Probabilistic Engineering Mechanics 44 (2016) 66–76

Contents lists available at ScienceDirect

Probabilistic Engineering Mechanics

journal homepage: www.elsevier.com/locate/probengmech

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

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

[a] Faculty of Civil Engineering and Geodetic Science, Leibniz University Hannover, Callinstr. 34, 30167 Hannover, Germany
[b] Institute for Risk and Uncertainty, University of Liverpool, Liverpool L69 3GH, UK
[c] Department of Civil Engineering and Engineering Mechanics, Columbia University, New York, NY 10027, USA

ARTICLE INFO

Article history:
Received 17 August 2015
Accepted 29 September 2015
Available online 1 October 2015

Keywords:
Compressive sensing
Stochastic process
Missing data

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

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 [4–6] 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

* Corresponding author.
E-mail addresses: l.comerford@liv.ac.uk (L. Comerford), iak2115@columbia.edu (I.A. Kougioumtzoglou), m.beer@liv.ac.uk (M. Beer).

http://dx.doi.org/10.1016/j.probengmech.2015.09.015

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 [12–16] or chirplet (e.g. [17]) transforms, the short-time Fourier transform, Gabor transform (e.g. [12]) and Wigner–Ville distribution (e.g. [18–20]). 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. [24–26])

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

where $Z(\omega)$ has the properties

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

and

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

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 $0 \le \Phi_k < 2\pi$. 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. [30–32]).

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

$$\Psi^G_{(m,n),k}(\omega) = \begin{cases} \dfrac{1}{(n-m)\Delta\omega} \exp\left(-i\omega\dfrac{kT_0}{n-m}\right), & m\Delta\omega \le \omega \le n\Delta\omega \\ 0 & \text{otherwise}, \end{cases} \tag{6}$$

where m, n and k are considered to be positive integers and $\Delta\omega = 2\pi/T_0$; 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

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

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

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

Eq. (10) represents a localized process at scale (m, n) and translation (k) defined in the intervals $[m\Delta\omega, n\Delta\omega]$ and $[kT_0/(n-m), (k+1)T_0/(n-m)]$, whereas $S_{(m,n),k}$ 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])