Wind data extrapolation and stochastic field statistics estimation via compressive sampling and low rank matrix recovery methods

George D. Pasparakis[a], Ketson R.M. dos Santos[b], Ioannis A. Kougioumtzoglou[c,*], Michael Beer[a,d,e]

[a]*Institute for Risk and Reliability, Leibniz Universität Hannover*
[b]*Earthquake Engineering and Structural Dynamics Laboratory (EESD), École Polytechnique Fédérale de Lausanne (EPFL)*
[c]*Department of Civil Engineering and Engineering Mechanics, Columbia University*
[d]*Institute for Risk and, Uncertainty and School of Engineering, University of Liverpool*
[e]*International Joint Research Center for Engineering Reliability and Stochastic Mechanics, Tongji University*

Abstract

A methodology based on compressive sampling is developed for incomplete wind time-histories reconstruction and extrapolation in a single spatial dimension, as well as for related stochastic field statistics estimation. This relies on l_1-norm minimization in conjunction with an adaptive basis re-weighting scheme. Indicatively, the proposed methodology can be employed for monitoring of wind turbine systems, where the objective relates to either reconstructing incomplete time-histories measured at specific points along the height of a turbine tower, or to extrapolating to other locations in the vertical dimension where sensors and measurement records are not available. Further, the methodology can be used potentially for environmental hazard modeling within the context of performance-based design optimization of structural systems.

Unfortunately, a straightforward implementation of the aforementioned approach to account for two spatial dimensions is hindered by significant, even prohibitive in some cases, computational cost. In this regard, to address computational challenges associated with higher-dimensional domains, a methodology based on low rank matrices and nuclear norm minimization is developed next

*Corresponding author
Email address: iak2115@columbia.edu (Ioannis A. Kougioumtzoglou)

for wind field extrapolation in two spatial dimensions. The efficacy of the proposed methodologies is demonstrated by considering various numerical examples. These refer to reconstruction of wind time-histories with missing data compatible with a joint wavenumber-frequency power spectral density, as well as to extrapolation to various locations in the spatial domain.

Keywords: Wind data, Stochastic field, Sparse representations, Compressive sampling, Low-rank matrix.

1. Introduction

Estimating wind field model related statistics relies, typically, on information provided by data acquisition systems such as distributed sensor networks and LIDAR acquisition systems (e.g., [1], [2]). In many real-life cases, however, the measured data are corrupted and incomplete. Also, it is often required to extrapolate relevant wind field information to points of interest, where there are no measurements due to limited equipment availability. Clearly, developing methodologies for accurate reconstruction and extrapolation of wind field data is of paramount importance to the analysis, design and monitoring of engineering systems such as wind turbines; see, for instance, [3]. Further, it is worth noting that such methodologies can be used potentially for environmental hazard modeling within the context of performance-based design optimization of structural systems (e.g., [4], [5]).

Indicatively, a wavelet-based fluid motion estimator was developed in [6] for estimating wind fields based on backscatter data. Further, a dimension reduction approach based on computational fluid dynamics data was applied in [7] for wind field reconstruction. Also, several machine learning approaches based on various neural network implementations and configurations were employed recently for wind data reconstruction and extrapolation (e.g., [8–10]). Moreover, surrogate modeling based on Kriging was proposed in [11], whereas Kalman filtering was used in [12] for wind field estimation based on a limited number of LIDAR measurements.

Nevertheless, most of the aforementioned approaches are characterized by significant limitations. For example, in many cases the techniques appear efficacious only for relatively small percentages of missing data, whereas results based on black-box approaches such as neural networks are not always interpretable. Alternatively, various methodologies based on compressive sampling (CS) have been developed recently, which appear promising for stochastic process statistics esti-

mation based on realizations with incomplete/missing data (e.g., [13–15]). The interested reader is also directed to the recent review paper by Kougioumtzoglou et al. [16] for a broad perspective on theoretical concepts and diverse applications of sparse representations and CS approaches in engineering mechanics.

In this paper, a methodology based on CS is developed, which relies on l_1-norm minimization in conjunction with an adaptive basis re-weighting scheme, for incomplete wind field time-histories reconstruction and extrapolation in a single spatial (vertical) dimension. Next, to address computational challenges associated with higher-dimensional domains, a methodology based on low rank matrices and nuclear norm minimization is developed for wind field extrapolation in two spatial dimensions. The efficacy of the proposed methodologies is demonstrated by considering various numerical examples. These refer to reconstruction of wind time-histories with missing data compatible with a joint wavenumber-frequency power spectral density (PSD), as well as to extrapolation to various locations in the spatial domain.

2. Wind field spectral representation

In various engineering applications, the wind field can be conveniently modeled as a stochastic wave (e.g., [17]). This facilitates the efficient simulation of realizations corresponding to a large number of points in the spatial domain, while circumventing the need for cross-PSD related calculations; see also [18]. In this section, following closely [19], the basic elements associated with Monte Carlo simulation (MCS) of a wind field compatible with a joint wavenumber-frequency PSD are reviewed for completeness.

2.1. Wind field time-histories simulation in a single spatial dimension

A homogeneous wind field in the vertical dimension is related to a wavenumber-frequency PSD (e.g., [17], [19]) given by

$$S^{(\mathrm{WF})}(k,\omega) = \frac{1}{2\pi}\int_{-\infty}^{\infty} S^{(\mathrm{F})}(\xi,\omega)e^{-ik\xi}\mathrm{d}\xi \tag{1}$$

where ω, ξ, k denote the frequency, spatial distance and wavenumber, respectively. Further,

$$S^{(\mathrm{F})}(\xi,\omega) = S_0(\omega)\rho(\xi,\omega) = S_0(\omega)\exp\left(-\frac{C_z}{2\pi U_{10}}|\omega||\xi|\right) \tag{2}$$

where $S_0(\omega)$ is the auto-PSD and $\rho(\xi,\omega)$ represents the coherence function. Combining Eq. (1) and Eq. (2), $S^{(\mathrm{WF})}(k,\omega)$ takes the form

$$S^{(\mathrm{WF})}(k,\omega) = S_0(\omega)\rho^{(\mathrm{WF})}(k,\omega) = S_0(\omega)\frac{C_z}{2\pi^2 U_{10}}\frac{|\omega|}{\left(\frac{C_z}{2\pi U_{10}}\right)^2\omega^2 + k^2} \tag{3}$$

where C_z is an exponential decay coefficient and U_{10} is the mean wind velocity at a height of 10m. Next, Eq. (3) can be used in conjunction with the spectral representation method (SRM) [20] for generating wind velocity records in the form

$$\begin{aligned} X(z,t) = \sum_{i=1}^{N_k}\sum_{j=1}^{N_\omega}\sqrt{4S^{(\mathrm{WF})}\left(k_i,\omega_j\right)\Delta k\Delta\omega} \\ \times\left[\cos\left(k_i z + \omega_j t + \varphi_{ij}\right) + \cos\left(k_i z - \omega_j t + \tilde{\varphi}_{ij}\right)\right] \end{aligned} \tag{4}$$

where φ_{ij} and $\tilde{\varphi}_{ij}$ represent two sets of independent random phase angles uniformly distributed over $[0, 2\pi]$; $k_i = i\Delta k$ with $\Delta k = k_u/N_k$ denotes the discretized wavenumber domain with an upper cut-off wavenumber k_u; and $\omega_j = j\Delta\omega$ with $\Delta\omega = \omega_u/N_\omega$ is the discretized frequency domain with an upper cut-off frequency ω_u. In the ensuing analysis and numerical examples, the Davenport PSD is considered (e.g. [20], [21]); that is,

$$S^{\text{Davenport}}(\omega) = 2.0u_*^2\frac{\left(\frac{1200}{2\pi U_{10}}\omega\right)^2}{|\omega|\left(1+\left(\frac{1200}{2\pi U_{10}}\omega\right)^2\right)^{4/3}} \tag{5}$$

where u_* denotes the shear flow velocity. Further, the parameters values used are $U_{10} = 31.88$ m/s, $u_* = 1.691$ m/s and $C_z = 10$.

2.2. Wind field time-histories simulation in two spatial dimensions

In this section, following closely [19], a generalization of the results outlined in section 2.1 is presented to account for a two-dimensional spatial domain. In

this regard, the joint wavenumber-frequency PSD of Eq. (3) takes the form

$$\begin{aligned} S^{(\mathrm{WF})}\left(k_y, k_z, \omega\right) &= S^{\text{Davenport}}\left(\omega\right) \cdot \rho^{(\mathrm{WF})}\left(k_y, k_z, \omega\right) \\ &= \frac{u_*^2}{\pi C_{1z} C_{1y} \left(\frac{1}{2\pi U_{10}}|\omega|\right)^2} \frac{\left(\frac{1200}{2\pi U_{10}}\omega\right)^2}{|\omega| \left(1 + \left(\frac{1200}{2\pi U_{10}}\omega\right)^2\right)^{4/3}} \\ &\times \frac{1}{\left(1 + \left[\left(\frac{1}{C_{1y}}k_y\right)^2 + \left(\frac{1}{C_{1z}}k_z\right)^2\right] / \left(\frac{1}{2\pi U_{10}}|\omega|\right)^2\right)^{\frac{3}{2}}} \end{aligned} \tag{6}$$

where

$$\begin{aligned} \rho^{(\mathrm{WF})}\left(k_y, k_z, \omega\right) &= \frac{1}{2\pi C_{1z} C_{1y}} \frac{1}{\left(\frac{1}{2\pi U_{10}}|\omega|\right)^2} \\ &\times \frac{1}{\left(1 + \left[\left(\frac{1}{C_{1y}}k_y\right)^2 + \left(\frac{1}{C_{1z}}k_z\right)^2\right] / \left(\frac{1}{2\pi U_{10}}|\omega|\right)^2\right)^{\frac{3}{2}}} \end{aligned} \tag{7}$$

In Eqs. (6)-(7), C_{1z} and C_{1y} are the exponential decay coefficients corresponding to the vertical and horizontal directions, respectively, and k_z and k_y denote the respective wavenumbers. Next, realizations compatible with the PSD of Eq. (6) can be generated based on the SRM (e.g., [17], [19]). In this context, Eq. (4) becomes

$$\begin{aligned} X(z,y,t) = &\sum_{i=1}^{N_{k_z}} \sum_{j=1}^{N_{k_y}} \sum_{m=1}^{N_\omega} \sqrt{4 S^{(\mathrm{WF})}\left(k_i^{(z)}, k_j^{(y)}, \omega_m\right) \Delta k_i^{(z)} \Delta k_j^{(y)} \Delta \omega^{(m)}} \\ &\cdot \Big[\cos\left(k_i^{(z)} z + k_j^{(y)} y + \omega_m t + \varphi_{ijm}^{(1)}\right) \\ &+ \cos\left(k_i^{(z)} z + k_j^{(y)} y - \omega_m t + \varphi_{ijm}^{(2)}\right) \\ &+ \cos\left(k_i^{(z)} z - k_j^{(y)} y + \omega_m t + \varphi_{ijm}^{(3)}\right) \\ &+ \cos\left(k_i^{(z)} z - k_j^{(y)} y - \omega_m t + \varphi_{ijm}^{(4)}\right)\Big] \end{aligned} \tag{8}$$

$k_j^{(y)} = j\Delta k_y, j = 1, 2, \ldots, N_{k_y}$ and $k_i^{(z)} = i\Delta k_z, i = 1, 2, \ldots, N_{k_z}$ are the discretized wavenumber domains in y and z directions with a number of points N_{k_y}

and N_{k_z}, respectively. Further, $\varphi_{ijm}^{(1)}$, $\varphi_{ijm}^{(2)}$, $\varphi_{ijm}^{(3)}$ and $\varphi_{ijm}^{(4)}$ represent four different sets of independent random phase angles uniformly distributed in $[0, 2\pi]$.

Regarding computational implementation aspects, it is readily seen that the Davenport PSD of Eq. (5) exhibits a singularity at the origin, which can be addressed, however, based on a frequency shift scheme (e.g., [17], [22]). In this regard, it is reasonable to consider an uneven discretization scheme, which is denser near the origin of the wavenumber domains.

3. Wind field reconstruction and extrapolation in the joint space-time domain: A compressive sampling treatment

Research efforts during the past fifteen years have focused on identifying and exploiting low-dimensional representations of high-dimensional data, as well as on establishing conditions guaranteeing unique representation in the low-dimensional space (e.g., [23], [24]). These theoretical results, coupled with potent convex optimization numerical algorithms, have triggered the birth of the currently expanding field of CS and have led to numerous impactful contributions in a wide range of application areas (e.g., [25]). The interested reader is also directed to the recent review paper by Kougioumtzoglou et al. [16] and to references therein for a broad perspective on theoretical concepts and diverse applications of sparse representations and CS approaches in engineering mechanics.

In this section, first, a CS approach based on l_1-norm minimization in conjunction with an adaptive basis re-weighting scheme is developed for wind field time-histories reconstruction and extrapolation in a single spatial dimension. Second, to address computational challenges associated with multi-dimensional domains, a CS approach based on nuclear norm minimization is developed for wind field extrapolation in two spatial dimensions.

3.1. Reconstruction and extrapolation in a single spatial dimension: l_1-norm minimization in conjunction with an adaptive basis re-weighting scheme

3.1.1. Theoretical aspects

In this section, the CS-based methodology developed in [26] for extrapolation of random wave field data in the joint space-time domain is adapted and extended to account for wind stochastic field extrapolation. Indicatively, this methodology can be employed for cases of monitoring wind turbine systems, where the objective is to either reconstruct incomplete time-histories measured at specific points along the height of a turbine tower, or to extrapolate to other locations in the vertical dimension where sensors and measurement records are not available.

Consider an $n_0 \times 1$ column vector $\mathbf{y}_0$ denoting a measured time history, which can be expanded by employing a basis matrix $\mathbf{A}_0$ of dimensions $n_0 \times n_0$ in the form $\mathbf{y}_0 = \mathbf{A}_0\mathbf{x}$, where $\mathbf{x}$ is the corresponding coefficient vector. Clearly, the sparsity degree of the coefficient vector $\mathbf{x}$ (i.e., number of non-zero elements) depends on the choice of the basis matrix $\mathbf{A}$ (e.g, polynomial, Fourier, etc). Resorting to a CS-based solution approach and considering a relatively high degree of sparsity, the coefficient vector $\mathbf{x}$ can be determined with satisfactory accuracy even if the system of equations $\mathbf{y}_0 = \mathbf{A}_0\mathbf{x}$ is underdetermined. Specifically, consider an $(n_0 - n_m) \times 1$ column vector $\mathbf{y}$ representing an under-sampled time history at a specific location along the height of a wind turbine. n_0 denotes the original sample and n_m is the number of randomly missing data. Further, considering a sampling matrix $\mathbf{A}$ of dimensions $(n_0 - n_m) \times n_0$ leads to the underdetermined system of equations $\mathbf{y} = \mathbf{A}\mathbf{x}$, where the objective relates to determining the coefficient vector $\mathbf{x}$ assumed to be sparse. According to CS theory (e.g., [16]), the problem can be cast in a convex optimization setting via minimizing the l_1-norm of $\mathbf{x}$; that is,

$$\mathbf{min} \mid\mid \mathbf{x} \mid\mid_{l_1} \text{ subject to } \mathbf{y} = \mathbf{A}\mathbf{x} \tag{9}$$

Clearly, the use of the l_1-norm promotes sparsity, whereas Eq. (9) can be readily solved by standard gradient optimization algorithms (e.g., [16]). Alternatively, the constraint in Eq. (9) can be relaxed and replaced by $\mid\mid \mathbf{y} - \mathbf{A}\mathbf{x} \mid\mid_{l_2} \leq \varepsilon$ to account for possible presence of noise.

Nevertheless, as also shown in [27], [4], [15], an adaptive basis re-weighting scheme can further promote sparsity and yield solution estimates of enhanced accuracy. In this regard, Eq. (9) becomes

$$\mathbf{min} \mid\mid \mathbf{x} \mid\mid_{l_1} \text{ subject to } \mathbf{y} = \mathbf{A}\mathbf{W}\mathbf{x} \tag{10}$$

where $\mathbf{W}$ is a re-weighting diagonal matrix. The rationale of the scheme relates to matrix $\mathbf{W}$ being used to appropriately weigh the columns of the basis matrix $\mathbf{A}$. To this aim, the entries of $\mathbf{W}$ correspond to the magnitudes of the components of $\mathbf{x}$. This promotes sparsity as it reduces the contribution of the smaller components of vector $\mathbf{x}$. Concisely, the mechanization of the re-weighting scheme is shown in Fig. (1), whereas the interested reader is directed to [4], [15] for a detailed presentation and discussion.

Clearly, the scheme is best suited to problems where an ensemble of time-histories with incomplete data points is available and the objective relates to determining statistics based on the ensemble average, such as estimating the underlying process PSD. Nevertheless, even in cases where only a single (relatively

• Initialize re-weighting matrix $\mathbf{W}, \mathbf{W}_2 = \mathbf{I}_{n_0}$ where $\mathbf{I}$ is the identity matrix
while $\mid \mathbf{W}_2 - \mathbf{W} \mid > threshold$ **do**
$\mathbf{W} = \mathbf{W}_2$
$\mathbf{W}_2 = \text{zeros}(n_0, n_0)$
• Generate re-weighting matrix via least squares
for $i = 1$ to m **do** $\triangleright$ m = number of available time-histories
$\mathbf{x} = (\mathbf{AW})^{\mathrm{T}} \left((\mathbf{AW})(\mathbf{AW})^{\mathrm{T}}\right) \mathbf{y}_i$
$\mathbf{x} = [\mathrm{x}_1, \mathrm{x}_2, \mathrm{x}_3, \ldots, \mathrm{x}_{n_0-1}, \mathrm{x}_{n_0}]$
$\mathbf{W}_2 = \mathbf{W}_2 + \text{diag}\big(\big[|| \mathrm{x}_2, \mathrm{x}_1 ||, || \mathrm{x}_2, \mathrm{x}_1 ||, || \mathrm{x}_3, \mathrm{x}_2 ||, || \mathrm{x}_3, \mathrm{x}_2 ||, \ldots, || \mathrm{x}_{n_0/2}, \mathrm{x}_{n_0/2-1} ||, || \mathrm{x}_{n_0/2}, \mathrm{x}_{n_0/2-1} ||\big]\big)$
end for
$\mathbf{W}_2 = \dfrac{\mathbf{W}_2}{\text{Mean}(\mathbf{W}_2)} + \text{bias} \cdot I_{n_0}$
end while
• Use l_1 minimization to compute the coefficient vector $\mathbf{x}$
$\mathbf{x} = \min \mid \mathbf{x} \mid_{l_1}$ subject to $\mathbf{y} = \mathbf{AW}_2\mathbf{x}$

Fig. 1: Mechanization of l_1-norm minimization with an adaptive basis

long) time-history is available, the scheme can be still implemented under the assumptions of stationarity and ergodicity by considering a partition of the record into smaller time intervals.

Further, it is shown that the above methodology can be readily adapted to be used for extrapolating wind time-histories corresponding to specific points along the vertical dimension based on measured data at neighboring locations. Specifically, consider a number of $N + M$ wind velocity records, each composed of length n data points and corresponding to a vertical height z_i. The objective refers to inferring wind velocity time histories at M distinct points from N measurement locations as depicted in Fig. (2). In this regard, the measured records can be represented as a column vector $\mathbf{y}$ of dimension $nN \times 1$ to be used in the formulation of Eq. (9), or alternatively, of Eq. (10). Next, the sampling matrix $\mathbf{A}$ can be constructed based on the SRM of Eq. (4). In particular, matrix $\mathbf{A}$ is formed as the tensor product of trigonometric basis functions (e.g., [28]) spanning the frequency and wavenumber domains. For the frequency domain a basis can be constructed as

$$\mathbf{B}_1 = \left[\cos(\omega_0 \mathrm{t}) \;\; \sin(\omega_1 \mathrm{t}) \;\; \cdots \;\; \sin(\omega_{\frac{\mathrm{n}}{2}-1} \mathrm{t}) \;\; \cos(\omega_{\frac{\mathrm{n}}{2}} \mathrm{t})\right] \tag{11}$$

with $\omega_\ell = 2\pi\ell/n, \;\; \ell = 0, \cdots, n/2$, whereas for the wavenumber domain the

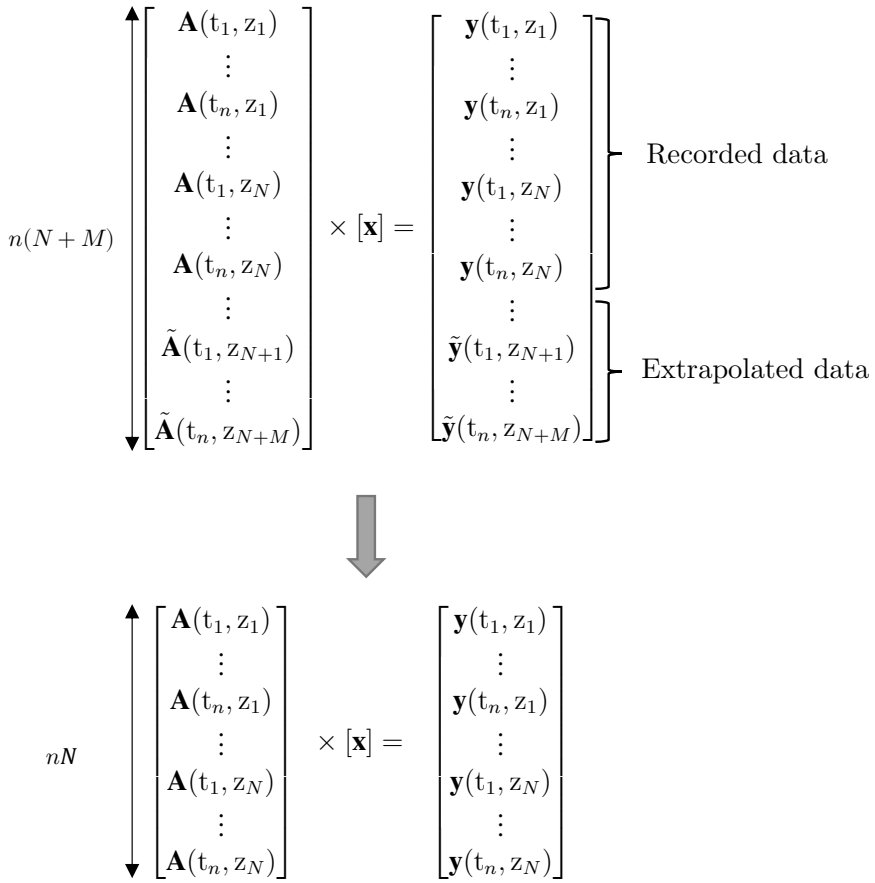


Fig. 2: Sampling matrix construction for CS-based extrapolation in a single spatial dimension.

basis becomes

$$\mathbf{B}_2 = \begin{bmatrix} \cos(k_0 z) & \sin(k_1 z) & \cdots & \sin(k_{N/2-1} z) & \cos(k_{N/2} z) \end{bmatrix} \tag{12}$$

with $k_m = 2\pi m/N, \quad m = 0, \cdots, N/2$. The tensor product of $\mathbf{B}_1$ and $\mathbf{B}_2$ produces a new basis for the joint wavenumber-frequency domain in the form

$$\mathbf{B} = \mathbf{B}_1 \otimes \mathbf{B}_2 \tag{13}$$

The complete matrix $\mathbf{B}$ is of dimensions $n(N+M) \times n(N+M)$. However, considering that time-histories at M locations of interest are not available, $\mathbf{B}$ becomes an $nN \times n(N+M)$ sampling matrix $\mathbf{A}$ to be used in Eq. (9) (or, alternatively, in Eq. (10)) for extrapolating for the M locations; see also Fig. (2). Obviously, com-

paring with the dimensions of matrix $\mathbf{A}$ defined in Eq. (9) yields $n_0 - n_m = nN$ and $n_0 = n(N+M)$. In passing, it is noted that over-complete dictionaries can be employed as well for constructing matrices $\mathbf{B}_1$ and $\mathbf{B}_2$ to enhance the respective domain resolution (e.g., [29], [30]).

3.1.2. Numerical examples

In this section, the efficacy of the proposed CS-based approach is assessed, first, in conjunction with the problem of reconstructing wind time-histories with missing data, and second, in conjunction with extrapolating to various locations in the one-dimensional spatial domain.

To this aim, 50 time-histories compatible with the PSD of Eq. (3) are generated by the SRM of Eq. (4). The parameter values used are: time duration T_0 = 255.75 s; upper cut-off frequency $w_u = 8\pi$ rad/s; $dt = 2\pi/w_u = 0.0125$ s; $\Delta\omega = 2\pi/T_0 = 0.0246$ rad/s; upper cut-off wavenumber $k_u = \pi$ rad/m; $\Delta k = 0.002$ rad/m. Next, for each time-history, 40% of missing data are introduced in uniformly random gaps. The records are then reconstructed based on the adaptive basis re-weighting approach by utilizing Eq. (10).

Two indicative wind velocity time-histories are shown in Fig. (3) corresponding to vertical heights of 50m and 62m, together with their reconstructed counterparts. It is seen that although there are notable discrepancies between the target and the reconstructed records, the main features of the time-histories are estimated satisfactorily in an average sense. This is very encouraging considering that the proposed CS-based approach in this paper focuses on estimating stochastic field statistics (e.g., PSD) defined as averages over an ensemble of realizations. Indeed, in Fig. (4) the PSD corresponding to a vertical height of 62m is shown, estimated as the ensemble average of the Fourier transform of the reconstructed time-histories with 40% missing data. Similarly, the cross-correlation and the coherence function referring to the two locations at 50m and 62m are also estimated based on the ensemble average of the reconstructed time-histories and are shown in Fig. (5). Clearly, in all cases, comparisons with the target quantities demonstrate the capability of the CS-based methodology to estimate stochastic field statistics based on an ensemble of realizations with a relatively high degree of accuracy considering the rather large amount of missing data.

Next, the configuration shown in Fig. (6) is considered where the objective is to extrapolate for the entire time-history at a height of 62m and determine related statistics by utilizing measured records corresponding to heights of 50, 56, 68 and 74m. In this regard, Fig. (7) shows an indicative target time-history at 62m generated via Eq. (4) together with its estimated counterpart based on Eq. (9).

Further, the CS-based estimated PSD at 62m is plotted in Fig. (8) and compared with the target PSD, whereas the estimated cross-correlation and the coherence function between the extrapolation point at 62m and two other indicative points are shown in Figs. (9) and (10), respectively. It is readily seen that the accuracy degree exhibited by the proposed CS-based methodology is, in general, satisfactory. As anticipated, however, and also observed in Fig. (10), the accuracy degree of the technique in estimating coherence values decreases for increasing distance between the considered points in the spatial domain.

3.2. Reconstruction and extrapolation in two spatial dimensions: Low-rank matrices and nuclear norm minimization

3.2.1. Theoretical aspects

In this section, an alternative CS approach based on nuclear norm minimization is developed to account for wind field extrapolation in two spatial dimensions. The rationale relates to the fact that a straightforward application of the CS approach proposed in section 3.1 based on l_1-norm minimization becomes computationally intensive, and even prohibitive in some cases, for an increasing number of dimensions. In particular, adapting the methodology in section 3.1 for addressing two spatial dimensions yields a sampling matrix $\mathbf{A}$ in Eq. (13) with a prohibitively large number of elements. Clearly, this renders the subsequent numerical implementation of the methodology at least a rather daunting, if not impossible, task. Thus, there is a need for developing alternative, more computationally efficient, approaches to address more sophisticated and realistic wind field modeling accounting for two spatial dimensions.

In this regard, the concept of a matrix norm is invoked in this section, and specifically, the nuclear norm of a matrix is employed (given as the sum of its singular values), which can be construed as a generalization of the l_1-norm to account for matrices (e.g., [31], [32]). In fact, minimizing the nuclear norm of low rank matrices can be viewed as an extension of minimizing the l_1-norm of sparse vectors, and thus, low rank matrices with a relatively large number of missing entries can be reconstructed with high probability [33]. Based on the above argument, the data extrapolation problem in two spatial dimensions can be recast as a "basis-free" matrix completion problem at each time instant by appropriately re-arranging the measured data in matrix form. In passing, it is worth mentioning that nuclear norm minimization in conjunction with low rank matrices has been used recently in various civil engineering applications. Indicatively, by proposing a matrix reshape scheme, a low-rank representation of large-scale structural seismic and typhoon responses was identified in [34], which proved to be beneficial

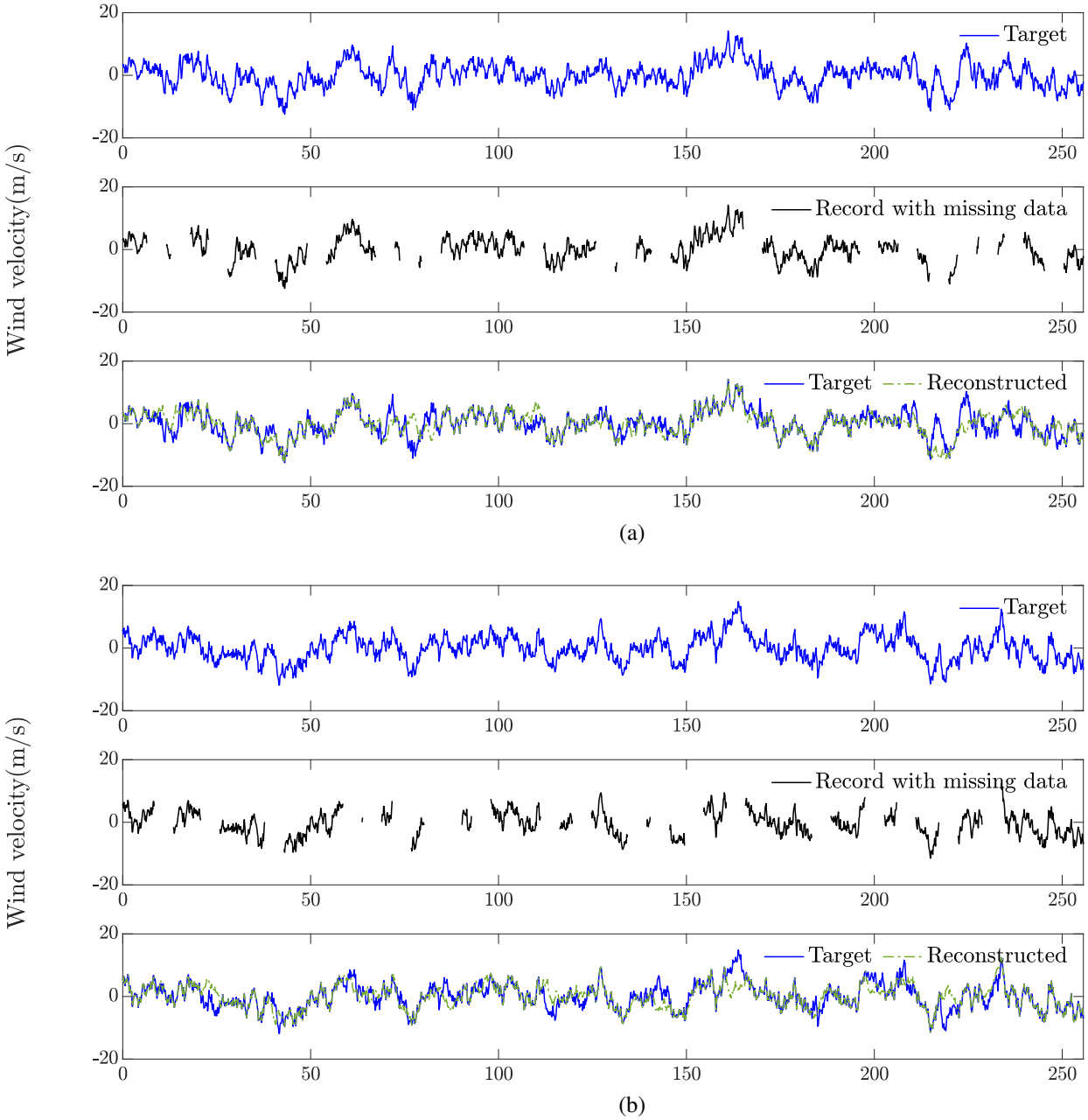


Fig. 3: Indicative wind velocity time-histories at a height of (a) 50m and (b) 62m; comparisons between the target and the CS-based reconstructed records considering 40% missing data.

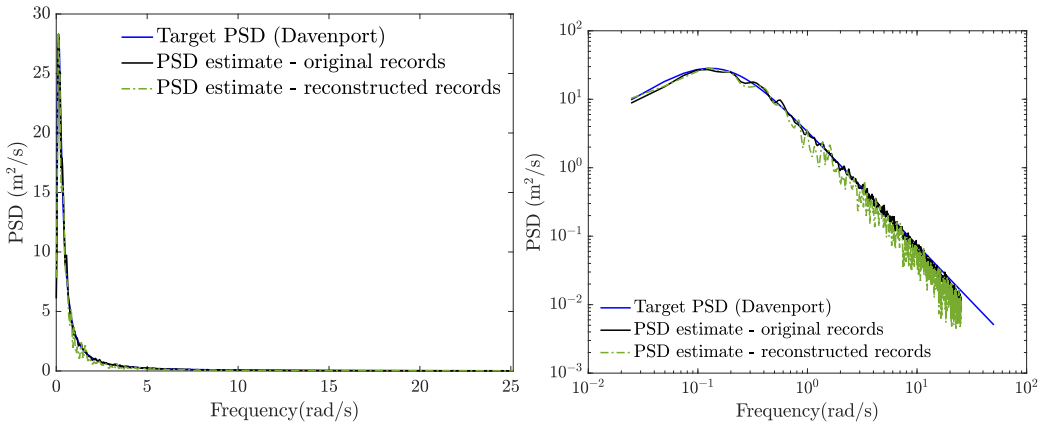


Fig. 4: Estimated PSD corresponding to a vertical height of 62m based on the ensemble average of reconstructed time-histories with 40% missing data; (a) linear scale and (b) logarithmic scale.

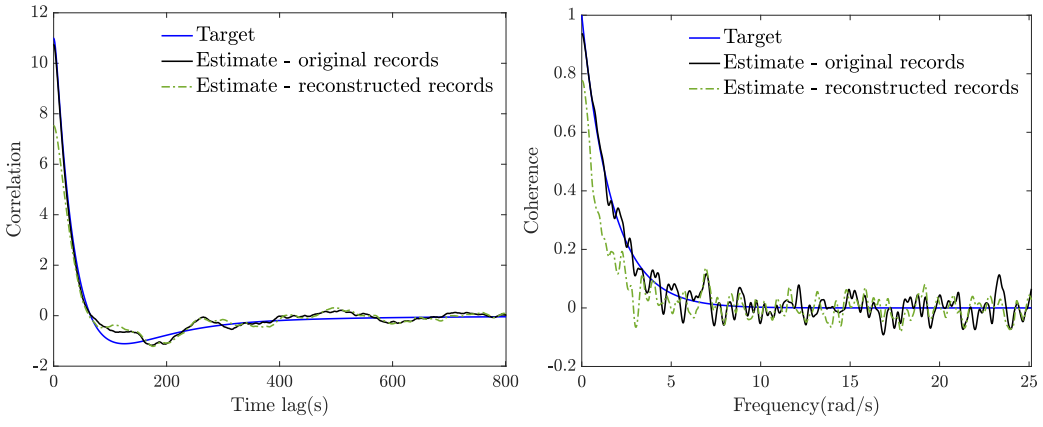


Fig. 5: (a) Cross-correlation and (b) coherence function estimated based on the ensemble average of reconstructed time-histories with 40% missing data corresponding to vertical heights of 50 and 62m.

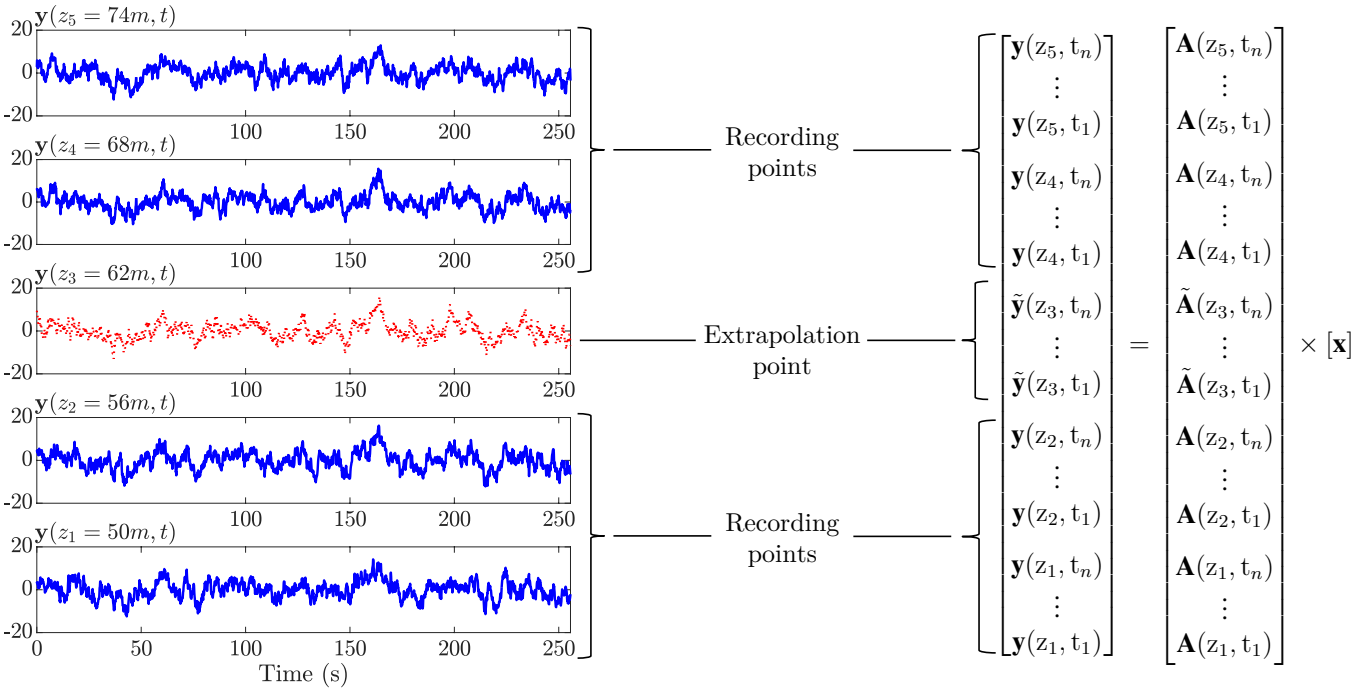


Fig. 6: Schematic representation of the extrapolation scheme in a single spatial dimension via compressive sampling.

for efficient data compression. The scheme was coupled in [35] with a nuclear norm minimization algorithm for recovering of multi-channel structural response time-histories with randomly missing data.

Next, considering a matrix $\mathbf{M} \in \mathbb{R}^{n \times n}$ of rank $r << n$, with only $l < n^2$ of its entries observed, the problem of reconstructing the complete matrix can be formally expressed as

$$\begin{aligned} &\text{minimize} \quad \|\mathbf{Y}\|_* \\ &\text{subject to} \quad Y_{k,l} = M_{k,l}, \quad (k,l) \in \Omega, \end{aligned} \tag{14}$$

where $\|\cdot\|_*$ denotes the nuclear norm and Ω is the index set of observed entries. In general, a smaller value of rank r dictates fewer required observed matrix entries for successful matrix completion. More specific relationships between n, r, and l can be found, indicatively, in [33]; see also [16] for a broader perspective. Eq. (14) represents a convex optimization problem, and a variety of algorithms have been developed for its solution; see, for instance, [36], [37] as well as [16] and references therein. In the ensuing analysis, the Augmented Lagrangian Method (ALM)

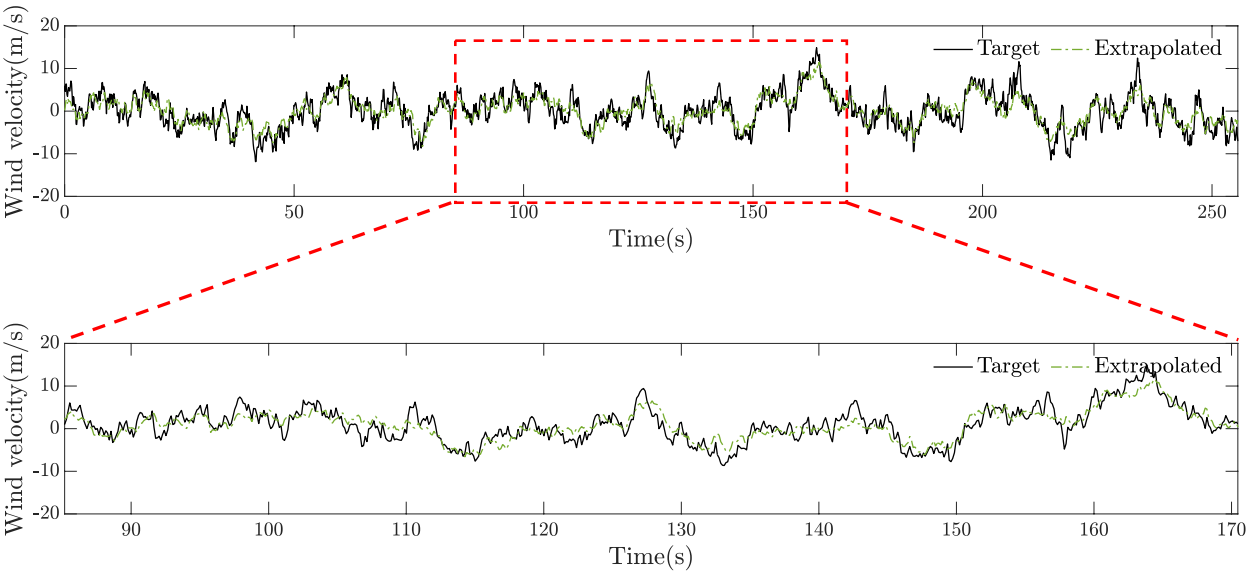
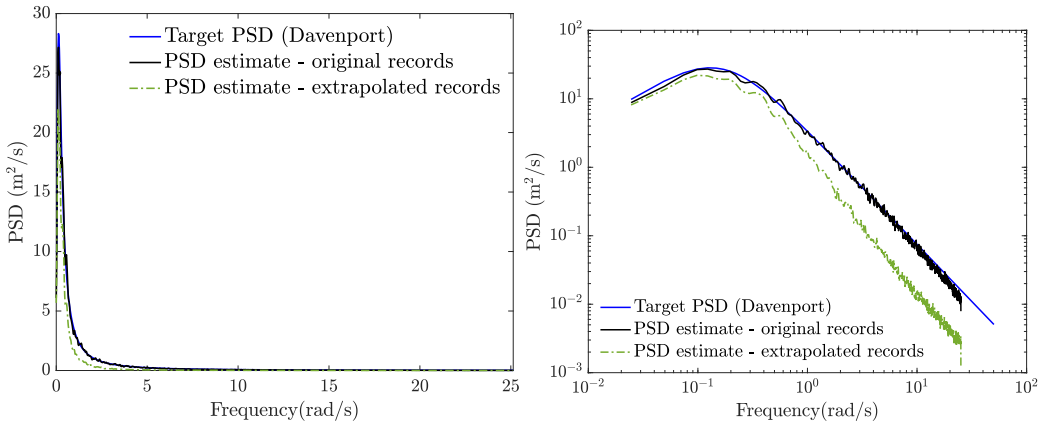


Fig. 7: Indicative wind velocity time-history at a height of 62m; comparisons between the target and the extrapolated records.



(a) (b)

Fig. 8: Estimated PSD corresponding to a vertical height of 62m based on the ensemble average of CS-based extrapolated time-histories; (a) linear scale and (b) logarithmic scale.

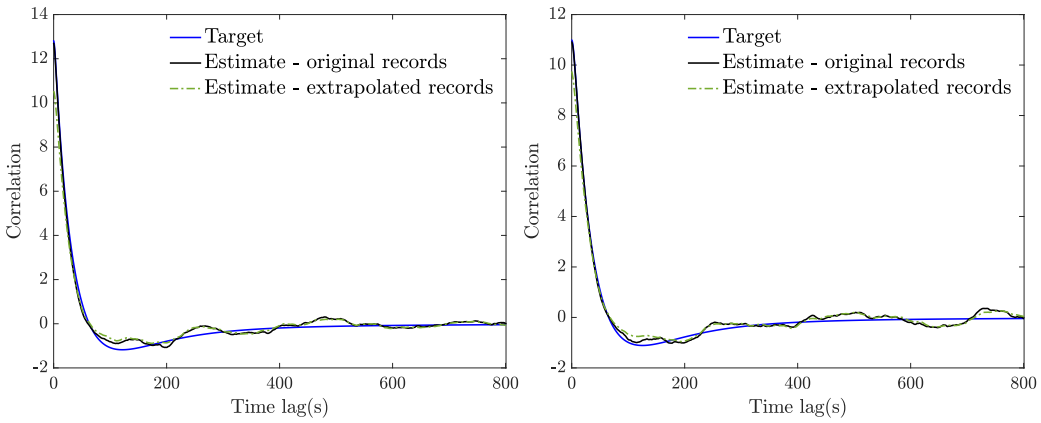


Fig. 9: Cross-correlation function estimated by employing CS-based extrapolated time-histories at 62m and original records at (a) 56m and (b) 74m.

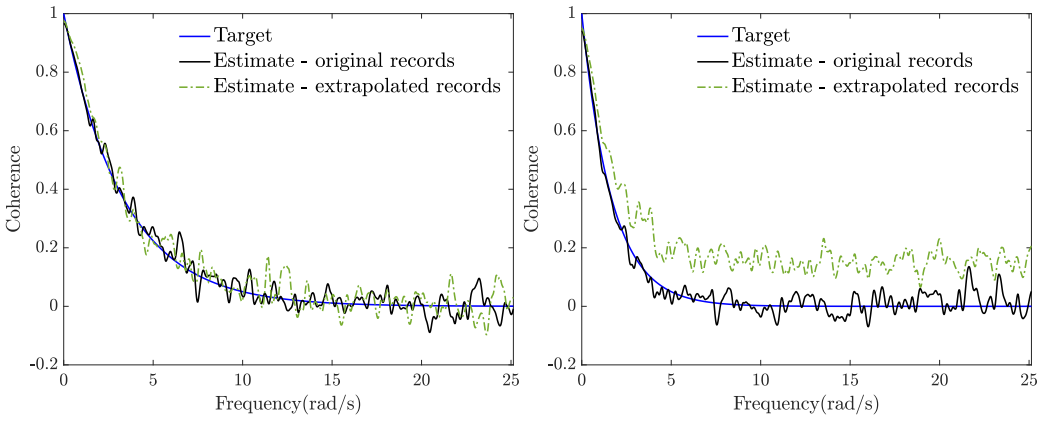


Fig. 10: Coherence function estimated by employing CS-based extrapolated time-histories at 62m and original records at (a) 56m and (b) 74m.

is employed [37] for recasting Eq. (14) into an unconstrained optimization problem and for determining the missing matrix entries; see also [38–40]. An indicative mechanization of ALM is shown in Fig. (11), whereas the interested reader is directed to [37] for more details.

3.2.2. Numerical examples

In this section, the efficacy of the proposed CS methodology based on low-rank matrices and nuclear norm minimization is assessed in conjunction with extrapolating to various locations in the two-dimensional spatial domain.

Specifically, the extrapolation configuration considered in this section is shown in Fig. (12), where 50 two-dimensional wind velocity realizations compatible with the PSD of Eq. (6) are generated based on the SRM of Eq. (8). Next, time-histories at 12 out of the 36 grid points are considered missing. It is readily seen that, at a given time instant, extrapolating for the aforementioned 12 locations in the two-dimensional domain can be formulated as a matrix completion problem in the form of Eq. (14), where $\mathbf{Y}$ represents a matrix with 36 elements. Note that the low-rank assumption, required for a successful implementation of the nuclear norm minimization solution approach, can be adequately justified by considering the dependence (by construction based on Eqs. (6)-(7)) between time-histories corresponding to different locations in the two-dimensional spatial domain. In this regard, applying the ALM shown in Fig. (11), the time-histories corresponding to the 12 grid points are determined. An indicative extrapolated record associated with location P_2 in Fig. (12) is shown in Fig. (13). Next, the PSD estimated based on the ensemble average of the extrapolated time-histories corresponding to location P_2 is shown in Fig. (14), whereas the estimated cross-correlation and the coherence function between points P_1 and P_3 are presented in Fig. (15). Clearly, the proposed methodology exhibits a high degree of accuracy in estimating related statistics. To further assess its performance, the more challenging configuration shown in Fig. (16) is considered, where only 18 out of the 36 grid points are measurement locations. Indicatively, Fig. (17) shows the estimated PSD at location P_1 as shown in Fig. (16). The estimated cross-correlation and coherence functions between points P_1 and P_2 and between points P_1 and P_3 are shown in Figs. (18) and Figs. (19), respectively. The accuracy degree exhibited remains satisfactory, although it tends to deteriorate as the distance between the locations increases.

Input: observation set Ω, sampled entries $\mathcal{P}_\Omega(\boldsymbol{M_j})$ **Output:** A_k, E_k

$Y_0 = 0; E_0 = 0;$

while not converged **do**

// solve: $\mathbf{A}_{k+1} = \arg\min_{\boldsymbol{A}} L\left(\boldsymbol{A}, \boldsymbol{E}_k, \boldsymbol{Y}_k, \mu_k\right)$

$[\boldsymbol{U}, \boldsymbol{S}, \boldsymbol{V}] = \text{svd}\left(\boldsymbol{D} - \boldsymbol{E}_k + \mu_k^{-1}\boldsymbol{Y}_k\right)$

$A_{k+1} = \boldsymbol{U}\mathcal{S}_{\mu_k^{-1}}[\boldsymbol{S}]V^T$ ▷ $\mathcal{S}$: soft thresholding (shrinkage) operator

// solve: $\boldsymbol{E}_{k+1} = \arg\min_{\pi_\Omega(\boldsymbol{E})=0} L\left(\boldsymbol{A}_{k+1}, \boldsymbol{E}, \boldsymbol{Y}_k, \mu_k\right)$

$\boldsymbol{E}_{k+1} = \pi_{\bar{\Omega}}\left(\boldsymbol{M}_j - \boldsymbol{A}_{k+1} + \mu_k^{-1}\boldsymbol{Y}_k\right)$

$\boldsymbol{Y}_{k+1} = \boldsymbol{Y}_k + \mu_k\left(\boldsymbol{M}_j - \boldsymbol{A}_{k+1} - \boldsymbol{E}_{k+1}\right); \mu_{k+1} = \rho\mu_k$

$k \quad k+1$

end while

Fig. 11: Augmented Lagrange Multipliers (ALM) Method based on [37]

.

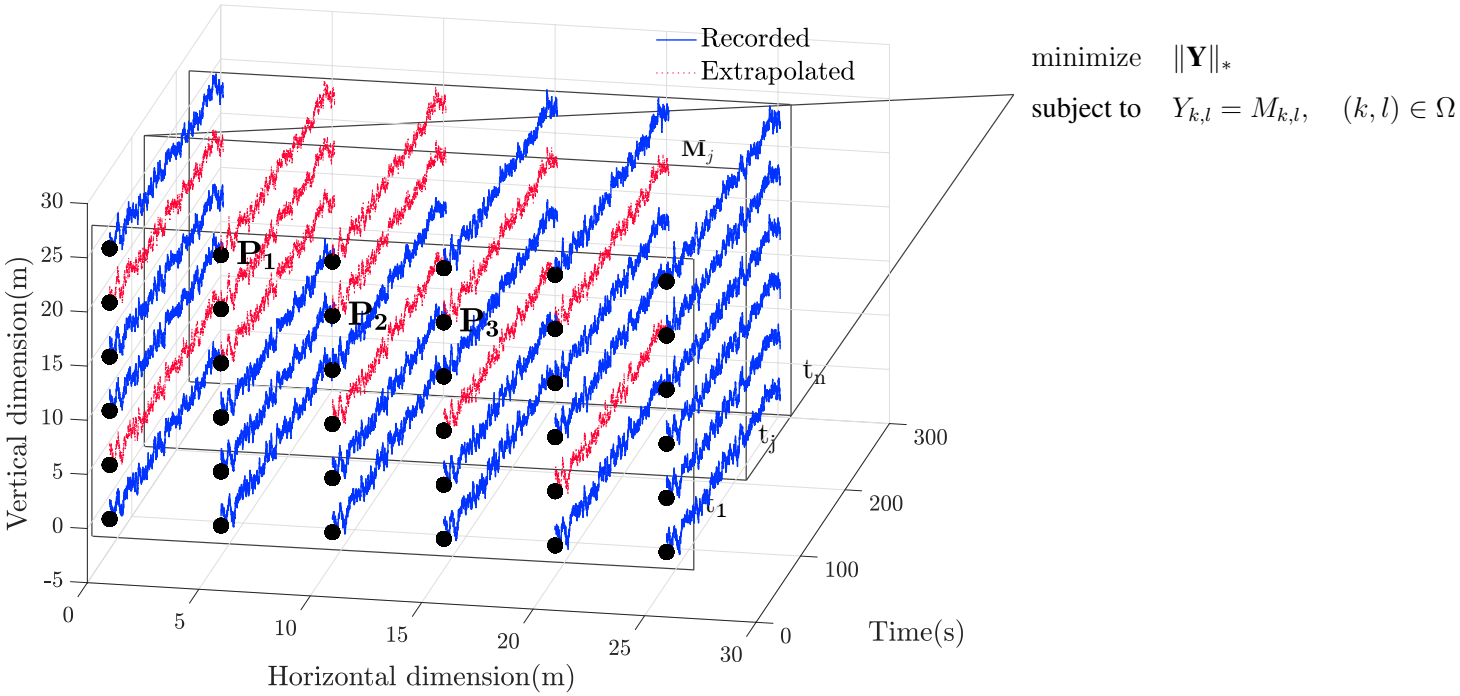


Fig. 12: Schematic representation of the extrapolation scheme in two dimensions via nuclear norm minimization; extrapolating to 12 grid points based on 24 measurement locations

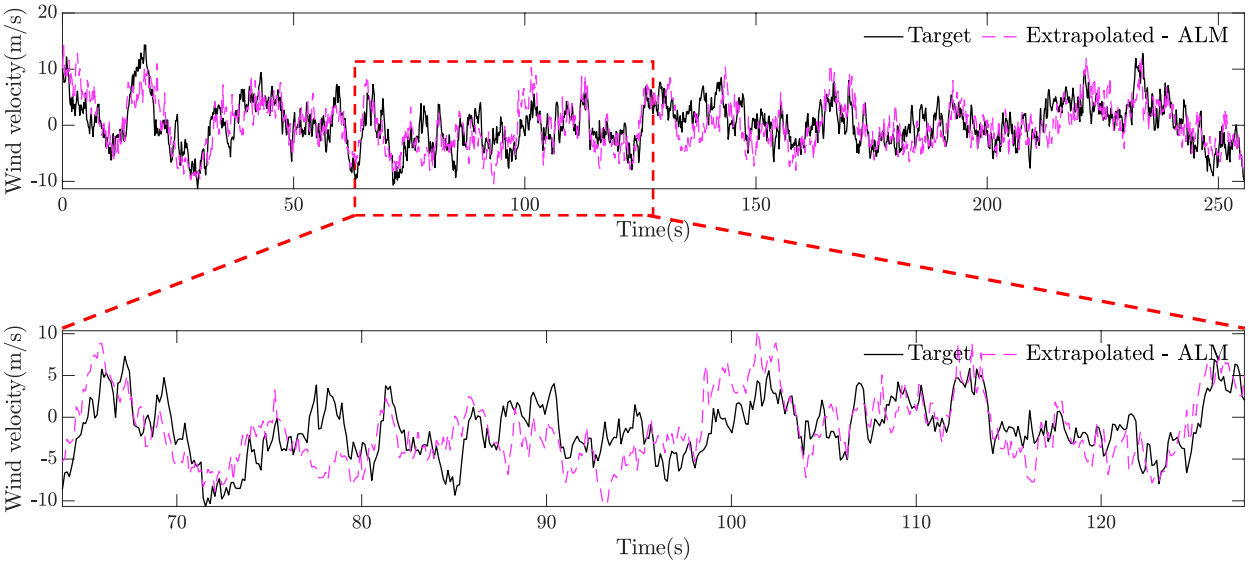


Fig. 13: Extrapolated time history at point P_2 of Fig. (12)

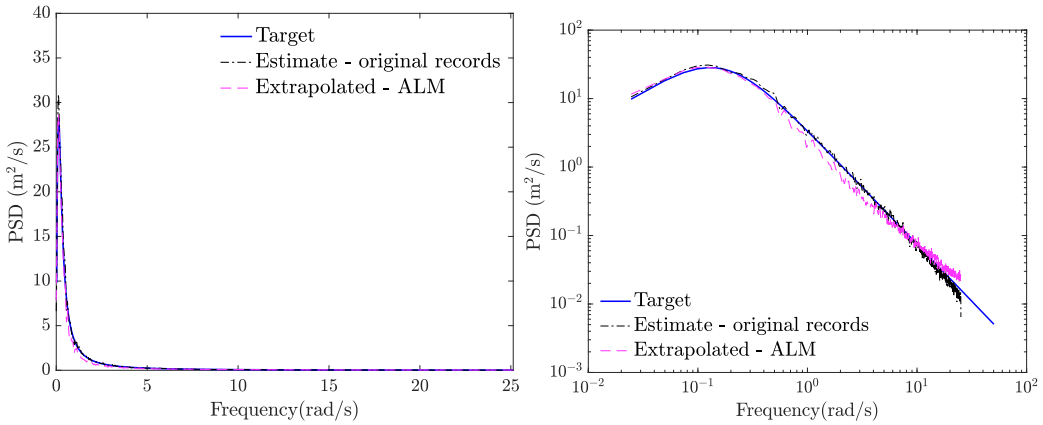


Fig. 14: Estimated PSD corresponding to point P_2 as shown in Fig. (12) based on the ensemble average of ALM-based extrapolated time-histories; (a) linear scale and (b) logarithmic scale.

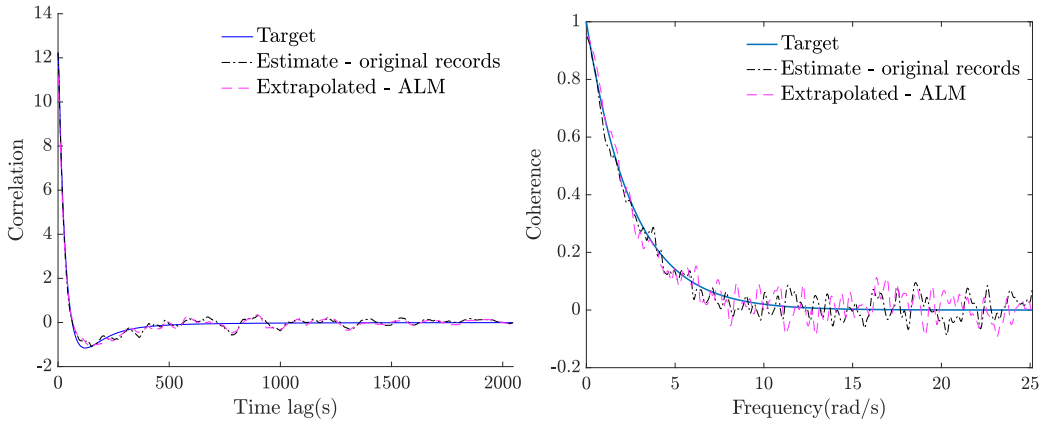


Fig. 15: (a) Cross-correlation and (b) coherence function between the ALM-based extrapolated time histories of points P_1 and P_3 as shown in Fig. (12)

4. Concluding remarks

In this paper, first, a CS approach based on l_1-norm minimization in conjunction with an adaptive basis re-weighting scheme has been developed for wind field time-histories reconstruction and extrapolation in a single spatial dimension. Second, to address computational challenges associated with higher-dimensional domains, a CS approach based on nuclear norm minimization has been developed for wind field extrapolation in two spatial dimensions. Various numerical examples have been considered for demonstrating the reliability of the proposed methodologies regarding reconstruction and extrapolation of wind field data compatible with a joint wavenumber-frequency PSD. It has been shown that the methodologies exhibit a relatively high degree of accuracy in estimating relevant statistics of the underlying stochastic field based on the ensemble of the reconstructed/extrapolated realizations, even for a large percentage of missing data. However, as anticipated, the accuracy degree in estimating coherence values decreases for increasing distance between the considered locations in the spatial domain. Finally, it is worth noting that the developed methodologies can be used potentially for environmental hazard modeling within the context of performance-based design optimization of structural systems.

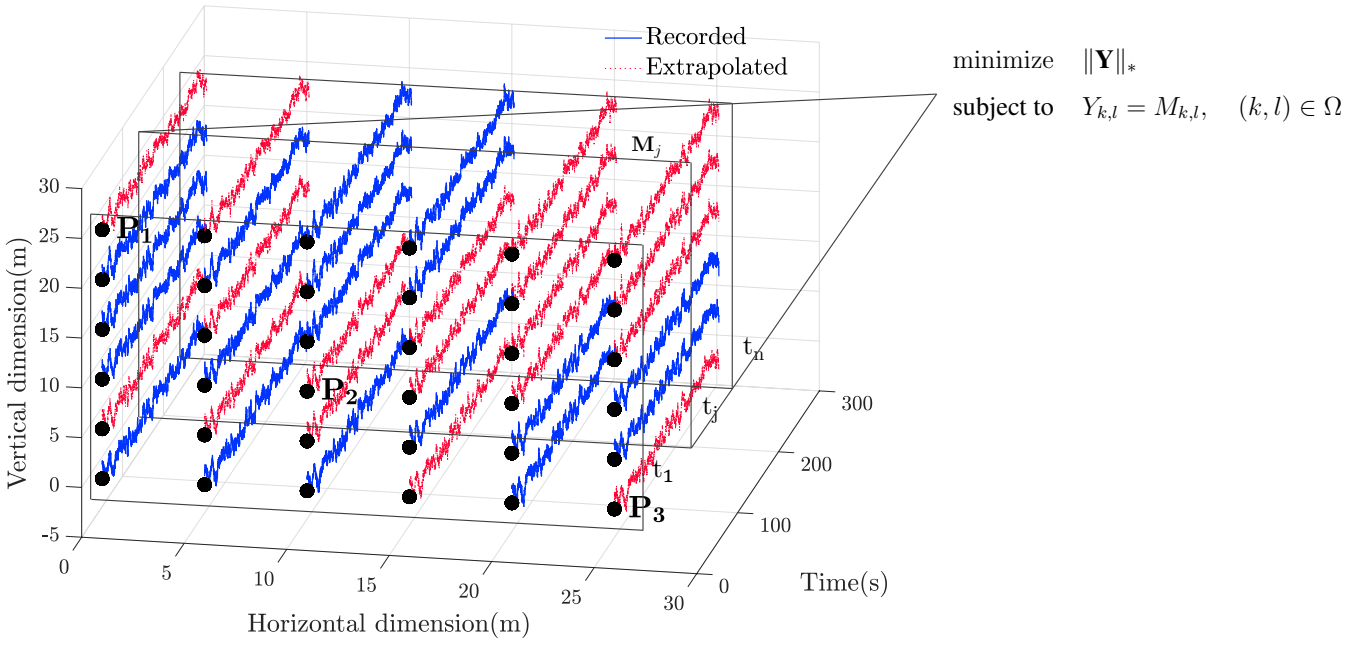


Fig. 16: Schematic representation of the extrapolation scheme in two dimensions via nuclear norm minimization; extrapolating to 18 grid points based on 18 measurement locations

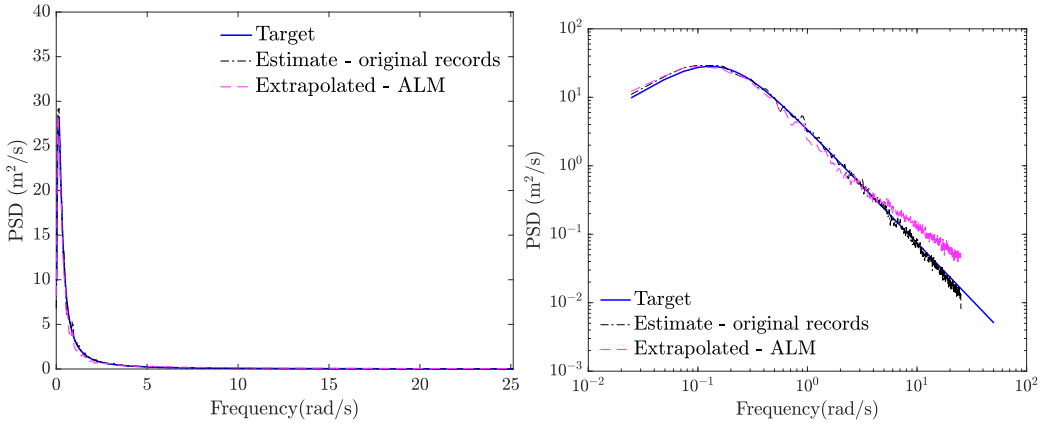


Fig. 17: Estimated PSD corresponding to point P_1 as shown in Fig. (16) based on the ensemble average of ALM-based extrapolated time-histories; (a) linear scale and (b) logarithmic scale.

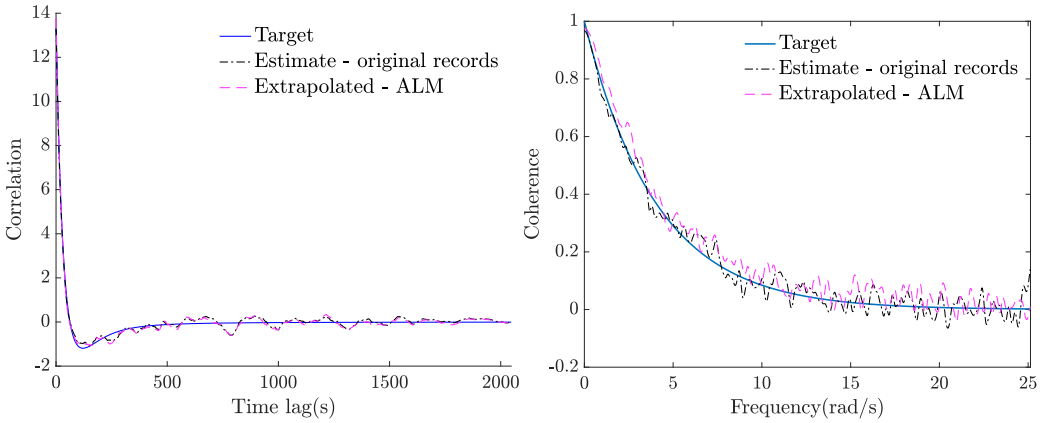


Fig. 18: (a) Cross-correlation and (b) coherence function between the ALM-based extrapolated time histories of points P_1 and P_2 as shown in Fig. (16).

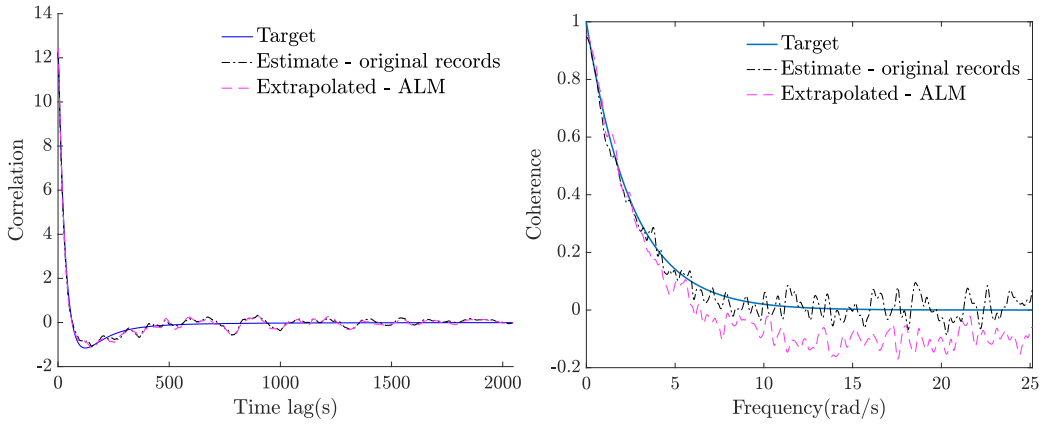


Fig. 19: (a) Cross-correlation and (b) coherence function between the ALM-based extrapolated time histories of points P_1 and P_3 as shown in Fig. (16).

CRediT authorship contribution statement

George Pasparakis: Conceptualization, Methodology, Software, Writing - original draft, Visualization, Validation. Ketson dos Santos: Conceptualization, Methodology. Ioannis A. Kougioumtzoglou: Conceptualization, Methodology, Writing - review & editing, Supervision, Project administration, Funding acquisition. Michael Beer: Supervision, Funding acquisition.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgement

G. D. Pasparakis and M. Beer gratefully acknowledge the support and funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No 764547.

I. A. Kougioumtzoglou gratefully acknowledges the support by the CMMI Division of the National Science Foundation, USA (Award number: 1724930).

References

[1] R. Frehlich, S. M. Hannon, S. W. Henderson, Coherent doppler lidar measurements of wind field statistics, Boundary-Layer Meteorology 86 (2) (1998) 233–256.

[2] M. Harris, M. Hand, A. Wright, Lidar for turbine control, National Renewable Energy Laboratory, Golden, CO, Report No. NREL/TP-500-39154 (2006).

[3] L. Carassale, G. Solari, Monte carlo simulation of wind velocity fields on complex structures, Journal of Wind Engineering and Industrial Aerodynamics 94 (5) (2006) 323–339.

[4] L. Comerford, H. Jensen, F. Mayorga, M. Beer, I. Kougioumtzoglou, Compressive sensing with an adaptive wavelet basis for structural system response and reliability analysis under missing data, Computers & Structures 182 (2017) 26–40.

[5] I. P. Mitseas, I. A. Kougioumtzoglou, M. Beer, An approximate stochastic dynamics approach for nonlinear structural system performance-based multi-objective optimum design, Structural Safety 60 (2016) 67–76.

[6] P. Dérian, C. F. Mauzey, S. D. Mayor, Wavelet-based optical flow for two-component wind field estimation from single aerosol lidar data, Journal of Atmospheric and Oceanic Technology 32 (10) (2015) 1759–1778.

[7] L. Qin, S. Liu, T. Long, M. A. Shahzad, H. I. Schlaberg, S. A. Yan, Wind field reconstruction using dimension-reduction of cfd data with experimental validation, Energy 151 (2018) 272–288.

[8] F. Qu, J. Liu, Y. Ma, D. Zang, M. Fu, A novel wind turbine data imputation method with multiple optimizations based on gans, Mechanical Systems and Signal Processing 139 (2020) 106610.

[9] Y. Ni, M. Li, Wind pressure data reconstruction using neural network techniques: A comparison between bpnn and grnn, Measurement 88 (2016) 468–476.

[10] M. A. Mohandes, S. Rehman, Wind speed extrapolation using machine learning methods and lidar measurements, IEEE Access 6 (2018) 77634–77642.

[11] Q. Lin, C. Li, Kriging based sequence interpolation and probability distribution correction for gaussian wind field data reconstruction, Journal of Wind Engineering and Industrial Aerodynamics 205 (2020) 104340.

[12] P. Towers, B. L. Jones, Real-time wind field reconstruction from lidar measurements using a dynamic wind model and state estimation, Wind Energy 19 (1) (2016) 133–150.

[13] 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.

[14] V. Laface, I. A. Kougioumtzoglou, G. Malara, F. Arena, Efficient processing of water wave records via compressive sensing and joint time-frequency analysis via harmonic wavelets, Applied Ocean Research 69 (2017) 1–9.

[15] Y. Zhang, L. Comerford, I. A. Kougioumtzoglou, M. Beer, Lp-norm minimization for stochastic process power spectrum estimation subject to incomplete data, Mechanical Systems and Signal Processing 101 (2018) 361 – 376. doi:https://doi.org/10.1016/j.ymssp.2017.08.017.

[16] I. A. Kougioumtzoglou, I. Petromichelakis, A. F. Psaros, Sparse representations and compressive sampling approaches in engineering mechanics: A review of theoretical concepts and diverse applications, Probabilistic Engineering Mechanics 61 (2020) 103082.

[17] B. A. Benowitz, G. Deodatis, Simulation of wind velocities on long span structures: A novel stochastic wave based model, Journal of wind engineering and industrial aerodynamics 147 (2015) 154–163.

[18] M. Shinozuka, G. Deodatis, Simulation of multi-dimensional gaussian stochastic fields by spectral representation (1996).

[19] J. Chen, Y. Song, Y. Peng, P. D. Spanos, Simulation of homogeneous fluctuating wind field in two spatial dimensions via a joint wave number–frequency power spectrum, Journal of Engineering Mechanics 144 (11) (2018) 04018100.

[20] G. Deodatis, M. Shinozuka, Simulation of seismic ground motion using stochastic waves, Journal of engineering mechanics 115 (12) (1989) 2723–2737.

[21] E. Simiu, R. H. Scanlan, Wind effects on structures: fundamentals and applications to design (1996).

[22] A. Zerva, Seismic ground motion simulations from a class of spatial variability models, Earthquake engineering & structural dynamics 21 (4) (1992) 351–361.

[23] E. J. Candes, J. K. Romberg, T. Tao, Stable signal recovery from incomplete and inaccurate measurements, Communications on Pure and Applied Mathematics: A Journal Issued by the Courant Institute of Mathematical Sciences 59 (8) (2006) 1207–1223.

[24] D. L. Donoho, Compressed sensing, IEEE Transactions on Information Theory 52 (4) (2006) 1289–1306. doi:10.1109/TIT.2006.871582.

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

[26] G. Malara, I. A. Kougioumtzoglou, F. Arena, Extrapolation of random wave field data via compressive sampling, Ocean Engineering 157 (2018) 87 – 95.

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

[28] A. F. Psaros, I. Petromichelakis, I. A. Kougioumtzoglou, Wiener path integrals and multi-dimensional global bases for non-stationary stochastic response determination of structural systems, Mechanical Systems and Signal Processing 128 (2019) 551 – 571. doi:https://doi.org/10.1016/j.ymssp.2019.04.014.

[29] S. S. Chen, D. L. Donoho, M. A. Saunders, Atomic decomposition by basis pursuit, SIAM review 43 (1) (2001) 129–159.

[30] E. Van Den Berg, M. P. Friedlander, Probing the pareto frontier for basis pursuit solutions, SIAM Journal on Scientific Computing 31 (2) (2009) 890–912.

[31] C. D. Meyer, Matrix analysis and applied linear algebra, Vol. 71, Siam, 2000.

[32] S. Friedland, L.-H. Lim, Nuclear norm of higher-order tensors, Mathematics of Computation 87 (311) (2018) 1255–1281.

[33] E. J. Candès, B. Recht, Exact matrix completion via convex optimization, Foundations of Computational mathematics 9 (6) (2009) 717.

[34] Y. Yang, S. Nagarajaiah, Y.-Q. Ni, Data compression of very large-scale structural seismic and typhoon responses by low-rank representation with matrix reshape, Structural Control and Health Monitoring 22 (8) (2015) 1119–1131.

[35] Y. Yang, S. Nagarajaiah, Harnessing data structure for recovery of randomly missing structural vibration responses time history: Sparse representation versus low-rank structure, Mechanical Systems and Signal Processing 74 (2016) 165–182.

[36] J.-F. Cai, E. J. Candès, Z. Shen, A singular value thresholding algorithm for matrix completion, SIAM Journal on optimization 20 (4) (2010) 1956–1982.

[37] Z. Lin, M. Chen, Y. Ma, The augmented lagrange multiplier method for exact recovery of corrupted low-rank matrices, arXiv preprint arXiv:1009.5055 (2010).

[38] M. R. Hestenes, Multiplier and gradient methods, Journal of optimization theory and applications 4 (5) (1969) 303–320.

[39] R. T. Rockafellar, The multiplier method of hestenes and powell applied to convex programming, Journal of Optimization Theory and applications 12 (6) (1973) 555–562.

[40] M. J. Powell, A fast algorithm for nonlinearly constrained optimization calculations, in: Numerical analysis, Springer, 1978, pp. 144–157.