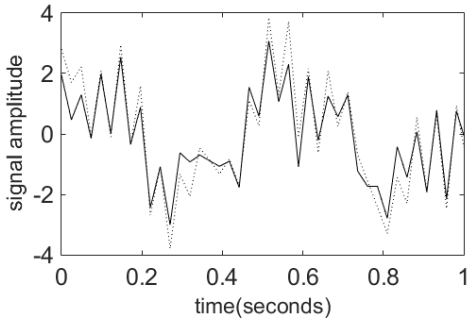


Figure 4: Comparison of first second of Figure 1 against the CS estimation in the time domain from Figure 2

3.2. Restricted Isometry Property

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

$$(1-\delta_K)\left|\tilde{x}\right|_{l2}^2 \leq \left|A\tilde{x}\right|_{l2}^2 \leq (1+\delta_K)\left|\tilde{x}\right|_{l2}^2, \tag{17}$$

for every vector $\tilde{x}$ with at least K non-zero entries [38]. 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 [39]. 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

$$m \geq CK\log\left(\frac{N}{M}\right), \tag{18}$$

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. [40]).

3.3. Sparse solutions 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,

$$y = Ax \tag{19}$$

where x is the measurement vector assumed to be sparse. Eq.19 represents an under-determined system with infinite solutions. This problem may be solved easily under the constraint that x must be minimized in the least squares sense, i.e.

$$\min |x|_{l2} = A^T \left(AA^T\right)^{-1} y. \tag{20}$$

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.20) does not lead to a sparse solution. The sparsest solution of Eq.19 occurs when the $L0$ norm is minimized, often referred to as a pseudo-norm [41] and defined as

$$|x|_{l0} = \begin{cases} 1, & x > 0 \ or \ x < 0 \\ 2, & x = 0 \end{cases} \tag{21}$$

This optimization problem is non-convex with no known exact solution [23]. However, a viable alternative exists in minimizing the $L1$ norm instead. $L1$ norm minimization promotes sparsity and will often yield the same result as $L0$ norm minimization in many cases [41]. Further, the problem becomes convex and may be set in a convenient linear programming form, i.e.

$$\min |x|_{l1} \ \ subject \ to \ y = Ax \tag{22}$$

Eq.22 describes a basis pursuit optimization problem and can be easily solved via a gradient-based optimization method, e.g. [42]. Figure 5 shows how $L1$ minimization gives sparse solutions by comparing both $L1$ and $L2$ 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 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 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

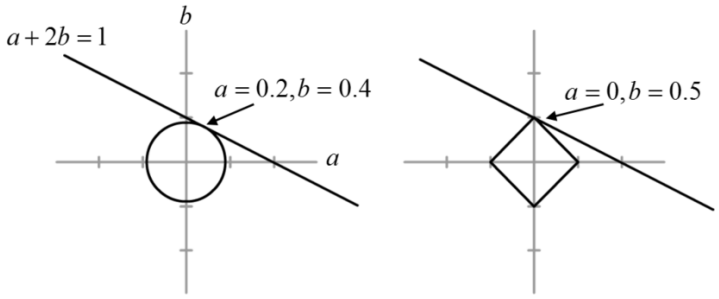


Figure 5: Minimum $L2$ and $L1$ solutions to the equation, $a + 2b = 1$

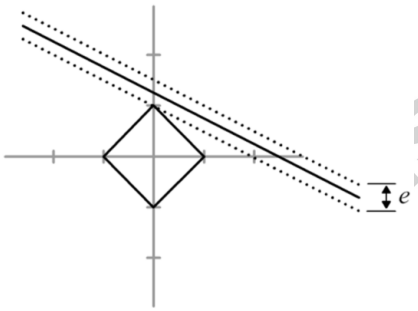


Figure 6: Minimum $L1$ solution to the equation, $a + 2b = 1$ with tolerance, e for noise vector, z

coefficients across most bases. With a small modification to Eq.22 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. For a noisy signal,

$$y = Ax + z, \tag{23}$$

where z is some noise vector. Given a tolerance, e, relative to the variance of the noise, Eq.22 may be re-cast in the form,

$$\min |x|_{l1} \;\; subject\ to \;\; |Ax - y|_{l2} \leq e. \tag{24}$$

This modification has the effect of applying intervals to the solutions (Figure 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 7. Figure 7a shows radiosonde wind speed data recorded from a weather balloon launched from Halley Antarctic Research Station in January, 2014 [43]. This data has a very sparse representation in the GHW domain and is therefore ideal for CS. With 50% of the data removed at random locations Figures 7c and 7d show $L1$ and $L2$ wavelet reconstructions respectively.

3.4. Compressive sensing for 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. [44, 45]). 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. Nevertheless, applying compressive sensing theory to the problem of missing data differs primarily in one respect; i.e., missing data are not necessarily intentional. Unfortunately this removes control over one important step of CS: the arrangement of the sampling matrix. CS relies on the choice of an appropriate sampling matrix. Uniform random Fourier matrices obey the RIP with high probability when data are sparse ([22, 23]); similarly, random GHW 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 GHW matrices, regular or large gaps of missing data leads to lower orthogonality between random columns of the sampling matrix. The result is that greater numbers of measurements are required

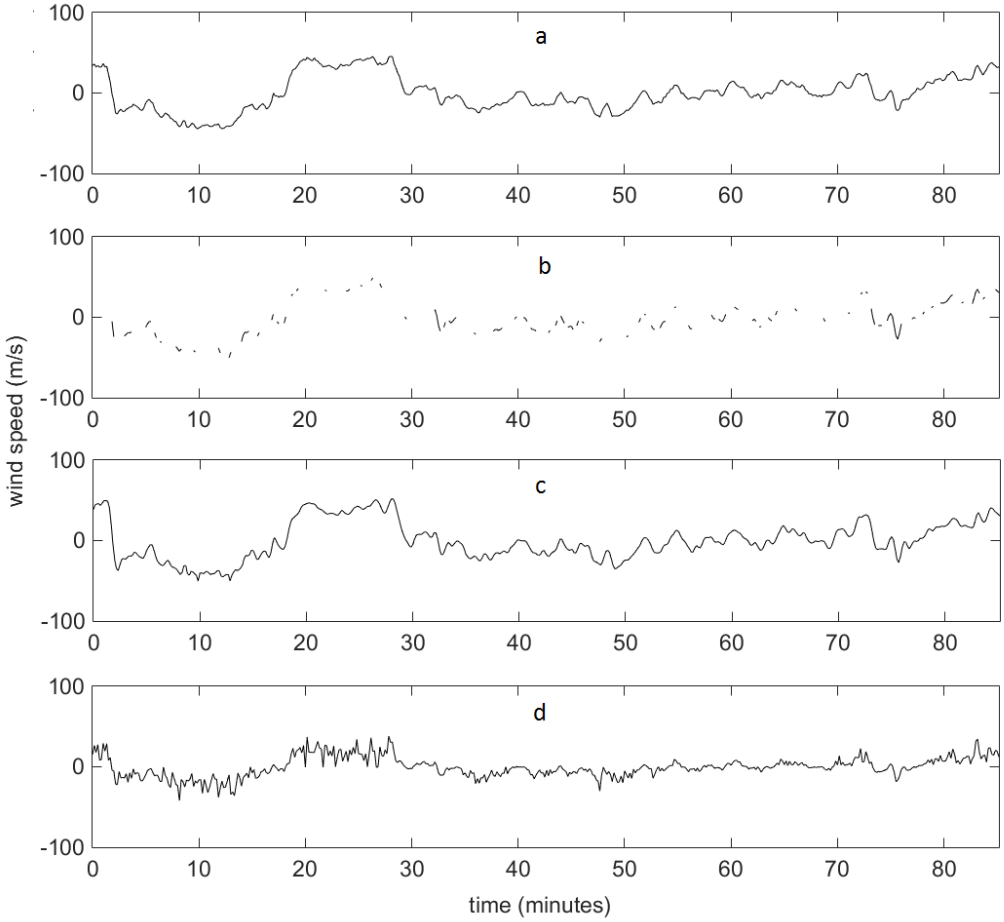


Figure 7: (a) Full windspeed record. (b) Full record with missing data. (c,d) Records with 50% missing data in random locations, down-sampled and reconstructed via $L1$ and $L2$ minimization of harmonic wavelets respectively. Data provided by [43]

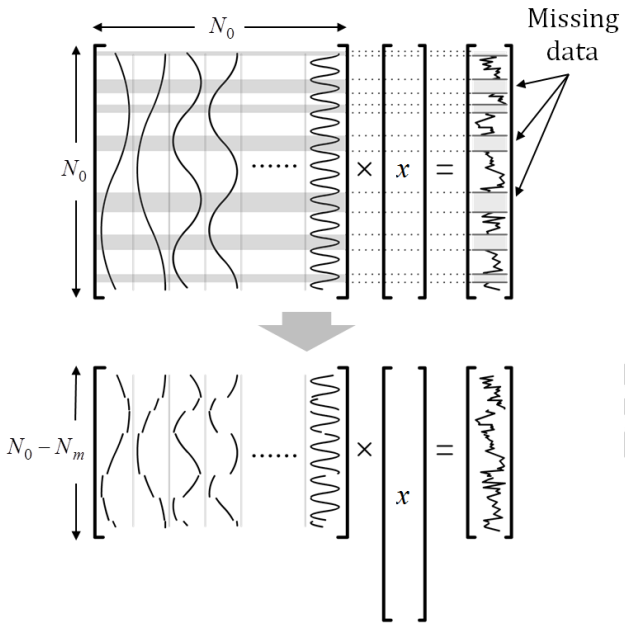


Figure 8: Fourier sampling matrix construction with missing data

for reliable reconstruction. Despite these problems, CS reconstruction based on the assumption of sparsity may still be 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 key frequencies.

3.5. Basis matrix construction

3.5.1. Stationary case

For stationary stochastic processes, represented by Eq.1, the power spectrum is estimated based on the mean square value of the Fourier transform over an ensemble of time-histories, e.g. [15]. 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 8.

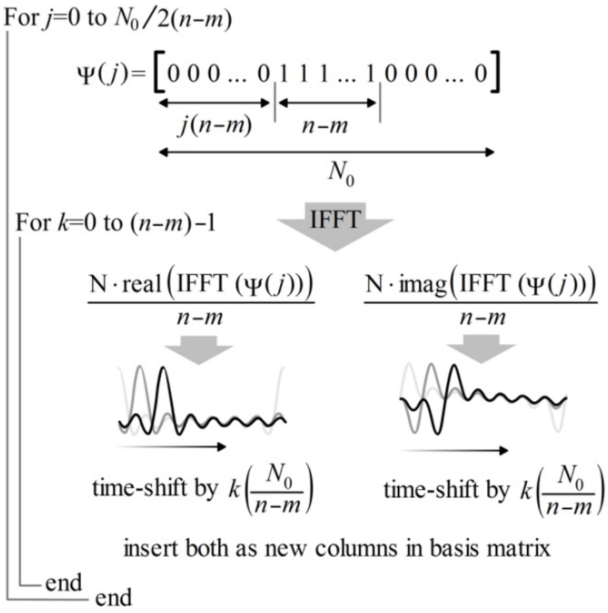


Figure 9: Harmonic wavelet basis construction using IFFT and nested for-loops

3.5.2. 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.6. 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. The process used to build this harmonic wavelet basis matrix is depicted in Figure 9. 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 10).

3.5.3. Cases with uneven sampling

The previously described methods for forming stationary and non-stationary basis matrices assume that the original data has been regularly sampled. This

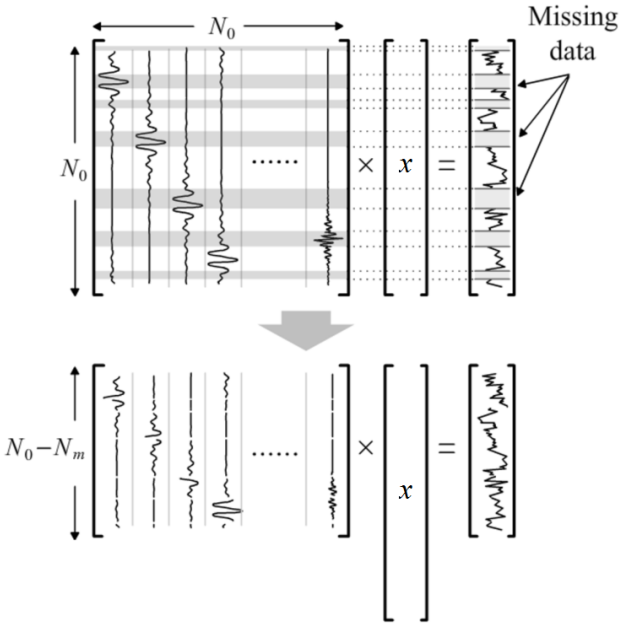


Figure 10: Harmonic wavelet sampling matrix construction with missing data

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.7.

With the basis formed, the CS reconstruction may be solved via an appropriate minimization algorithm. Examples include linear programming basis pursuit [46, 23] and greedy algorithms [47].

4. Numerical examples

Examples are provided for stationary and non-stationary processes, utilizing Fourier and GHW bases, respectively. 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). Scaled zero padding is a common spectral reconstruction technique when dealing with missing data (e.g. [48]). The result is the same as that of a least squares solution but the resulting spectrum is scaled up relative to the amount of missing data in the signal. It is shown that even with this additional scaling, estimates often fall short of the original target spectrum.

4.1. Simulation of missing data

In the ensuing examples, 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,

$$f_0(t) = \begin{cases} f(t), & r_a(t) \geq m \\ 0, & r_a(t) < m \end{cases} \tag{25}$$

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,

$$f_0(t) = \begin{cases} f(t), & v(t) = 1 \\ 0, & v(t) = 0 \end{cases} \tag{26}$$

where $v(t)$ is given by

$$v(u) = \begin{cases} M, & r_b(u) \leq 1/k \\ 0, & r_b(u) > 1/k \end{cases} \tag{27}$$

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

4.2. Stationary simulated processes

Process records are generated compatible with the JONSWAP spectrum to simulate wave height over time (Eq.28) [49]. This spectrum typically has a very sharp, strong peak and it is relatively sparse (narrow-band) in the

frequency domain. Further, JONSWAP spectrum compatible time-histories of length $N_0 = 256$ are generated by utilizing Eq.4. The JONSWAP spectrum has the form,

$$S(\omega) = \frac{ag^2}{\omega^5} e^{-\frac{5}{4}\left(\frac{\omega_p}{\omega}\right)^4} \gamma^r; \quad r = e^{-\left(\frac{\omega-\omega_p}{2\sigma\omega_p}\right)^2} \tag{28}$$

where $\alpha = 0.03$, $\omega_p = 0.5$, $\gamma = 3.3$, 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 a Kanai-Tajimi earthquake spectrum [50, 51] 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.4, based on the spectrum

$$S(\omega) = \frac{1 + \alpha\omega^2}{\left(\omega_g^2 - \omega^2\right) + \left(2\zeta\omega\omega_g^2\right)^2}, \tag{29}$$

where the natural frequency, $_g = 12rad/s$, damping ratio $\zeta = 0.4$ and $\alpha = 20$. Furthermore, Gaussian white noise is added to the process realizations of magnitude $1/5^{th}$ the standard deviation of the record in both cases. Three examples are shown, two for the JONSWAP spectrum and one for the Kanai-Tajimi spectrum. For each example, ten process realizations are generated and data is removed. 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 11 shows a full example process realization alongside the same with 65% missing data. Figure 12 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, zero padded reconstruction it is clear that the CS approach is far superior to a scaled least squares which has a very low spectral peak and a long tail of noise across all other frequencies.

The second example (Figure 13) uses the same spectrum but this time with 50% missing data at intervals of length $N_0/32$ (Eq.26). 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 recon-

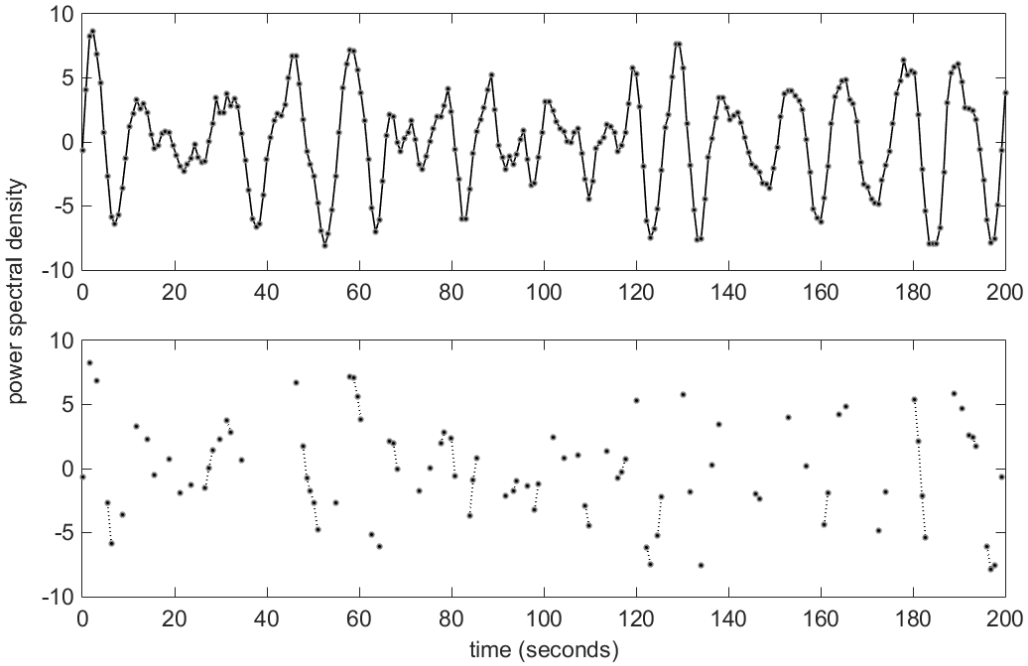


Figure 11: Example JONSWAP process (top) and the same process with 65% missing data (bottom)

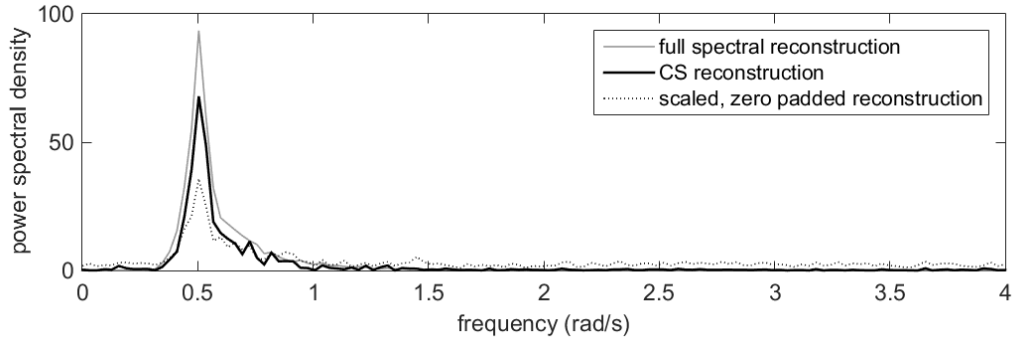


Figure 12: JONSWAP power spectrum reconstruction from 10 stationary process records via $L1$ minimization and scaled, zero-padding for 65% missing data at uniform random locations

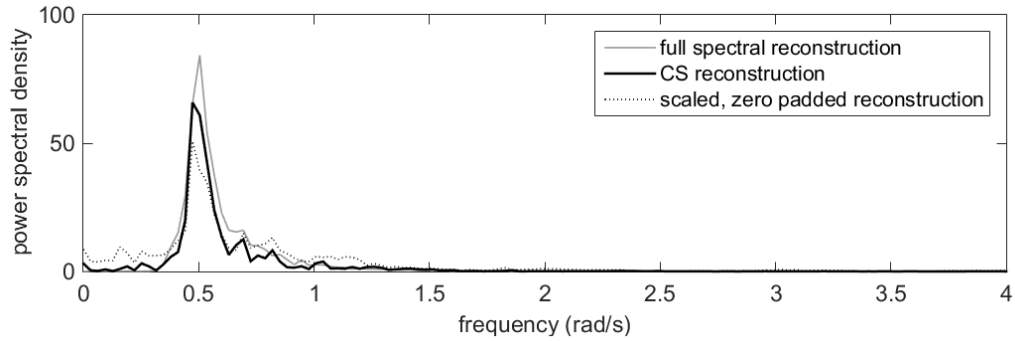


Figure 13: JONSWAP power spectrum reconstruction from 10 stationary process records via $L1$ minimization and scaled, zero-padding for 50% missing data over randomly located fixed intervals of length $N_0/32$

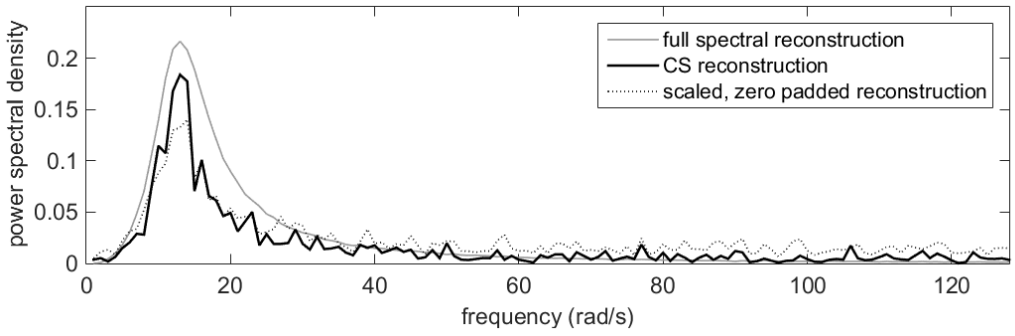


Figure 14: Kanai-Tajimi power spectrum reconstruction from 10 stationary process records via $L1$ minimization and scaled, zero-padding for 50% missing data at uniform random locations

struction when the signal is not as sparse in the Fourier domain. For the Kanai-Tajimi process, the number of missing data was reduced to 50% to account for the broader spectrum, by which point a least-squares solution became a viable option (although it still retains higher noise at large values of the frequency domain). The differences between the two approaches shown in Figure 14 are far less significant than in previous examples. In this case either reconstruction method could be considered.

4.3. 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 [52] of the form,

$$S(\omega,t) = g(t)^2 S_0 \frac{\omega^4}{\left(\omega_f^2-\omega^2\right)+4\zeta_f^2\omega_f^2\omega^2} \cdot \frac{\omega_g^4+4\zeta_g^2\omega_g^2\omega^2}{\left(\omega_g^2-\omega^2\right)+4\zeta_g^2\omega_g^2\omega^2}, \tag{30}$$

where $S_0 = 0.06$, $\omega_f = 1$, $\zeta_f = 0.6$, $\omega_g = 10$, $\zeta_g = 0.4$ and $g(t)$ is the envelope function,

$$g(t) = ke^{-at} - e^{-bt}, \tag{31}$$

where $k = 4$, $a = 0.3$ and $b = 0.6$. Finally, for the non-separable case, the method is tested on an earthquake process of the form,

$$S(\omega,t) = \left(\frac{\omega}{5\pi}\right)^2 e^{-0.15t^2} e^{-\left(\frac{\omega}{5\pi}\right)^2 t}. \tag{32}$$

Eq.32 (Figure 17) has a broad spectrum of power at the start of the time history. High frequency powers are significant during the first second of the process, quickly reducing and resembling more of a separable process during the final ten seconds. Naturally, for the non-stationary reconstructions, a GHW basis is used. For the separable spectrum (Eq.30), the bandwidth of the wavelets is set at $1/32^{nd}$ the length of the corresponding Fourier power spectrum (or $8rad/s$ 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.32), a larger bandwidth of $1/16^{th}$ the length of the power spectrum is used ($16rad/s$). This allows the GHWT to better capture the more rapid changes in power over time. Figures 15 & 17 show the target spectra drawn directly from Eq.30 & Eq.32. Comparatively, Figures 16 & 18 show the average power of the GHWT of 25 time-histories

of the separable and non-separable processes (generated using Eq.14). 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 16 & 18 still present a useful result, showing the location of the peak frequencies and the trends over time. Because of these limitations of the HWT with no missing data, Figures 16 & 18 are considered to be the target spectra for reconstruction in the separable and non-separable examples respectively. In the following examples 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 least-squares (Figure 19) and CS (Figure 20) 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 least squares. Figures 21 and 22 show the same reconstructions but with uniformly distributed intervals of missing data of length $N_0/32$. Here the differences are not so apparent as although the CS reconstruction remains relatively unchanged, the scaled least squares solution has much lower noise than in Figure 19. 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 21, the least-squares 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.30). 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 24) over scaled least squares (Figure 23). The low-power, high frequency spike that occurs at the very beginning of the process is lost in a sea of noise in Figure 23. The significance of this difference between the two reconstructions is made even clearer when using them to generate new process realizations. Figure 27.a shows a non-stationary process realization compatible with Eq.32, note the high frequency oscillations at the start, slowing over time. Figure 27.b shows a process realization generated from Figure 24 (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 27.c shows a process realization generated from Figure 23 (the scaled least-squares reconstruction). Here it is rather difficult to identify visually any frequency dependent change

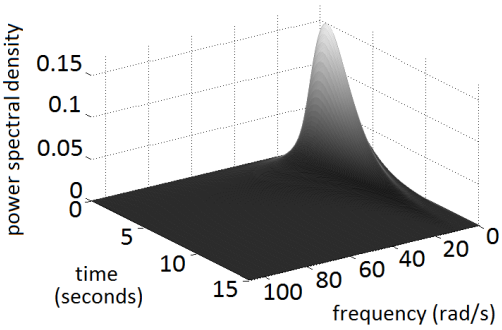


Figure 15: Separable target spectrum drawn from Eq.30

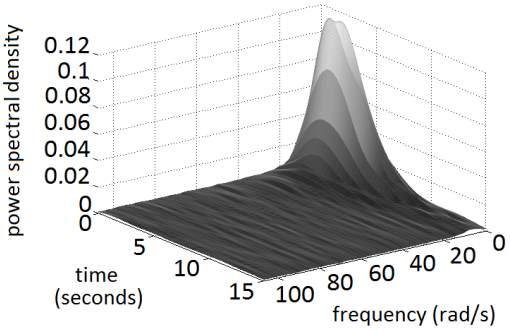


Figure 16: Averaged HWT spectrum estimation with no missing data using 25 time-histories compatible with Eq.30

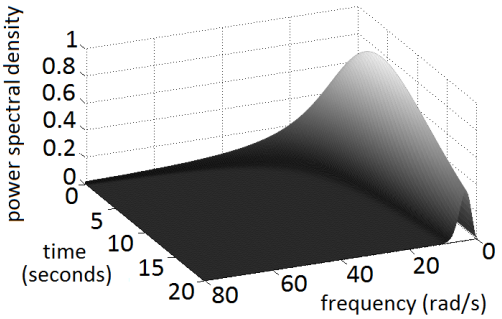


Figure 17: Separable target spectrum drawn from Eq.32

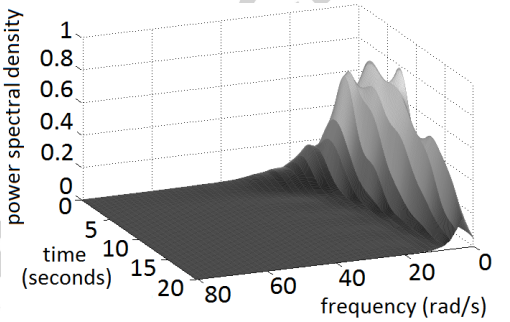


Figure 18: Averaged HWT spectrum estimation with no missing data using 25 time-histories compatible with Eq.32

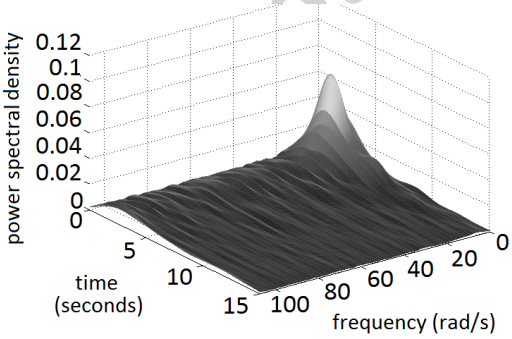


Figure 19: Separable earthquake power spectrum reconstruction from 25 stationary process records via least squares for 50%/ missing data at uniform random locations

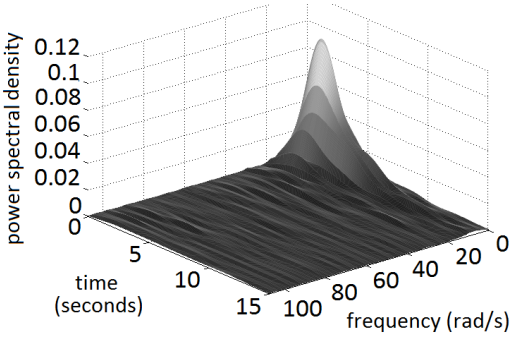


Figure 20: Separable earthquake power spectrum reconstruction from 25 stationary process records via $L1$ minimization for 50% missing data at uniform random locations

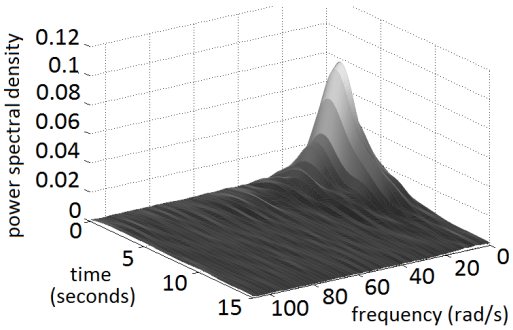


Figure 21: Separable earthquake power spectrum reconstruction from 25 stationary process records via least squares for 50% missing data over randomly located fixed intervals

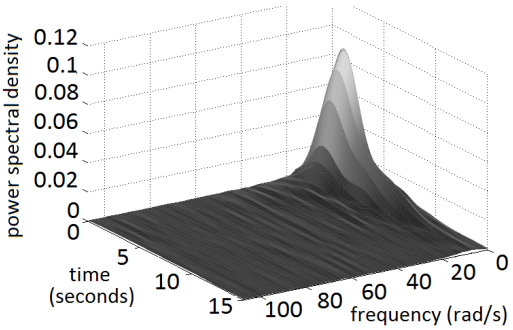


Figure 22: Separable earthquake power spectrum reconstruction from 25 stationary process records via $L1$ minimization for 50% missing data over randomly located fixed intervals

over time, further high frequency noise has a significant presence throughout. For the interval-based missing data examples (Figures 25 & 26), as in the separable case, the differences are less apparent. However, Figure 26 does still present less noise at higher frequencies and has a more defined shape than Figure 25, more closely matching that of Figure 18.

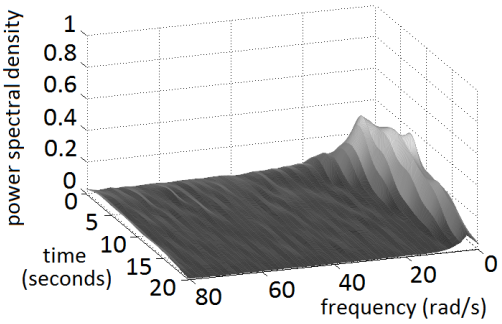


Figure 23: Non-separable earthquake power spectrum reconstruction from 25 stationary process records via least squares for 50%/ missing data at uniform random locations

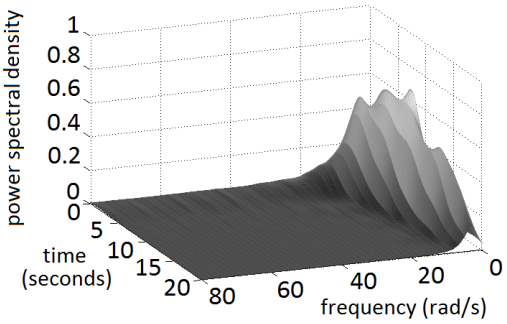


Figure 24: Non-separable earthquake power spectrum reconstruction from 25 stationary process records via $L1$ minimization for 50% missing data at uniform random locations

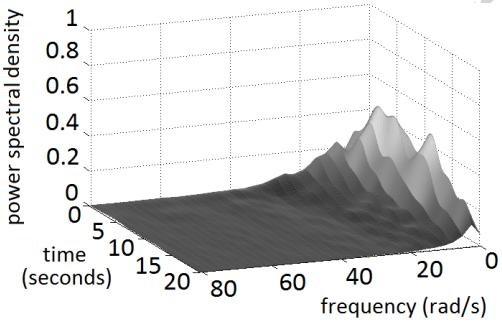


Figure 25: Non-separable earthquake power spectrum reconstruction from 25 stationary process records via least squares for 50% missing data over randomly located fixed intervals

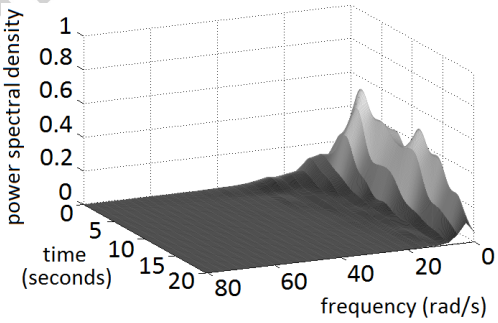


Figure 26: Non-separable earthquake power spectrum reconstruction from 25 stationary process records via $L1$ minimization for 50% missing data over randomly located fixed intervals

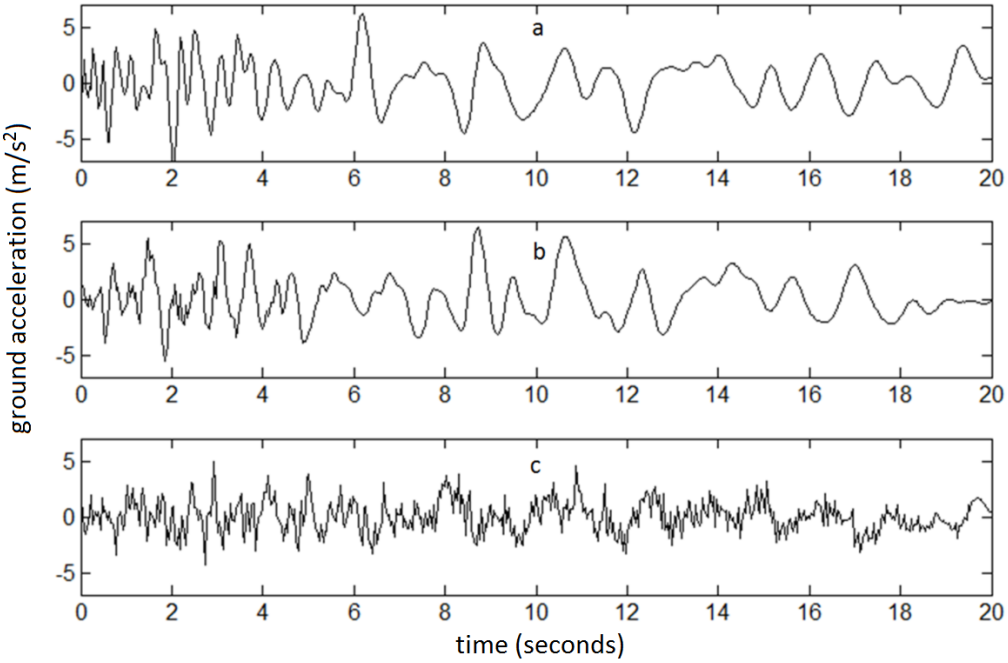


Figure 27: (From top to bottom) single realization of Eq.32, single realization from Figure 26 and a single realization of Figure 25

5. Conclusion

In this paper, a CS based approach for stationary and non-stationary stochastic process power spectrum estimation subject to missing data has been developed. 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 has been shown to be highly effective in reconstructing the power spectrum in many cases. The approach appears efficient and superior to least-squares 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 is capable of providing with accurate power spectrum estimates under a range of missing data configurations. Of particular note is the extent to which CS-reconstruction with harmonic wavelets is able to produce non-separable spectra with up to 50% missing data. The accuracy of non-separable estimated spectra is comparable to that of the separable spectra results, suggesting that the herein developed approach can be efficacious in a range of problems in which very little a priori information is known about the shape of the spectrum.

References

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

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

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

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

[5] 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.

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

[7] 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.

[8] 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.

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

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

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

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

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

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

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

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

[17] 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.

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

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

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

[21] 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.

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

[23] 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.

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

[25] H. Cramèr, R. Leadbetter, Stationary and related stochastic processes: sample function properties and their applications, Wiley, 1967, lc66026738.

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

[27] P. D. Welch, The use of fast fourier transform for the estimation of power spectra: A method based on time averaging over short, modified periodograms, Audio and Electroacoustics, IEEE Transactions on 15 (2) (1967) 70–73.

[28] M. Priestley, Wavelets and time-dependent spectral analysis, Journal of Time Series Analysis 17 (1) (1996) 85–103.

[29] 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.

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

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

[32] H. Ombao, J. Raz, R. von Sachs, W. Guo, The slex model of a non-stationary random process, Annals of the Institute of Statistical Mathematics 54 (1) (2002) 171–200.

[33] 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.

[34] 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.

[35] 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.

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

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

[38] 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.

[39] A. S. Bandeira, E. Dobriban, D. G. Mixon, W. F. Sawin, Certifying the restricted isometry property is hard, arXiv preprint arXiv:1204.1580.

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

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

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

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

[44] 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.

[45] 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.

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

[47] 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.

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

[49] 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).

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

[51] 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.

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