Compressive sensing with an adaptive wavelet basis for structural system response and reliability analysis under missing data

L. Comerford[a], H.A. Jensen[b], F. Mayorga[b], M. Beer[a,c,d], I. A. Kougioumtzoglou[e],

[a]*Institute for Risk and Reliability, Leibniz University Hannover, Germany*
[b]*Dept. of Civil Engineering, Santa Maria University, Valparaiso, Chile*
[c]*Institute for Risk and Uncertainty, University of Liverpool, Liverpool, L69 3GH, UK*
[d]*School of Civil Engineering & Shanghai Institute of Disaster Prevention and Relief, Tongji University, China*
[e]*Department of Civil Engineering and Engineering Mechanics, Columbia University, USA*

Abstract

The challenge of determining response and reliability statistics of large-scale structural systems under earthquake induced stochastic excitations is considered where the source load data records are incomplete. To this aim, a compressive sensing based framework in conjunction with an adaptive wavelet basis is presented for reconstructing the samples with missing data and estimating the underlying process EPS. In this regard, novel insights are provided whereas certain conceptual, numerical, and practical implementation aspects of the technique are presented in detail. A numerical example pertaining to the stochastic response and reliability analysis of an eight floor reinforced concrete building structural system demonstrates the effectiveness of the proposed methodology.

Keywords: Compressive Sensing, Harmonic Wavelets, Power Spectrum, Missing Data, Advanced Simulation Techniques, Reliability Analysis

Email address: comerford@irz.uni-hannover.de (L. Comerford)

1. Introduction

Numerical simulations for the analysis of structural systems subject to random dynamic excitations require realistic stochastic models of the system excitation processes. Such systems may be highly sensitive to the nature of these excitations and so simulation accuracy is dependant upon a reliable excitation process model.

A reliable spectral model providing frequency dependant information can be of significant importance in investigating the response of an engineering system to stochastic input such as earthquakes. Further, spectral models may be utilized for generating stochastic process records, fitting with the frequency dependant statistics of the given model, for use in numerical Monte-Carlo analyses e.g. [1, 2, 3]. However, a basic spectral model such as that based on a non-windowed discrete Fourier transform (DFT) may only describe a stationary process, i.e. one in which the spectral content does not change over time. This assumption of stationarity may give a poor approximation of the true process, especially in the case of earthquake excitations in which the frequency content can change dramatically over their duration. Hence, in many cases, realization of time-dependant properties of stochastic processes is also considered central to defining reliable spectral models. In this regard, the concept of the evolutionary power spectrum (EPS) [4, 5] provides an appealing model for capturing the statistics and the time-varying frequency content of the underlying non-stationary stochastic processes. Further, they can be used as a basis for joint time-frequency system response analysis [6, 7], or efficient stochastic simulation utilizing advanced Monte Carlo techniques.

In an ideal scenario, such a model could be avoided entirely in the case where extremely large data banks of real recorded excitation processes of interest were available. Unfortunately, particularly in the field of earthquake engineering, this is seldom the case. Instead, process models are often estimated, based partially or entirely on a small set of relevant recorded processes. Numerous approaches exist for EPS estimation based on time records including short-time Fourier transforms, wavelet [8, 9, 10] & chirplet [11] transforms. Harmonic wavelets [12, 13] are concentrated on in this paper due in part to their box-shaped frequency spectrum, ideal for identifying specific bands of energy and for the fact that they constitute an orthogonal basis, which is ideal for the compressive sensing approach applied herein.

It is logical to assume that the more data upon which such a model is

based, the more statistically accurate/relevant numerical simulation results are likely to be. As the available data may be quite limited, it is important that it is utilized to the fullest extent, which in some cases includes working with problematic data sets. In this regard, when analyzing real earthquake excitation data, coupled with the problem of limited numbers of samples or shorter than ideal sample lengths, is the potential major issue of missing data. Practical reasons for having limited data include for example, equipment failure (if a sensor becomes damaged, perhaps even as a result of the process itself, data may be lost) and sensor thresholding limitations (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). Numerous other issues including sensor maintenance, bandwidth limitations, usage & data acquisition restrictions as well as data corruption may also lead to missing data.

Under these conditions, when working with limited and/or missing data, standard Fourier techniques for spectral estimation, will frequently demonstrate poor performance. Although there exist many algorithms and procedures in the literature that provide spectral estimates in the presence of missing data, these alternatives come with certain drawbacks and often impose significant assumptions on the statistics of the underlying stochastic process. For instance, autoregressive methods may be applied under the assumption that source time-histories are relatively long and that the missing data are grouped [14, 15]. Further, least-squares sinusoid fitting and zero-padded gaps [16, 17, 18] 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. Similar issues are faced when applying these methods to wavelet transforms in the case of EPS estimation, and specific approaches for non-stationary signal reconstruction are uncommon. However, recent developments have been made in the area of EPS estimation subject to missing data including applications of artificial neural networks [19] and compressive sensing (CS) [20]. The latter is applied herein, utilizing the relative band-limited nature of evolutionary earthquake spectra. To further improve the spectrum estimation, the CS approach is applied in conjunction with an adaptive basis re-weighting procedure, building on ideas introduced in [21], which is useful in the case where process record ensembles are available.

The organization of this contribution is as follows. A brief introduction to CS theory is provided in Section 2 with references to further reading.

Section 3 highlights the differences between the methods applied herein and established CS theory. The novel adaptive basis re-weighting procedure for signal reconstruction from multiple records is detailed in Section 4. In Section 5, discrete orthogonal Harmonic Wavelets are introduced briefly, along with commonly encountered practical wavelet transform issues, before being set in the context of the CS-based reconstruction problem. Section 6 deals with the estimation of ground excitation power spectra from simulated earthquake records with various missing data configurations for which response statistics and system reliability of a large structural model subject to such excitations are compared. The work closes with some conclusions and final remarks.

2. Compressive sensing

CS [22, 23] 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 several important assumptions to be made concerning the nature of the process of interest. However, in many problem cases, especially those related to environmental processes (and in particular spectral representation of earthquake excitations), 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 alternative reconstruction methods.

2.1. CS background

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 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 [24].

2.2. CS requirements

For robust compressive sensing, several properties concerning the source signal and transformation basis are required. Most importantly the signal must be sparse in a known basis, and obey properties of incoherence and restricted isometry (RIP). The last two requirements are discussed in detail

in any introductory text on CS theory (eg. [25].) For clarity and completeness in notation a brief description of sparsity is provided in the following subsections.

2.2.1. Signal sparsity

For a sampled signal to exhibit sparsity in some known basis, it must be possible to represent that 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.,

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

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.,

$$x = \sum_{i=1}^{K} A_{n_i} y_{n_i}, \tag{2}$$

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,

$$|y|_{L0} = K, \tag{3}$$

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

$$|y|_{Lp} = \left(\sum_i |y_i|^p \right)^{\frac{1}{p}}. \tag{4}$$

The $L0$ norm used in Eq.3 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,

$$|y|_{L0} = \sum_i \begin{cases} 1 & y_i \neq 0 \\ 0 & otherwise \end{cases} \tag{5}$$

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.

2.2.2. 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 (absolute value) minimization.

In principle, the sparsest solution of an under-determined system of equations occurs when the $L0$ norm is minimized (the minimum number of non-zero coefficients). 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 [26]. Further, the problem becomes convex, and may be set in a convenient linear programming form, i.e.

$$\min |y|_{L1} \;\; subject\ to\ x = Ay \tag{6}$$

Eq. 6 describes a basis pursuit optimization problem and can be easily solved via a gradient-based optimization method, e.g. [27]. Unfortunately, real signals are rarely ever truly sparse; even low levels of noise will produce small coefficients across most bases. Hence, a tolerance, e, relative to the variance of the noise is included and Eq. 6 may be re-cast in the form,

$$\min |y|_{L1} \;\; subject\ to\ \; |Ay - x|_{L2} \leq e. \tag{7}$$

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.

3. CS application to 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. [28, 29, 30, 31, 32]). 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.

3.1. Generalization of CS

It is clear at this point that CS techniques have a wide range of potential applications in the area of load sensing/modelling when considering that in many cases, particularly for environmental loads such as those produced by earthquakes (including structural responses to these effects), can be characterized by a relatively small space of dominant frequencies in the frequency domain.

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. When applying CS to pre-recorded data series with missing data that were not originally sampled in the context of a CS framework, the sampling matrix is fixed as a series pulses of magnitude 1 in the sampling domain. Hence in the sampling domain, the signal of interest must be non-sparse. Fortunately, harmonic signals sampled in the time domain exhibit this property, especially those with narrow band spectra due to the fact that the Fourier and time domains are maximally incoherent.

3.2. Illustrative example

As an example, consider a situation in which a harmonic signal (Eq. 8) is captured in the time domain but where some of the samples are randomly missing.

$$x_f(t) = \cos(6t) + \cos(12t) \tag{8}$$

For compression, the DFT would be highly applicable in this case, producing only two coefficients. Figure 1 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 discrete time domain, x_f is the full signal shown in Eq. 8 and A_f is a real Fourier matrix (composed of cosines). If the sampling matrix is arranged to simulate

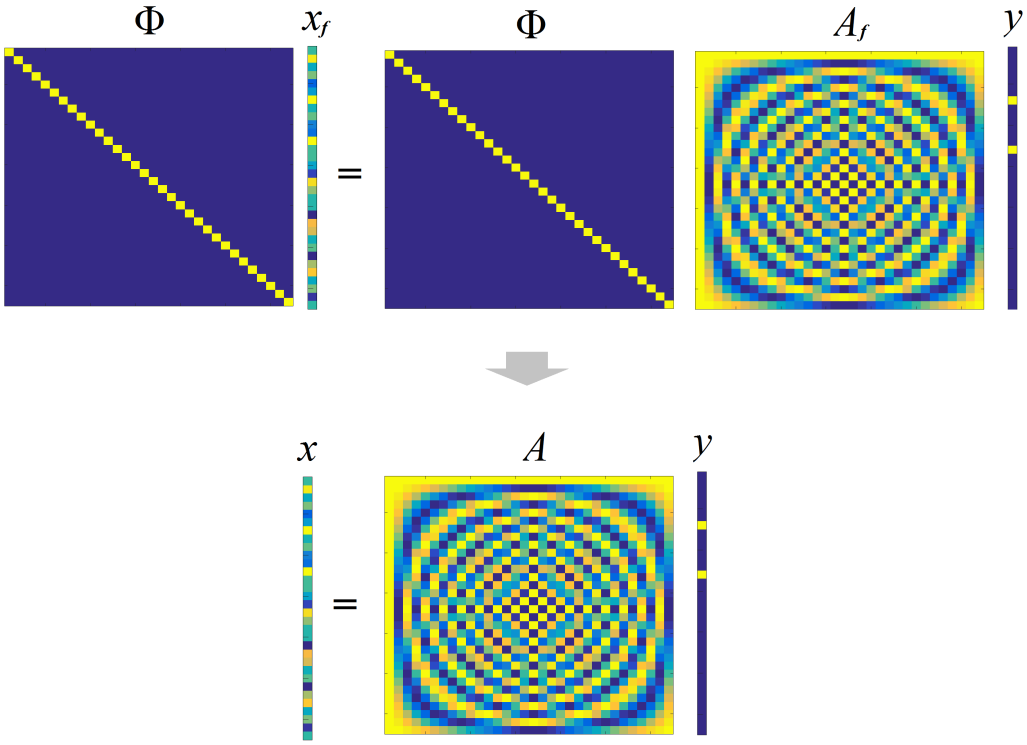

Figure 1: Signal acquisition of Eq. 8 at every point in N. Note that x_f is sparse in the Fourier matrix A_f, hence y has only two entries (one representing $\cos(6t)$ and one representing $\cos(12t)$)

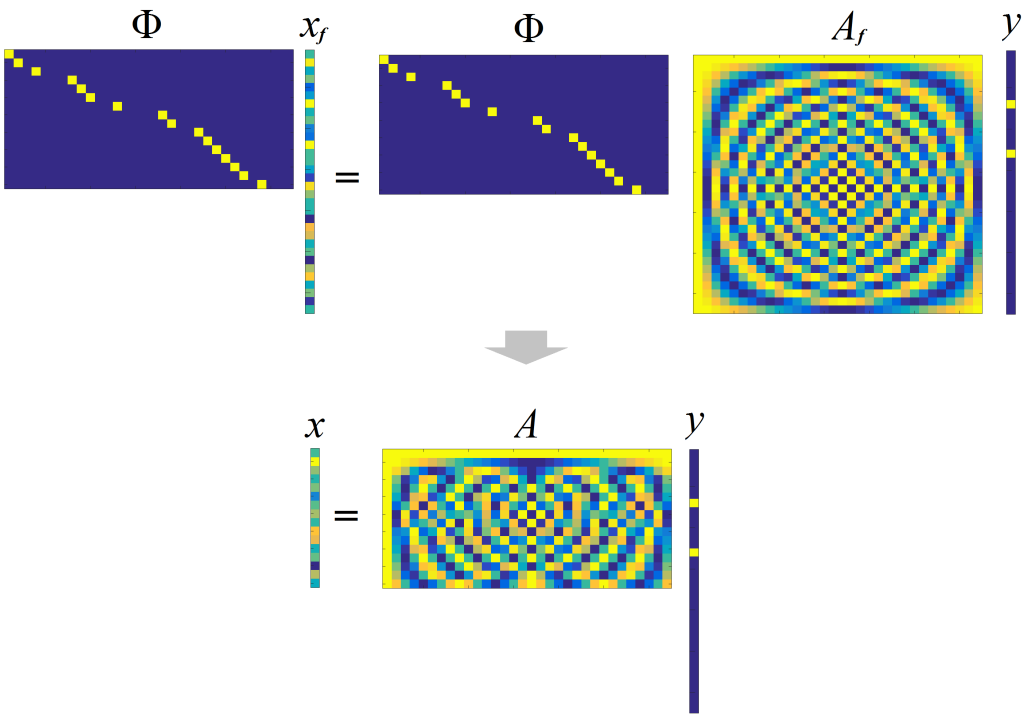

Figure 2: Eq. 8 with $\frac{N}{2}$ uniformly distributed missing data over x_f set in CS framework with a Fourier basis

missing data in a recorded process realization, it would appear with ordered rows as in Figure 2. Uniform random Fourier matrices obey the RIP with high probability when data are sparse [22, 23]; 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. 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. Adaptive basis re-weighting procedure

Herein, 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 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.

4.1. Formulation based on weighting matrix

It is shown in [20] 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 and can be found in [21]. This section expands on the aforementioned presentation of the adaptive basis method for additional clarity, while at the same time provides with novel insights and details on practical implementation issues.

The motivation behind this adaptive basis approach comes from the fact that when utilizing an ensemble of process records for producing a single power spectrum estimate, for sharp (narrow-banded) spectra, CS should assume sparsity in the mean sense, rather than sparsity for each individual record. Further, $L1$ minimization, unlike L0, does not guarantee the sparsest solution and will yield different solutions if columns of the basis matrix are assigned different weights. Hence it was proposed that weights could be used

to alter individual $L1$ solutions to missing data problems, possibly reducing sparsity of individual solutions but enhancing sparsity in the mean. [21] demonstrates the effect of re-weighting the basis matrix visually for a simple 3-dimensional problem. By introducing a diagonal matrix of weights W, Eq. 6 is recast in the form,

$$\min |y|_{L1} \;\; subject\ to\ x = AWy; \tag{9}$$

The weighting matrix is updated iteratively by solving Eq. 9 in a least-squares sense rather than with $L1$ minimization. When compared to $L1$ minimisation, not only does this decrease the computational effort of the re-weighting to a large extent, but also reduces the possibility of very large false peaks affecting the weights; 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. 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 approach taken 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.

4.2. Proposed procedure

The re-weighting procedure is described in detail in the following with reference to Figure 3

1. Two temporary re-weighting matrices are required for the iterative procedures. These are both initialized as square identity matrices.
2. The re-weighting matrix is 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 as well.
3. 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.

4. 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 transformed signals across the ensemble (e.g. for wavelet based EPS estimation see Eq. 14). Hence, the next iteration of the re-weighting matrix is built up over a loop of all process records in the ensemble.
5. 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 where A is the reduced basis matrix.
6. Here, the basis matrix A is assumed to be made up of odd and even functions (e.g. DFT or "generalized harmonic wavelet transform" (GHWT)). As the power spectrum model is not dependant on phase, the weighting of odd and even function components separately is not desired. Hence, the individual coefficients are split into pairs.
7. The Euclidean norm for these pairs is taken to convert them into frequency-only dependant coefficients.
8. 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 odd and even components of the same function equally on the next iteration. These diagonal coefficients are added for each record in the ensemble.
9. Once the loop at point 4. has ended, the re-weighting matrix is normalized by its mean, and summed with a pre-defined constant bias. Next, the re-weighting matrix is set equal to the filled W_2 and then W_2 is once again set to zero and the loop repeats. 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.

In Figure 3, step 9 shows that a bias has been added to the weights, to prevent zero valued coefficients. Zero-valued coefficients impose sparsity by removing low-valued peaks. Once these are removed, the iterative algorithm

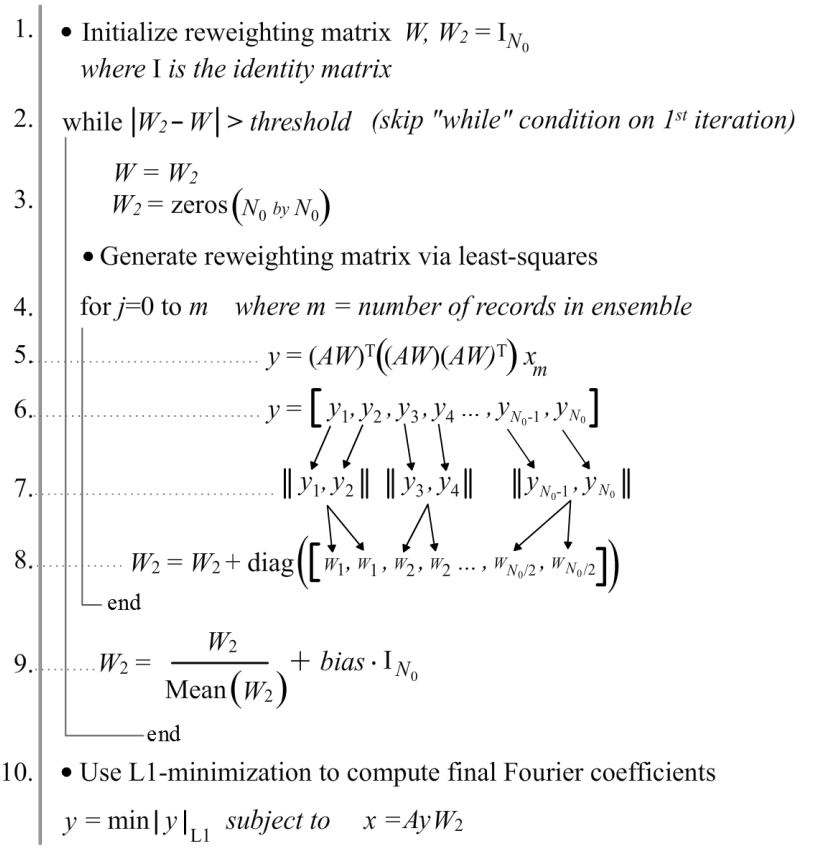

Figure 3: Mechanization of approach based on CS with adaptive basis

is forced to use remaining coefficients to represent the records. In many cases this will lead to false peaks in seemingly random locations, except when the signal is extremely sparse in the transform domain. The optimal choice of bias is problem dependant; in some cases a very low bias can lead to highly accurate spectrum estimates. However, as accuracy cannot be validated in real problems, a bias of unity is used in the upcoming examples. This, effectively prevents weights from decreasing in magnitude with the possibility of growing to one standard deviation from the weight mean when combined with the mean normalization, also shown in step 9. Note that as the bias increases, the solution tends towards that of non-re-weighted CS.

5. Evolutionary power spectrum representation via harmonic wavelets

In the previous section CS was applied with a basis matrix in which the signal must have a sparse representation. In order to apply CS for reconstruction of non-stationary process records, an appropriate non-stationary basis must be defined. In this regard, [33] developed a framework for representing non-stationary stochastic processes by utilizing a time/frequency-localized wavelet basis, as opposed a stationary Fourier decomposition. The representation reads

$$X\left(t\right)=\sum_{j}\sum_{k}w_{j,k}\psi_{j,k}\left(t\right)\xi_{j,k}, \tag{10}$$

where $\psi_{j,k}\left(t\right)$ is the chosen family of wavelets, j and k represent the different scales and translation levels respectively and $\xi_{j,k}$ is a stochastic orthonormal increment sequence (in which pairs of $\xi_{j,k}$ where j and k are not equal are uncorrelated). The local contribution to the variance of the process of Eq. 10 is given by $\left|w_{j,k}\right|^{2}$.

The wavelet-based model of Eq. 10 relies on the theory of locally stationary processes (see also [34]).

Next, utilizing the family of generalized harmonic wavelets [12, 13], it can be shown that Eq. 10 may be written in terms of its evolutionary power spectral density function (see [35]). Defined in the frequency domain, generalized harmonic wavelets exhibit the following form,

$$\Psi_{(m,n),k}^{G}\left(\omega\right)=\begin{cases}\frac{1}{(n-m)\Delta\omega}e^{\left(-i\omega\frac{kT_{0}}{n-m}\right)}, & m\Delta\omega\leq\omega\leq n\Delta\omega\\ 0, & otherwise\end{cases}, \tag{11}$$

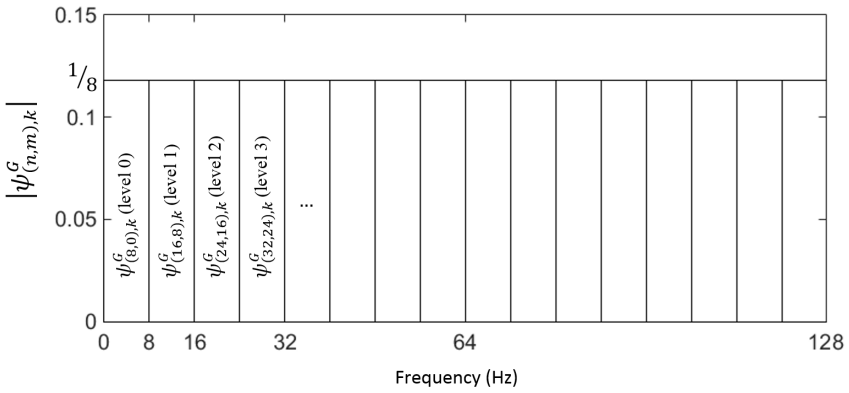

Figure 4: Harmonic wavelets in the frequency domain with $n-m=8Hz$

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. 11 span frequency bands defined by m and n as shown in Figure 4. An orthogonal set of harmonic wavelets are produced when n and m define adjacent non-overlapping intervals for all wavelets in the set. Substituting generalized harmonic wavelets into Eq. 10 (see [35]), the process may be written as,

$$X_{(m,n),k}(t) = \sqrt{S^X_{(m,n),k}(n-m)\Delta\omega}\psi_{(m,n),k}(t)\,\xi_{(m,n),k}. \tag{12}$$

Eq. 12 represents a localized process at scale (m, n) and translation (k) defined in the intervals $[m\Delta\omega, n\Delta\omega]$ and $\left[\frac{kT_0}{n-m}, \frac{(k+1)T_0}{n-m}\right]$, whereas $S^X_{(m,n),k}$ represents the spectrum $S_X(\omega, t)$ at scale (m, n) and translation (k). Further, it has been shown that realizations compatible with $S_X(\omega, t)$ can be generated in the following way (see [5])

$$X(t) = \sum_{j=0}^{N-1} \sqrt{4S_X(\omega_j, t)\Delta\omega}\sin(\omega_j t + \Phi_j), \tag{13}$$

where Φ_j are uniformly distributed random phase angles in the range $0 \leq \Phi_j < 2\pi$.

5.1. 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 [36, 35]

$$S_X(\omega,t) = S^X_{(m,n),k} = \frac{E\left(\left|W^G_{(m,n),k}[X]\right|^2\right)}{(n-m)\Delta\omega}, \quad m\Delta\omega \le \omega \le n\Delta\omega, \quad \frac{kT_0}{n-m} \le t \le \frac{(k+1)T_0}{n-m}, \tag{14}$$

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.

5.2. Practical issues relating to wavelet transforms

5.2.1. Time-frequency resolution trade-off with the wavelet transform

As previously shown, harmonic wavelets are defined by their bounds n and m in the frequency domain (Figure 4). 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 5). 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.

5.2.2. Mitigation of 'end effects' due to the wavelet transform

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 discrete linear transforms performed by circular convolution (such as the DFT and GHWT), 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 transform to give coefficients directly representing only the original components of the signal. The consequence of this fact is sometimes known as spectral leakage.

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

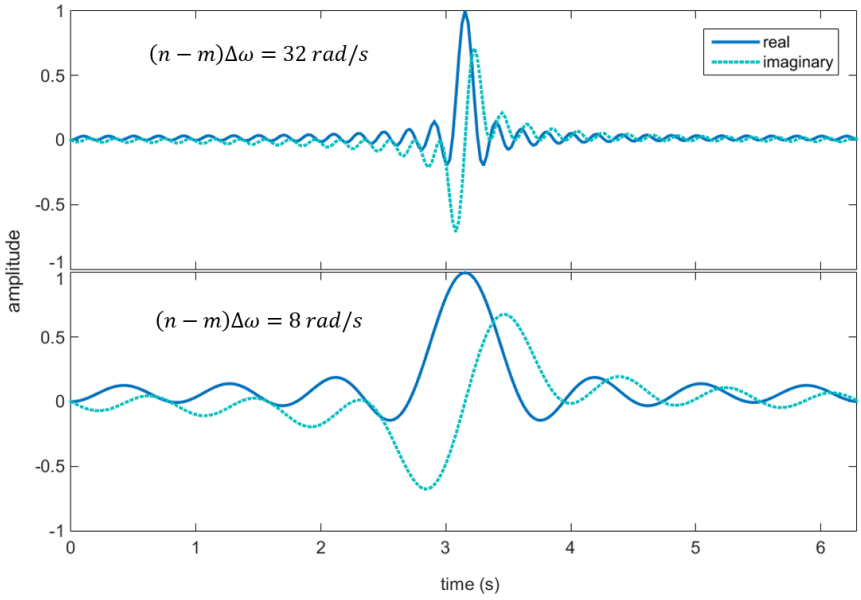

Figure 5: Comparison of Harmonic wavelets in the time domain for high (top) and low (bottom) resolution in time

by these end effects is referred to as the "cone of influence" [37]. Basic practical methods of preventing the signal from wrapping around onto itself include padding the time domain signal with zeros at the beginning and end or to pad the original signal with its reverse on both sides. For high frequency resolution wavelets, padding with zeros can significantly reduce the mean estimated spectral power, reverse signal padding is less likely to suffer from this problem, hence, the latter is utilized in the upcoming numerical examples.

5.3. Harmonic wavelet basis matrices for compressive sensing

In this paper, for CS process reconstruction, harmonic wavelet basis matrices are utilized. 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.

For harmonic wavelet basis construction, wavelet scales must first be defined; that is, a set of non-overlapping frequency intervals corresponding to $n - m$. In most cases these are chosen to be equally spaced with an interval

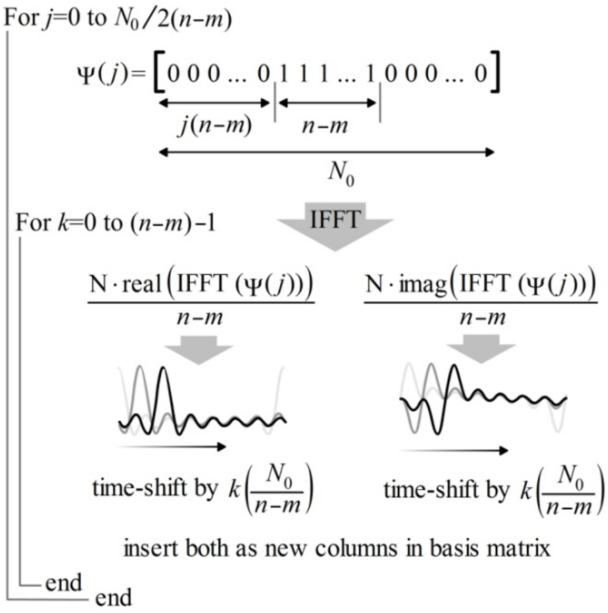

Figure 6: Non-redundant orthogonal Harmonic wavelet basis construction using IFFT and nested for-loops

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 basis matrix can be altered accordingly. The harmonic wavelet basis components may be generated efficiently via the Inverse Fast Fourier Transform (IFFT) as shown in [38]. 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 square harmonic wavelet basis matrix is depicted in Figure 6. Rows must be removed corresponding to the missing data, yielding a sampling matrix with more columns than rows (Figure 7).

With the basis formed, the CS reconstruction may be solved. Note that due to the requirement of incoherence between the signal domain and transform domain for CS, the harmonic wavelets should occupy relatively narrow frequency bands. I.e., as n-m increases, incoherence decreases.

6. Application Problem

The objective of the application problem is to evaluate the performance of the proposed methodology in treating missing data. In particular, the

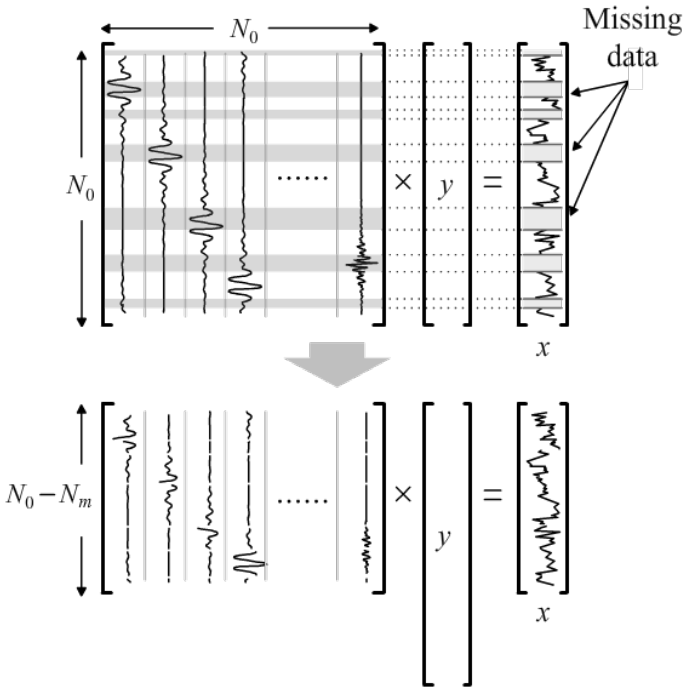

Figure 7: Harmonic wavelet sampling matrix construction with missing data

Floor	1	2	3	4
Mass (ton)	204.7	202.7	198.9	195.6
Floor	5	6	7	8
Mass (ton)	192.2	188.9	186.7	175.5

Table 1: Total mass of the different floors

effect of estimated spectra on the response statistics and reliability of a large structural model under earthquake excitation is investigated.

6.1. Model Description

The structural model shown in Figure 8 is considered for analysis. It consists of an eight floor, three-dimensional reinforced concrete building model under stochastic ground acceleration. Material properties of the reinforced concrete structure have been assumed as follows: Young's modulus $E = 2.56 \times 10^{10}$ N/m^2; Poisson ratio $\nu = 0.2$; and mass density $\rho = 2500$ kg/m^3. The total mass of the different floors is given in Table 1. The height of each floor is 3.5 m leading to a total height of 28.0 m for the structure. The floors are modeled with shell elements with a thickness of 0.2 m. Additionally, beam and column elements are used in the finite element model, which has 102960 degrees of freedom. A 5% of critical damping for the modal damping ratios is introduced to the model. The building is excited horizontally by a ground acceleration applied at 35^o with respect to the axis x. The ground excitation is modeled as indicated in the following section. For an improved earthquake performance the structural system is enforced with twelve vibration control devices. In particular, vibration control devices composed of a series of metallic U-shaped flexural plates (UFP) are considered in this study [39]. The vibration control devices are connected to the structure every four floors as indicated in Figure 8. A typical configuration of the vibration control device is shown in Figure 9 (left picture). It consists of brace and plate elements where a series of UFP's (40 in total) are located between the plates as shown in Figure 9 (right picture).

Each UFP exhibits a one-dimensional hysteretic type of non-linearity modeled by the restoring force law

$$r_d(t) = \alpha\, k_e\, \delta(t) + (1-\alpha)\, k_e U^y\, z(t) \tag{15}$$

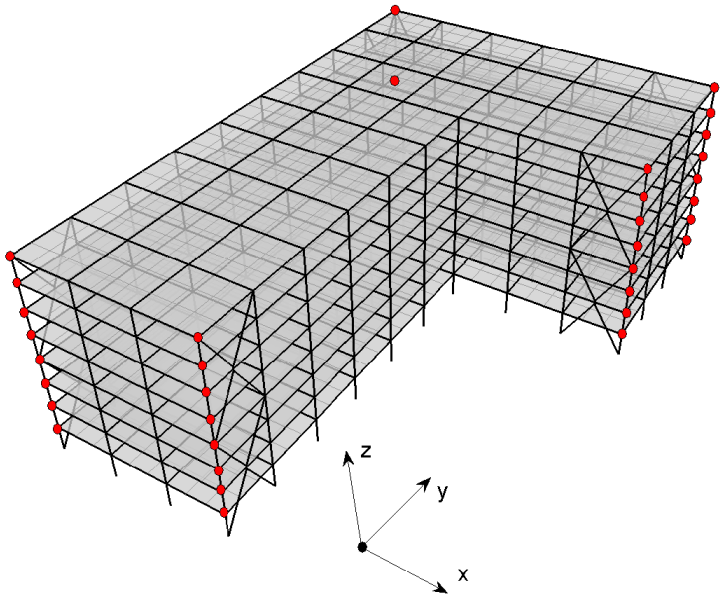

Figure 8: Isometric view of the finite element model

Figure 9: Left picture: typical configuration of a vibration control device. Right picture: dissipators in the form of metallic U-shaped flexural plates (UFP).

where k_e is the pre-yield stiffness, U^y is the yield displacement, α is the factor which defines the extent to which the restoring force is linear, $z(t)$ is a dimensionless hysteretic variable, and $\delta(t)$ is the relative displacement between the upper and lower surfaces of the device. The hysteretic variable $z(t)$ satisfies the first-order non-linear differential equation

$$\dot{z}(t) = \dot{\delta}(t) \left[\beta_1 - z(t)^2[\beta_2 + \beta_3 \text{sgn}(z(t)\dot{\delta}(t))]\right] \tag{16}$$

where β_1, β_2 and β_3 are dimensionless quantities that characterize the properties of the hysteretic behavior, sgn$(\cdot)$ is the sign function, and all other terms have been previously defined. The quantities β_1, β_1 and β_3 correspond to scale, loop fatness and loop pinching parameters, respectively. The above characterization of the hysteretic behavior corresponds to the Bouc-Wen type model [40]. The following values for the dissipation model parameters are used in this case: $k_e = 2.5 \times 10^6$ N/m; $U^y = 5 \times 10^{-3}$m; $\alpha = 0.1$; $\beta_1 = 1.5$; $\beta_2 = 0.5$; and $\beta_3 = 0.5$. A typical displacement-restoring force curve of one of the U-shaped flexural plates under seismic load is shown in Figure 10. The non-linear restoring force of each device acts between the floors where it is placed with the same orientation of the device.

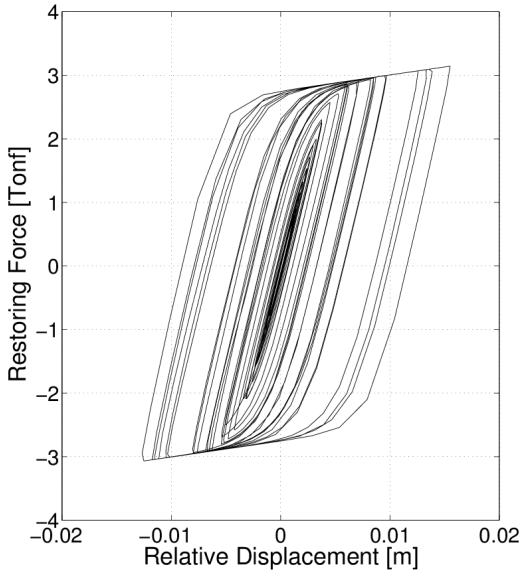

Figure 10: Typical displacement-restoring force curve of one of the U-shaped flexural plates

6.2. Excitation Model

A non-stationary ground acceleration process defined in terms of the Clough-Penzien power spectrum is used to generate synthetic ground motions in this application problem. Based on these records, a number of

scenarios with respect to the magnitude of missing data are constructed. The non-stationary ground acceleration process is characterized as $S_a(t,\omega) = h^2(t, M_m, r)\, S_a(\omega)$, where $h(\cdot)$ is an envelope function of time and $S_a(\cdot)$ is the Clough-Penzien power spectrum [41]. The envelope function suggested in [42] is considered in the present formulation. Such function is given by

$$h(t, M_m, r) = a_1 \left(\frac{t}{t_n}\right)^{a_2} \cdot \exp\left(-a_3 \cdot \frac{t}{t_n}\right) \tag{17}$$

where the parameters a_1, a_2 and a_3 are defined as

$$a_1 = \left(\frac{e}{\lambda}\right)^{a_2} \ , \ a_2 = \frac{-\lambda \ln(\eta)}{1 + \lambda \cdot (\ln(\lambda) - 1)} \ , \ a_3 = \frac{a_2}{\lambda} \tag{18}$$

The envelope function has a peak equal to unity at $t = \lambda \cdot t_n$ and equal to η at $t = t_n$ where $t_n = 2T_n$. The parameter T_n, which corresponds to the duration of ground motion, can be expressed as a sum of a path dependent and source dependent component [43]. More specifically, $T_n = 0.05\sqrt{r^2 + r_z^2} + \frac{0.5}{f_a}$ where r is the epicentral distance, r_z is the pseudo-depth given by $\log(r_z) = 0.15 \cdot M_m - 0.05$, where M_m is the moment magnitude, and f_a is the so-called corner frequency defined as $\log(f_a) = 2.181 - 0.496 \cdot M_m$ [44]. The values $\lambda = 0.2$, $\eta = 0.05$, $r = 25$ km, and $M_m = 7.0$ are considered in the present study. On the other hand, the Clough-Penzien power spectrum is characterized as

$$S_a(\omega) = S_0 \cdot \frac{\omega^4}{(\omega_f^2 - \omega^2)^2 + 4\,\xi_f^2\,\omega_f^2\,\omega^2} \frac{\omega_g^4 + 4\,\xi_g^2\,\omega_g^2\,\omega^2}{(\omega_g^2 - \omega^2)^2 + 4\,\xi_g^2\,\omega_g^2\,\omega^2} \tag{19}$$

where S_0 is the amplitude of the white noise bedrock excitation spectrum, ω_f, ξ_f, ω_g, and ξ_g are parameters related to soil conditions. The values $S_0 = 0.02[m^2/s^3]$, $\omega_f = 1.5[rad/s]$, $\xi_f = 0.6$, $\omega_g = 15.0[rad/s]$, and $\xi_g = 0.6$ are used in the current application. These values correspond to ground motions of regular intensity on firm soil conditions. The synthetic ground motions are generated in the time domain via Eq. 13 in which

$$S_X\left(\omega_j, t\right) = h(t, M_m, r)^2 S_a(\omega_j). \tag{20}$$

and where Eq. 13 is implemented with parameters $\Delta_\omega = 2.28$ and $N = 48$. The synthetic ground motions are discretized at time intervals equal to $\Delta t = 0.0293$s with a total duration of $T = 30.0$s given samples of length

equal to 1024 points. In [20], reconstruction of stationary and non-stationary power spectra via CS even without an adaptive basis procedure was shown to be robust in the presence of noise, due to a tolerance term, e in the $L1$ minimization procedure, shown in Eq. 7. Therefore, additional noise on the ground excitations is omitted for simplicity. The EPS and a sample realization are shown in Figures 11 and 12 respectively.

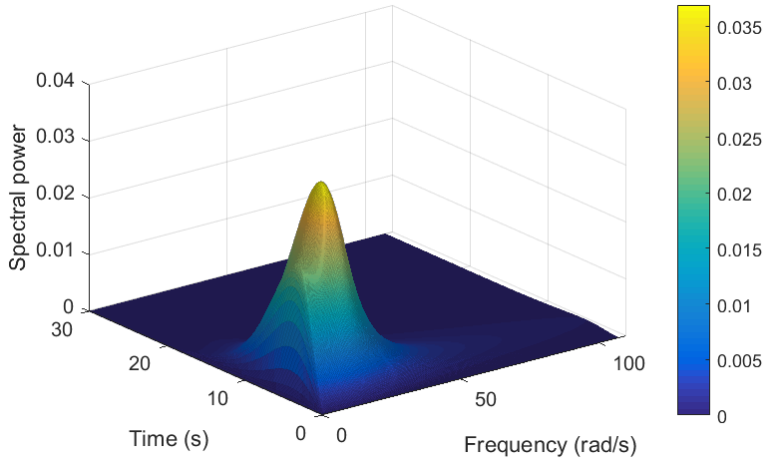

Figure 11: Evolutionary power spectrum based on the Clough-Penzien spectrum (Eq. 20)

6.3. Estimation of Power Spectrum

Recall that harmonic wavelets are used as a basis for the CS procedure, the coefficients of which may in turn be used to estimate the EPS. Due to the limitations previously discussed, such as time-frequency resolution trade-off and end-effects, the estimated EPS does not perfectly match the original evolutionary power spectrum. Further, the EPS estimation via the GHWT is drawn from expected values of wavelet coefficients as shown in Eq. 14. Hence, estimation accuracy is dependent upon the number of sample time histories available. To visualize the impact of these properties, the harmonic wavelet estimated EPS of the spectrum defined in Eq. 20 is shown in Figure 13 with a constant wavelet resolution, $n-m = 16$ and reverse signal padding on both sides totalling half the signal length for 80 time-histories. Although the shape is not as smooth, it is clear that Figure 13 is a good approximation

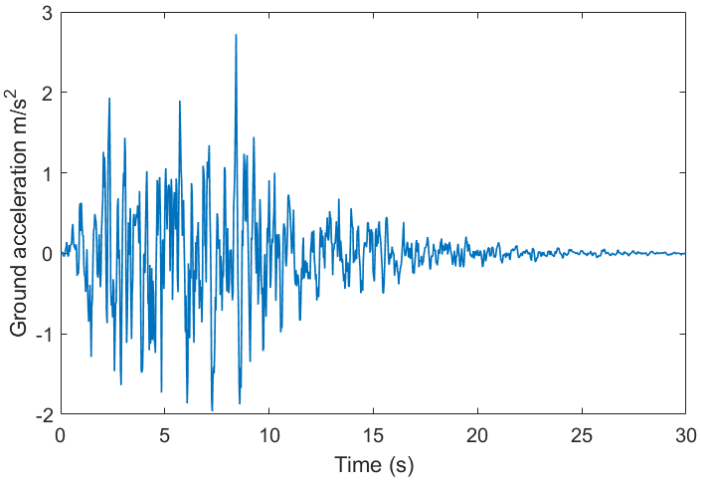

Figure 12: Time history sample realization of a ground motion obtained from the evolutionary power spectrum based on the Clough-Penzien spectrum (Eq. 20)

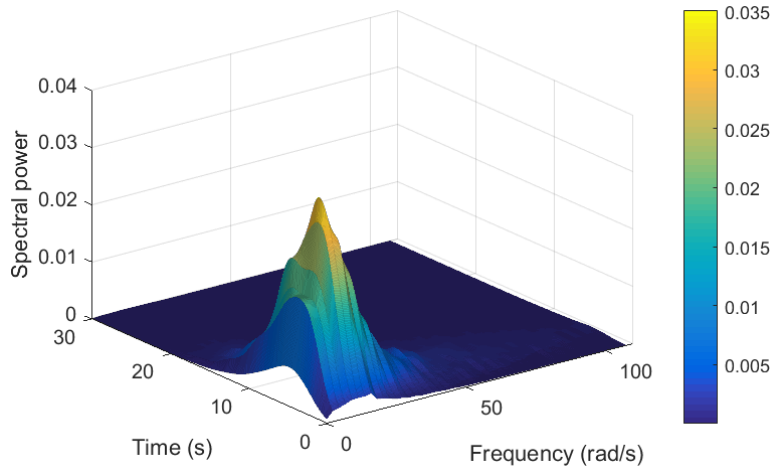

Figure 13: Harmonic wavelet based power spectrum estimation of the power spectrum defined in Eq. 20

to Figure 11, capturing the high frequency time-dependent tail and location of the peak power.

It is clear that even in the case where no missing data are present, the

choice of EPS estimation method will influence the results. Even within the presented CS framework, different resolutions and even alternative basis functions could have been utilized. The accuracy of such approaches in the context of EPS estimation has been the subject of extensive studies e.g. [36, 45, 46]. Therefore, it is important to note that the objective in this section is to assess the merit of the reconstruction technique through comparing estimated response statistics and system reliability. Hence the GHWT estimated EPS with no missing data is used as a control case against which the missing data reconstructions are compared, rather than against the original specified ground excitation spectrum model. In this manner, the effect of the EPS estimation is isolated for comparison purposes.

6.4. Simulation of Missing Data

Two different arrangements of missing data are considered. The first case simulates missing data at random locations while the second case simulates missing data that occur in groups (intervals) positioned at random locations. The random locations are defined by a procedure based on the generation of random numbers drawn from a uniform distribution of the time index [19]. In the context of the present application missing data is simulated for multiple process records used for the same spectral estimate. For the case of missing data at random locations the missing sample points are re-generated for each sample. However, for the case of missing data over intervals the missing data encompasses the intervals of all smaller interval cases for that set e.g. the 20% missing data case is an extension of missing data in the time domain from the 10% missing data case. Figures A.21 and A.20 show sample time histories with interval gaps with and without reconstructed data.

Sets of 80 time histories are generated from the aforementioned Clough-Penzien spectrum of length equal to 1024 points. From these sets of records, 8 separate configurations of missing data across the ensemble were simulated independently. In particular four cases, corresponding to 10%, 20%, 30 % and 40% of the data removed at random locations and over 10 constant intervals are considered. To evaluate the effectiveness of the reconstruction potential of CS, the estimated spectra corresponding to the cases of maximum missing data (40%) for the two different arrangements of missing data, i.e. randomly scattered and over 10 intervals, are shown in Figures 14 and 15, respectively. It is observed that even with a relatively large numbers of missing data the method produces non-stationary power spectra that fit very well with the shape of the target spectrum (Figure 13) over time as well as over the

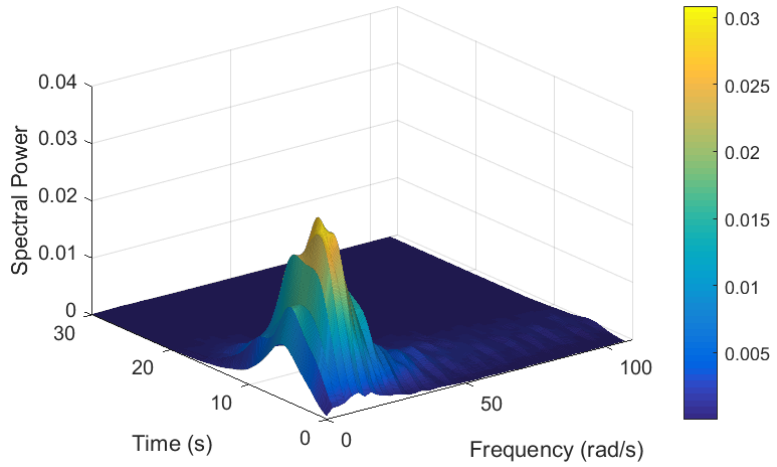

Figure 14: Harmonic wavelet based power spectrum estimation of the evolutionary power spectrum based on the Clough-Penzien spectrum (Eq. 20) with 40% missing data in uniformly distributed random locations

frequency domain. Based on the previous results it is seen that the spectra are reconstructed with sufficient accuracy.

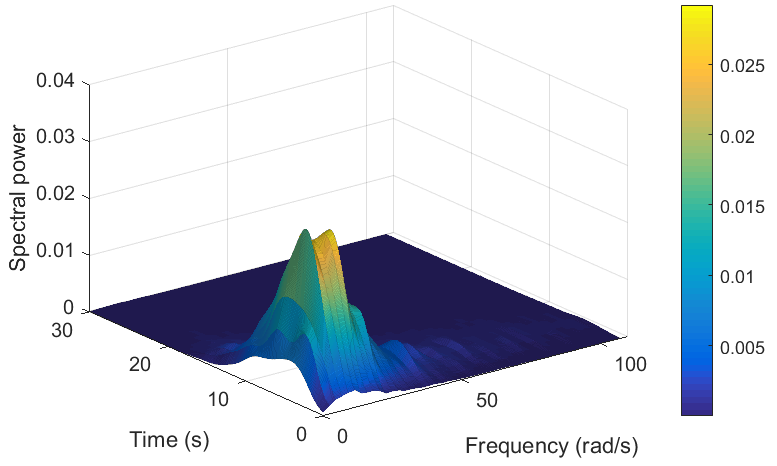

Figure 15: Harmonic wavelet based power spectrum estimation of the evolutionary power spectrum based on the Clough-Penzien spectrum (Eq. 20) with 40% missing data over 10 constant intervals

6.5. Effect on Response Statistics

The effect of estimated spectra on the response statistics of the structural system is investigated in this section. To this end the performance of the structural system is defined in terms of its response at 48 control points. The control points are located over the height of the structure at five different corners and at the center of mass of each floor. Some of these points, represented as dots, are shown in Figure 8. The responses to be monitored are the interstory drift ratios (absolute value of relative displacement between floors) in the x direction (δ_{x_i}) and y direction (δ_{y_i}) at the different control points ($i = 1, ..., 48$). Based on these responses a measure of the overall system performance r is defined. It is characterized as the maximum interstory drift response in time over the 48 control points, that is

$$r = \max_{t\in[0,T]} \max_{i=1,...,48} \{ \mid \delta_{x_i}(t, \boldsymbol{\phi}) \mid, \mid \delta_{y_i}(t, \boldsymbol{\phi}) \mid \} \tag{21}$$

where $\boldsymbol{\phi}$ is the vector of random variables involved in the characterization of the excitation (see Eq. 13). Note that the interstory drift ratios are functions of time (due to the dynamic nature of the excitation) and the random vector $\boldsymbol{\phi}$. For each arrangement of missing data and percentage of data removed synthetic seismic excitations are generated from the corresponding estimated spectra. Then, the overall system performance parameter r is computed for each ground motion generated. Based on a number of independent runs, 100 in this case, some statistics are estimated. In particular, the sample mean and standard deviation of the overall system performance parameter is considered in the present analysis. Tables 2 and 3 show the sample mean, standard deviation, and coefficient of variation of the overall system performance parameter obtained from the estimated spectra corresponding to the two arrangements of missing data defined in the previous section. That is, data removed in uniformly distributed random locations and over 10 constant intervals positioned at random locations.

Each table shows the results of the four cases corresponding to 10%, 20%, 30% and 40% of the data removed as well as the no-missing data case. Note that the no-missing data case corresponds to results obtained from the reconstructed spectrum defined in Eq. 20, and shown in Figure 13. For comparison purposes such spectrum is considered as the target spectrum. It is seen that the sample mean and standard deviation of the overall system performance parameter corresponding to the missing data cases are very similar to the

Data removed	Mean(cm)	Standard deviation(cm)	C.O.V
0%(no-missing data)	2.33	1.02	0.437
10%	2.33	1.02	0.437
20%	2.32	1.02	0.437
30%	2.31	1.01	0.436
40%	2.24	0.97	0.434

Table 2: Sample mean and standard deviation of the overall system performance parameter r. Case of data removed in uniformly distributed random locations.

ones obtained from the no-missing data case for both arrangements of missing data. In other words, the quality of the response statistics obtained from estimated spectra is very good as compared to the one obtained from the target spectrum (no-missing data case).

Data removed	Mean(cm)	Standard deviation(cm)	C.O.V
0%(no-missing data)	2.33	1.02	0.437
10%	2.42	1.06	0.438
20%	2.39	1.05	0.439
30%	2.37	1.04	0.438
40%	2.26	0.98	0.435

Table 3: Sample mean and standard deviation of the overall system performance parameter r. Case of data removed over 10 constant intervals positioned at random locations.

Based on the previous statistics the 5%, 68% and 95% quantiles of the mean value are estimated. Note that the 68% quantile corresponds to the interval defined by the mean value plus/minus one standard deviation. Figure 16 presents the 5%, 68% and 95% credible intervals for the mean of the overall system performance parameter. The four cases corresponding to 10%, 20%, 30% and 40% of the data removed at random locations as well as the no-missing data case are shown in the figure. In addition, the statistical mean is also indicated in the figure (represented as dot). The corresponding information for the cases of data removed over 10 constant intervals is given in Figure 17.

It is observed from both figures that the accuracy in the prediction of the 5%, 68% and 10% quantiles, compare with the one obtained from the no-

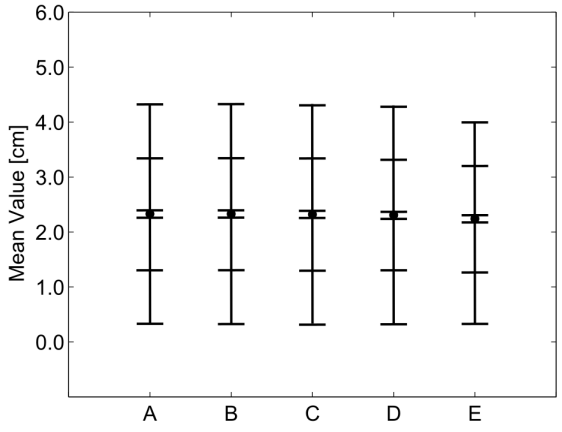

Figure 16: Statistics (sample mean, 5% quantile, 68% quantile and 95% quantile) for the mean of the overall system performance measure. A: no-missing data case. B,C,D, and E: 10%, 20%, 30% and 40% of missing data cases, respectively. Case of data removed in uniformly distributed random locations.

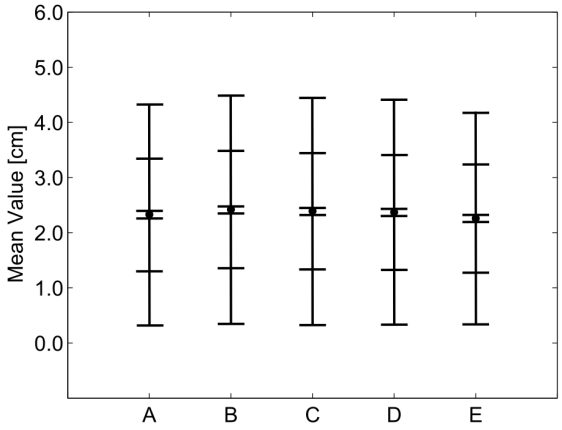

Figure 17: Statistics (sample mean, 5% quantile, 68% quantile and 95% quantile) for the mean of the overall system performance measure. A: no-missing data case. B,C,D, and E: 10%, 20%, 30% and 40% of missing data cases, respectively. Case of data removed over 10 constant intervals positioned at random locations

missing data case, is quite satisfactory even for high percentage of missing data (40%). Additional calculations have also shown the effectiveness of the proposed method for even more severe arrangements of missing data. For example an arrangement based on data removed over 5 constant intervals with 10%, 20%, 30% and 40% of missing data was also tested. The quality of the response statistics obtained from the estimated spectra was very good. Thus, the power spectrum reconstruction capability of the proposed methodology, in the context of response statistics, is apparent.

6.6. Effect on System Reliability

The impact of estimated spectra on the reliability of the structural system is considered in this section. For structural systems under stochastic excitation the probability that performance conditions are satisfied within a particular reference period T provides a useful reliability measure [47]. Such measure is referred as the first excursion probability. In this context, the failure event F is defined as $F = d(t, \boldsymbol{\phi}) > 1$, where d is the so-called normalized demand function defined as

$$d(t, \boldsymbol{\phi}) = \max_{t\in[0,T]} \max_{i=1,\ldots,48} \{ \frac{\mid \delta_{x_i}(t, \boldsymbol{\phi}) \mid}{\delta^*}, \frac{\mid \delta_{y_i}(t, \boldsymbol{\phi}) \mid}{\delta^*} \} \tag{22}$$

where as previously pointed out $\delta_{x_i}(t, \boldsymbol{\phi})$ and $\delta_{y_i}(t, \boldsymbol{\phi})$ represent the interstory drift response in the x and y direction at control point i, respectively, and δ^* is the corresponding acceptable response level (threshold). Note that there are 96 response functions associated with the failure event F. The probability of failure is formally defined as

$$P_F = P[\max_{t\in[0,T]} \max_{i=1,\ldots,48} \{ \frac{\mid \delta_{x_i}(t, \boldsymbol{\phi}) \mid}{\delta^*}, \frac{\mid \delta_{y_i}(t, \boldsymbol{\phi}) \mid}{\delta^*} \} > 1] \tag{23}$$

where $P[\cdot]$ is the probability that the expression in parenthesis is true. Equivalently, the failure probability can be written in terms of the multidimensional probability integral

$$P_F = \int_{d(t,\boldsymbol{\phi})>1} p(\boldsymbol{\phi}) d\boldsymbol{\phi} \tag{24}$$

where $p(\boldsymbol{\phi})$ is the probability density function that characterized the random variables $\boldsymbol{\phi}$. Note that the corresponding reliability problem is a high dimensional problem since there are 48 random variables that characterize the

excitation (number of random variables in Eq. 13). The failure probability is estimated by Subset simulation [48, 49]. In this well known advanced simulation technique the failure probabilities are expressed as a product of conditional probabilities of some chosen intermediate failure events, the evaluation of which only requires simulation of more frequent events. The intermediate failure events are chosen adaptively using information from simulated samples so that they correspond to some specified values of conditional failure probabilities. Therefore, a rare event simulation problem is converted into a sequence of more frequent event simulation problems. The method uses a Markov chain Monte Carlo method based on the Metropolis algorithm for sampling from the conditional probabilities [50]. This is the most widely applicable simulation technique because it is not based on any geometrical assumption about the topology of the failure domain. In fact, validation calculations have shown that subset simulation can be applied efficiently to a wide range of dynamical systems including general linear and non-linear systems [49, 51, 52]. For a detailed description of subset simulation the reader is referred to [48, 49]

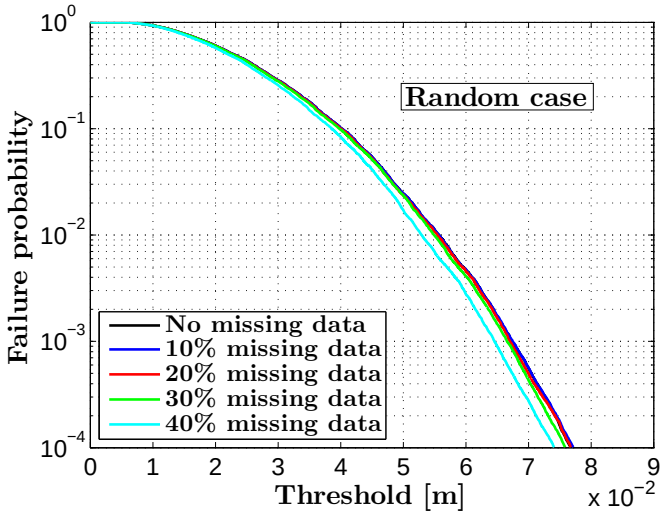

Figure 18: Failure probability in terms of the threshold level for different estimated spectra. Case of data removed in uniformly distributed random locations

Figure 18 shows the failure probability in terms of the response level (threshold) for the cases of 10%, 20%, 30% and 40% of the data removed at

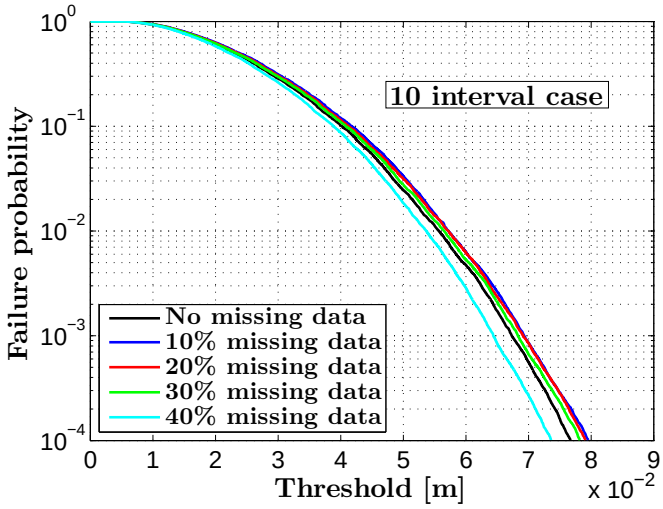

Figure 19: Failure probability in terms of the threshold level for different estimated spectra. Case of data removed over 10 constant intervals positioned at random locations

random locations. Also the results obtained from the no-missing data case are shown in the figure for comparison purposes. Similar information for the cases of 10%, 20%, 30% and 40% of the data removed over 10 constant intervals is shown in Figure 19. The different curves of the figures correspond to an average of 10 independent runs of subset simulation. It is seen that the reliability curves corresponding to the missing data cases almost coincide with the reliability curve of the no-missing data case (target curve) for both arrangements of missing data. This coincidence is particularly evident for the cases corresponding to 10%, 20% and 30% of missing data. Thus, the probability estimates corresponding to the missing data cases with missing data up 30% are quite accurate even for low failure probabilities, i.e less than 10^{-4}. For the case of 40% of data removed some minor differences are observed with respect to the target failure probability curve. Further, it can be seen that with some of the cases, the failure probability is over-estimated and in others under-estimated. This can be attributed in part to random effects contributed by the arrangement of missing data and nature of the specific generated signals. However, as the number of missing data becomes more significant in comparison to the sparsity of the process in the chosen wavelet basis, lower power estimates, comprising fewer wavelets will become more

likely in the proposed CS reconstruction framework. Realizations generated from these estimated spectra are in turn more likely to lead to lower failure probabilities. This can be observed by the fact that both of the 40% cases under-estimate failure probability.

As previously pointed out, the uncertainty involved in the problem is due to the characterization of the excitation model. In this regard, the potential effect of system uncertainty on the results is independent of the spectrum reconstruction process. In this situation, the effect of system variability should be considered explicitly in the estimation of the system reliability (or response statistics) for all cases, i.e. missing and no missing data cases. Thus, for comparison purposes such effect would be unimportant.

Based on the results obtained in this section it is concluded that the performance of the proposed methodology in treating missing data is quite satisfactory in the context of reliability analysis for intermediate levels of missing data. For more severe arrangements of missing data validation calculations have shown that the reconstructed spectra have some difficulties in reproducing reliability estimates with sufficient accuracy. These limitations are primarily due to poor spectrum estimates occurring as a result of the same difficulties previously discussed for the 40% missing data case. These issues with higher numbers of missing data call for further examination and enhancement of the proposed methodology in connection with reliability analysis. Such study is topic for future research.

7. Conclusions

The issue of determining structural system response and reliability statistics subject to excitation realizations with missing data has been addressed. To this aim, a compressive sensing based framework coupled with an adaptive basis has been presented for reconstructing the samples with missing data and estimating the underlying process. In this regard, novel insights have been provided whereas certain conceptual, numerical, and practical implementation aspects of the technique have been presented in detail.

Numerical results pertaining to the stochastic response and reliability analysis of a large-scale structural system have shown that the response statistics obtained by utilizing estimated spectra compare very well with results obtained by using the target spectrum (no-missing data case). Likewise, the reliability assessment with missing data is in very good agreement with the results derived from the target power spectrum. Based on these

results it is concluded that CS with adaptive basis is potentially an effective tool for performing stochastic response and reliability analyses of real-size structures under incomplete earthquake records with intermediate levels of missing data. Note that the main assumption of the methodology relates to the available records exhibiting relative sparsity on the harmonic wavelet domain. The effectiveness of the proposed approach in the context of more severe arrangements of missing data, including long gaps, as well as in the presence of real records with missing data is left for future research efforts.

Acknowledgments

The research reported here was supported in part by CONICYT under grant number 1150009 which is gratefully acknowledged by the authors.

References

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

[2] J. Chen, W. Sun, J. Li, J. Xu, Stochastic harmonic function representation of stochastic processes, Journal of Applied Mechanics 80 (1) (2013) 011001.

[3] J. Chen, J. Li, Optimal determination of frequencies in the spectral representation of stochastic processes, Computational Mechanics 51 (5) (2013) 791–806.

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

[5] 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.

[6] I. A. Kougioumtzoglou, Stochastic joint time-frequency response analysis of nonlinear structural systems, Journal of Sound and Vibration 332 (2013) 7153–7173.

[7] P. D. Spanos, F. Kong, J. Li, I. A. Kougioumtzoglou, Harmonic wavelets based excitation-response relationships for linear systems: A critical perspective, Probabilistic Engineering Mechanics (2016) doi:10.1016/j.probengmech.2015.09.021 (In Press).

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

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

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

[11] 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.

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

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

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

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

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

[17] 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.

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

[19] 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.

[20] L. Comerford, I. A. Kougioumtzoglou, M. Beer, Compressive sensing based stochastic process power spectrum estimation subject to missing data, Probabilistic Engineering Mechanics 44 (2016) 66–76.

[21] L. A. Comerford, M. Beer, I. A. Kougioumtzoglou, 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.

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

[25] Y. C. Eldar, G. Kutyniok, Compressed sensing: theory and applications, Cambridge University Press, 2012.

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

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

[28] 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.

[29] 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.

[30] 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.

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

[32] 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.

[33] 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.

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

[35] 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.

[36] 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.

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

[38] 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.

[39] H. Jensen, J. Sepulveda, On the reliability-based design of structures including passive energy dissipation systems, Structural Safety 34 (2011) 390–400.

[40] T. Baber, Y. Wen, Random vibration hysteretic, degrading systems., Journal of Engineering Mechanics 107 (6) (1981) 1069–1087.

[41] R. Clough, J. Penzien, Dynamics of structures., McGraw-Hill, New York 107 (6) (1975) 1069–1087.

[42] G. Saragoni, G. Hart, Simulation of artificial earthquakes., Earthquake Engineering and Structural Dynamics 2 (3) (1974) 249–267.

[43] D. Boore, Simulation of ground motion using the stochastic method., Pure and Applied Geophysics 160 (3-4) (2003) 635–676.

[44] G. Atkinson, W. Silva, Stochastic modeling of california ground motions., Bulletin of the Seismological Society of America 90 (2) (2000) 255–274.

[45] P. D. Spanos, G. Failla, Evolutionary spectra estimation using wavelets, Journal of Engineering Mechanics-Asce 130 (8) (2004) 952–960, pT: J; TC: 16; UT: WOS:000223001200010.

[46] I. A. Kougioumtzoglou, F. Kong, P. D. Spanos, J. Li, Some observations on wavelets based evolutionary power spectrum estimation, in: Proceedings of the Stochastic Mechanics Conference (SM12), Ustica, Italy, 2012, pp. vol. 3: 37–44, ISSN: 2035–679X.

[47] M. Grigoriu, T. Soong, Random vibration of mechanical and structural systems (1993).

[48] S.-K. Au, J. L. Beck, Estimation of small failure probabilities in high dimensions by subset simulation, Probabilistic Engineering Mechanics 16 (4) (2001) 263–277.

[49] K. M. Zuev, J. L. Beck, S.-K. Au, L. S. Katafygiotis, Bayesian post-processor and other enhancements of subset simulation for estimating failure probabilities in high dimensions, Computers & structures 92 (2012) 283–296.

[50] N. Metropolis, S. Ulam, The monte carlo method, Journal of the American statistical association 44 (247) (1949) 335–341.

[51] J. Ching, S.-K. Au, J. L. Beck, Reliability estimation for dynamical systems subject to stochastic excitation using subset simulation with splitting, Computer methods in applied mechanics and engineering 194 (12) (2005) 1557–1579.

[52] S.-K. Au, Y. Wang, Engineering risk assessment with subset simulation, John Wiley & Sons, 2014.

Appendix A. Example missing data records

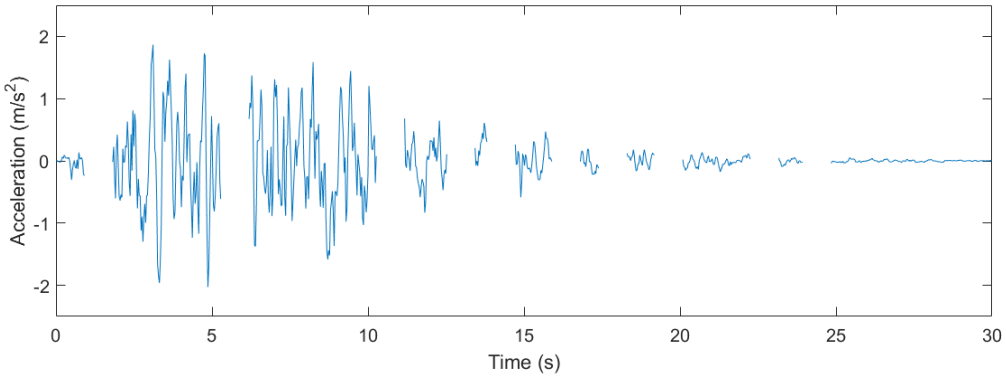

Figure A.20: Example earthquake realization with 30% missing data on 10 randomly positioned intervals

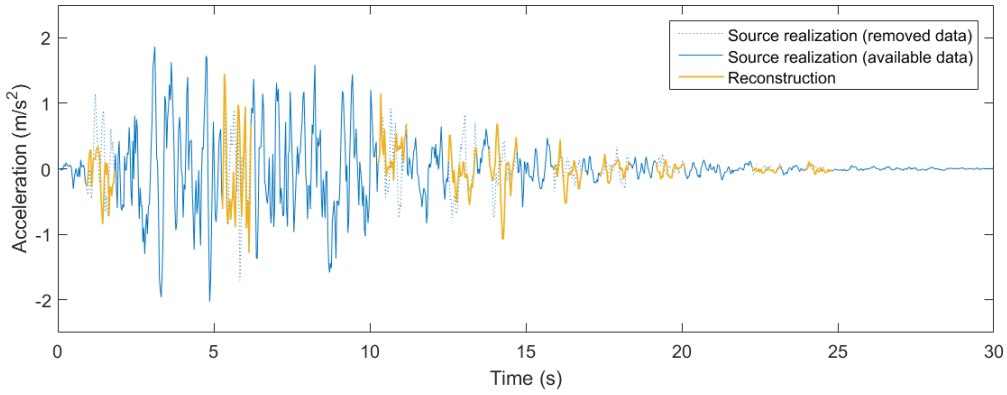

Figure A.21: Example earthquake realization with 30% missing data on 10 randomly positioned intervals including source and reconstructed data