

Bayesian Approach for main Frequency Identification on Extremely Under-sampled Signals

Chiara Nardin^{1,2[0000-0001-5860-0314]}, Stefano Zorzi^{1[0000-0002-3537-4649]}, Federica Zonzini^{3,4[0000-0002-2429-1469]}, Daniele Zonta^{1[0000-0002-7591-9519]}, Oreste Salvatore Bursi^{1[0000-0003-3072-7414]}, and Marco Broccardo^{1[0000-0003-4058-260X]}

¹ Department of Civil, Environmental, Mechanical Engineering, University of Trento,
Via Mesiano, 77, 38123 Trento, Italy

² Department of Civil, Environmental and Geomatic Engineering, Institute of
Structural Engineering, ETH, Stefano-Francini-Platz, 5, 8093 Zurich, Switzerland

³ Department of Electrical, Electronic, and Information Engineering, University of
Bologna, Viale del Risorgimento, 2, 40126 Bologna, Italy

⁴ Advanced Research Center on Electronic Systems “Ercole De Castro” (ARCES),
University of Bologna, Viale Pepoli 3/2, 40126, Bologna, Italy
chiara.nardin@unitn.it

Abstract. Structural Health Monitoring (SHM) is gaining a key role in ensuring the integrity and safety of civil infrastructures. In the last decades, the rapid shift to wireless sensor networks has given rise to new challenges, mainly related to the limitation of data transmission and payloads, energy autonomy, and computing power needed for extracting useful information. To tackle these issues in the framework of vibration assessment, diverse algorithms have been proposed, mainly inspired by the compressed sensing theory, taking advantage of their inherent sparse nature as a set of multiple exponentially damped sinusoids. However, those solutions usually entail a signal reconstruction step, which is computationally expensive, and demonstrated scarce real-field effectiveness due to significant limitations in the maximum achievable compression ratio. Instead, in this work, we bypass these constraints by proposing an alternative method, termed Bayesian Frequency Identification (BAY-FI), aimed at the identification of the main vibration frequency avoiding the decoding stage. The methodology is wrapped in a Bayesian formulation of an optimized, curve-fitting algorithm applied to random and extremely under-sampled measurement signal. BAY-FI is validated on a laboratory-scale, simply-supported beam and compared with conventional techniques, demonstrating significantly higher compression ratios while taking into consideration the importance of the sensor positioning.

Keywords: Bayesian identification method · Long-term structural health monitoring · Sparse signal measurement · Sub-Nyquist sampling.

1 Introduction

Long-term Structural Health Monitoring (SHM) is crucial to guarantee the safety and longevity of civil infrastructures [1]. Key aspects include data management, sensor resilience and powering, and extraction of damage-sensitive information [2]. Modal identification, involving the analysis of frequency-related parameters, like natural frequencies, mode shapes, and damping ratios, offers early-warning capabilities and is particularly appropriate for structures under dynamic conditions. Typically, these parameters are estimated through Operational Modal Analysis (OMA) techniques, which rely on ambient excitation tests without direct input measurement [3].

In recent decades, sensor networks for SHM have shifted from wired to wireless sensor networks to cut costs, enhance flexibility, and improve real-world use, [2]. However, challenges in data transmission, energy autonomy, and computing resources remain for long-term and real-time inspection functionalities. Data reduction techniques offer optimization at both the sensor and network levels, by minimizing transmission time and reducing bandwidth needs. Within the OMA perspective, these strategies can be divided into two main categories: i) *encoder-decoder* schemes, which reconstruct the original time series (e.g., see [4] and [5]); ii) *encoder-only* approaches, that compress data and extract the damage-sensitive information of interest from sub-sampled data, skipping the reconstruction stage (e.g., see [6] and [7]). Independently from their peculiarities, all these methods start from the common assumption that the analyzed system dynamic can be modeled as a time series resulting from the compositions of multiple exponentially damped signals (MEDS). Noteworthy, this condition applies to the majority of the industrial, mechanical, and automotive engineered structures and deserves particular attention, see for example the recent work of Huang *et al.* [7].

In the first category, Compressed Sensing (CS) is a leading methodology. Specifically, it reduces data dimensionality through a matrix-vector multiplication based on a compression matrix, assuming the signal is sparse in given representation domain. Thus, both the sensing matrix and the sparsity basis have to be chosen appropriately. For vibration data akin, exact reconstruction is possible in the spectral domain, considering that the MEDS representation can well capture the entire dynamics of the target facility. Over the past two decades, CS has been extensively researched in vibration contexts, tested in labs and real-world facilities. For example, Bao *et al.* [8,9] explored CS for vibration data compression for the real case of the Jinzhou Bridge.

Conversely, the second category aims at extracting damage-sensitive features directly in the compressed domain. Yang *et al.* [10] proposed an output-only approach that combines CS and blind source separation (BSS) for modal frequency and mode shape identification, offering an efficient alternative to traditional methods by accurately capturing these features with fewer samples than Nyquist-based techniques. Similarly, the study of [11] examines the potential of time-series compression versus modal properties estimation, using sub-Nyquist

non-uniform-in-time sampling, eventually built on a multi-co-set sub-Nyquist sampling technique.

On these premises, we build on the paradigm of extremely sub-Nyquist sampled signals to propose an innovative Bayesian algorithm for main frequency identification in dynamic systems. Specifically, leveraging the MEDS assumption, the method employs an optimized single-sinusoidal fitting algorithm guided by a signal-to-noise ratio (SNR) metric on a moving, randomly sub-sampled measurement window. The main frequency's posterior distribution is, then, determined by combining the likelihood of frequencies identified during optimization with an initial *a priori* distribution, enabling robust identification under extreme sub-sampling conditions.

The core contributions of this research are: (i) development of a Bayesian sinusoidal fitting algorithm for Frequency Identification (BAY-FI). Unlike traditional methods, BAY-FI eliminates the need for tuning and training phases, ensuring general-purpose applicability; (ii) adaptation of a global optimization algorithm (e.g., particle swarm) to determine the maximum target (a posteriori) frequency by applying the Bayes' theorem to the cumulative distribution function (CDF) of the SNR; (iii) analysis of the impact of the sensor position on the frequency identification capability in the representative case of a laboratory small-scale steel beam and comparison with a standard CS approach.

The remainder of the paper is organized as follows: Section 2 introduces the heuristic of the proposed Bayesian algorithm for frequency identification, i.e., BAY-FI. Next, in Section 3, the algorithm is tested on a simply supported beam, where its performance is compared against the classic CS-based method. Considerations on the sensibility of the method with respect to the optimal sensor placement are presented. Finally, Section 4 summarizes key findings and explores future research directions.

2 Problem Statement and Methodology

Consider the signal $s(t)$, representing the dynamic response of a structural system. This signal is zero-mean and locally dominated by a single sinusoidal component such that within a limited time window, it primarily exhibits a single frequency, referred to as the principal frequency. This behavior is typical in structural engineering, where natural frequencies are sparse in the frequency domain [7]. Additionally, during free oscillations, structures predominantly respond at their principal frequency, which corresponds to the mode associated with the largest proportion of the participating mass.

Each measurement \mathbf{x}_k of the signal s at time t_k is represented by the vector $\mathbf{x}_k = [t_k, s(t_k)]$ and is stored in the set of measurements $\Omega_{\mathbf{x}} = \{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_N\}$. The elements of $\Omega_{\mathbf{x}}$ are ordered according to time, i.e., $t_k < t_{k+1}$. Based on $\Omega_{\mathbf{x}}$, the ordered subset $\Omega_{\mathbf{x}}^{(i,n)}$ is defined. Note that the superscript (i,n) indicates the index of the subset (i) and the number of samples that the subset contains (n). Specifically, the elements of $\Omega_{\mathbf{x}}^{(i,n)}$ are extracted sequentially from $\Omega_{\mathbf{x}}$. Each subset $\Omega_{\mathbf{x}}^{(i,n)}$ represents a sliding frame of the measured signal with n sequen-

tial elements, moving along Ω_x . The number of possible sliding frames with n sequential elements is M_n . A graphical representation of the subset $\Omega_x^{(i,n)}$ is represented in Figure 1.

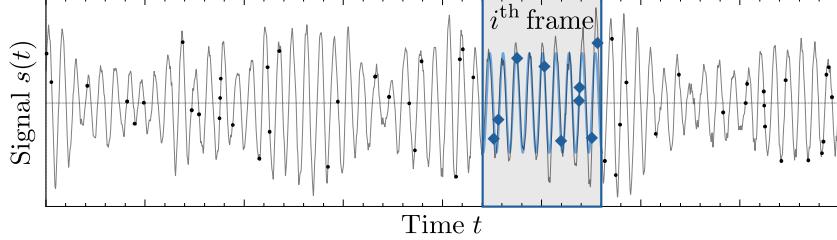


Fig. 1. Representation of the subset $\Omega_x^{(i,n)}$. The gray line represents the signal $s(t)$, whilst the black points the measurements in Ω_x . Within the i^{th} frame, the blue points represents the measurements in $\Omega_x^{(i,n)}$, while the line is the target function h , given $\Omega_x^{(i,n)}$.

Since the signal is assumed to be locally governed by a single frequency, the authors define the target function h , as

$$h(t) = A \sin(2\pi ft) + B \cos(2\pi ft), \quad (1)$$

where t is the time, A and B are, respectively, the amplitude of the sine and the cosine component, and f is the frequency of the target function h . Note that f represents the main frequency of the structural system. However, the main frequency is unknown in nature and needs to be estimated.

BAY-FI method aims to estimate the posterior probability of f in order to probabilistically identify the main frequency of the system. The target function in Equation 1 is fitted to the samples of each subset $\Omega_x^{(i,n)}$, obtaining the optimal values of $A_{\text{opt}}^{(i,n)}$ and $B_{\text{opt}}^{(i,n)}$, as

$$\left[A_{\text{opt}}^{(i,n)}, B_{\text{opt}}^{(i,n)} \right] = \underset{[A, B]}{\text{argmin}} \left\{ \text{MSE}(\Omega_x^{(i,n)} | A, B) \right\}, \quad (2)$$

where $\text{MSE}(\Omega_x^{(i,n)} | A, B) = \frac{1}{n} \sum_{k=1}^n (s(t_k) - h(t_k | A, B))^2$ is the mean squared error of the samples of $\Omega_x^{(i,n)}$. Based on the above definitions, a metric based on the signal-to-noise ratio is defined as

$$\delta_{\text{snr}}^{(i,n)} = \frac{\text{Var}[\mathbf{s}^{(i,n)}]}{\text{MSE}_{\text{opt}}^{(i,n)}}, \quad (3)$$

where $\text{Var}[\mathbf{s}^{(i,n)}]$ is the variance of the signal measurements in $\Omega_x^{(i,n)}$, while $\text{MSE}_{\text{opt}}^{(i,n)}$ is the optimum minimum value of the mean squared error, see Page

400 of [12]. The signal-to-noise ratio metrics defined in Equation 3 are calculated for all the M_n subsets $\Omega_{\mathbf{x}}^{(i,n)}$. The M_n values of $\delta_{\text{snr}}^{(i,n)}$ are collected in $\Omega_{\text{snr}}^{(n)}$ as

$$\Omega_{\text{snr}}^{(n)} = \left\{ \delta_{\text{snr}}^{(1,n)}, \delta_{\text{snr}}^{(2,n)}, \dots, \delta_{\text{snr}}^{(i,n)}, \dots, \delta_{\text{snr}}^{(M_n,n)} \right\}. \quad (4)$$

Next, based on the empirical CDF of the elements of $\Omega_{\text{snr}}^{(n)}$, $F(\delta_{\text{snr}}^{(n)})$, the observed quantity $y^{(n)}$ is defined as

$$y^{(n)} = \int_0^\infty (1 - F(\delta_{\text{snr}}^{(n)})) d\delta_{\text{snr}}^{(n)}, \quad (5)$$

Note that the observed quantity $y^{(n)}$ depends on the number of samples of each sliding frame (n) and represents the area under the empirical complementary CDF, as shown in Figure 2. In other words, the observed quantity $y^{(n)}$ is a

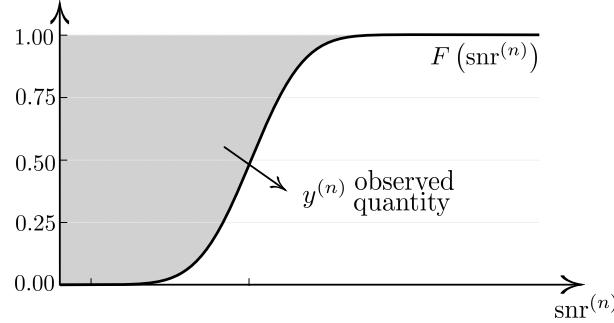


Fig. 2. Representation of the observed quantity $y^{(n)}$.

probabilistic metric that evaluates the quality of the fit, given the size n of the sliding frames. The authors expect that the frequency associated with the highest value of $y^{(n)}$ is, probably, the main frequency of the structural system. All observed quantities $y^{(n)}$ obtained from different dimensions n are collected in the observed quantity vector \mathbf{y} . The observed quantities are reasonably assumed to be statistically independent, i.e. $y^{(n_1)}$ and $y^{(n_2)}$ are i.i.d. if $n_1 \neq n_2$.

Next, the outlined framework is wrapped in a Bayesian perspective. Specifically, to value the prior knowledge that engineers generally have on the characteristics of the structure, we embed this information into a *a priori* distribution $p(f)$. This represents the engineer's belief regarding the system's main frequency.

Hence, the Bayes theorem is applied to calculate the posterior probability of f , defined as

$$p(f | \mathbf{y}) = \frac{p(\mathbf{y} | f) p(f)}{p(\mathbf{y})}, \quad (6)$$

where $p(f)$ is the prior distribution of f , $p(\mathbf{y} | f)$ is the likelihood, and $p(\mathbf{y})$ is the evidence, which assumes the role of normalization factor. Specifically, the

likelihood represents the probability of \mathbf{y} given the selected frequency f . According to the hypothesis of statistical independence of the observed quantities, the likelihood is defined as

$$p(\mathbf{y} | f) = \prod_n p(y^{(n)} | f), \quad (7)$$

where $p(y^{(n)} | f)$ is the likelihood of each $y^{(n)}$, given the selected frequency f . Notice that the authors consider the likelihood $p(y^{(n)} | f)$ to be proportional to $y^{(n)}$ itself, i.e., $p(y^{(n)} | f) \propto y^{(n)}$. Finally, since the evidence term $p(y)$ does not depend on f , the posterior probability is simply proportional to the product between the likelihood term and the prior distribution, i.e.,

$$p(f | \mathbf{y}) \propto p(\mathbf{y} | f) p(f) \propto p(f) \prod_n y^{(n)}, \quad (8)$$

Note that if computing the posterior distribution of f is computationally expensive, it is possible to consider its maximum-a-posteriori value, f_{MAP} , that can be identified using a global maximization algorithm. In the present contribution, the particle swarm algorithm [13] is considered.

3 Benchmark case: simple-supported lab beam

To test the BAY-FI method, we process acceleration data collected from a simple supported instrumented beam, tested within the Intelligent Sensors Systems (ISS) Laboratory of the University of Bologna.

3.1 Structure and sensors equipment

The structure, whose geometric and material properties are schematized in Figure 3 (base section $b = 60$ mm and thickness $h = 10$ mm), consists of a steel beam instrumented with a chain of six small footprint and lightweight accelerometers (S_1 to S_6) developed within the ISS Laboratory of the University of Bologna. Each sensing device is configured as a tight embedded system that integrates all the electronics necessary to sample, digitalize, and outsource vibration data measured through a triaxial MEMS inertial unit. A detailed description of the sensor and the network interface used to configure and download data can be found in [14]. The inter-segment distance between the sensors amounts to $d = 150$ mm, while the first sensor S_1 is positioned $d_o = 105$ mm apart from the left span of the beam. S_6 is applied in the mid-span of the beam. See Figure 3 for the layout of sensor positions and the material/mechanical properties of the structure, according to which the estimated analytical frequency amounts to 5.42 Hz.

3.2 Modal identification: BAY-FI vs CS algorithms

Free ambient vibrations are recorded using the installed MEMS sensors at a sampling frequency $f_s = 200$ Hz for an experiment duration of $d_e = 75$ seconds.

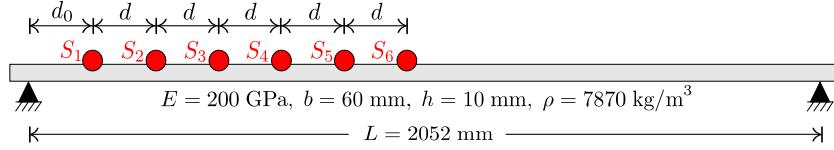


Fig. 3. Sketch of the simply-supported steel beam and the sensor installation plan, with enclosed the material properties.

Each time series contains $N_{\text{tot}} = 15000$ samples. To evaluate the performance of the BAY-FI algorithm, each time series is randomly undersampled using various *undersampling factors* α . The *undersampling factor* is defined as the ratio between the minimum number of samples required by the Nyquist-Shannon theorem N_{ny} and the samples considered N , i.e., $\alpha := N_{\text{ny}}/N$. In the analysis, we consider α values ranging from 1 to 15.

According to the heuristic described in Section 2, the definition of the prior distribution is required. Three different priors are explored, representing different levels of *a priori* beliefs: (U): a uniform distribution ranging from 0 Hz to 10 Hz, corresponding to the limits of the algorithm's search space. This case represents the absence of prior information, except for the constraints imposed by the pre-defined search space; (G_1): a Gaussian distribution with a mean of 5.42 Hz and a standard deviation of 0.50 Hz, reflecting frequencies derived from the known analytical model; (G_2): a Gaussian distribution with a mean of 2.50 Hz and a standard deviation of 1.00 Hz, designed to test the algorithm's ability to identify frequencies under incorrect prior assumptions that significantly differ from those predicted by the analytical model.

The BAY-FI method is applied on the undersampled time series considering frame lengths n from 4 to 30. The global optimization problem is solved using the function `pyswarms.single.GlobalBestPSO` included in the PySwarms library [15]. This optimizer is well-suited for problems involving real-valued parameters, such as the identification of the main frequency. The algorithm parameters are configured as follows: number of particles: 10; number of iterations: 100; cognitive parameter: 0.5; social parameter: 0.3; inertia parameter: 0.9.

To validate and compare performances between the BAY-FI and state-of-the-art CS competitor, we have assumed the following hypothesis in the encoding and decoding stage of the CS pipeline to ensure fair comparison: (i) the Discrete Cosine Transform (DCT) matrix is used as sparsifying operator, that is compatible with signals that have a few non zero components in the frequency domain; (ii) a binary sensing matrix having non-null entries only on the main diagonal has been employed to emulate the functioning of a random sampler selecting randomly the quantities to be transmitted to the decoder; (iii) the SPGL1 implementation of the recovery stage has been adopted⁵ since it allows to achieve the best recovery performances, independently from the available computational performances at the decoding stage, that are supposed unlimited.

⁵ <https://www.cs.ubc.ca/~mpf/spgl1/index.html>

3.3 Results

Figure 4 collects the results of the novel BAY-FI algorithm and the conventional CS encoding-decoding method in terms of identification performance. First, frequency identification is performed on the sensor data S_6 , i.e., the one installed in the middle of the beam that undergoes the highest vibration. The performance rate of the identification procedure, p , is calculated as a percentage of the correct frequency identification given α . The frequency identification is achieved

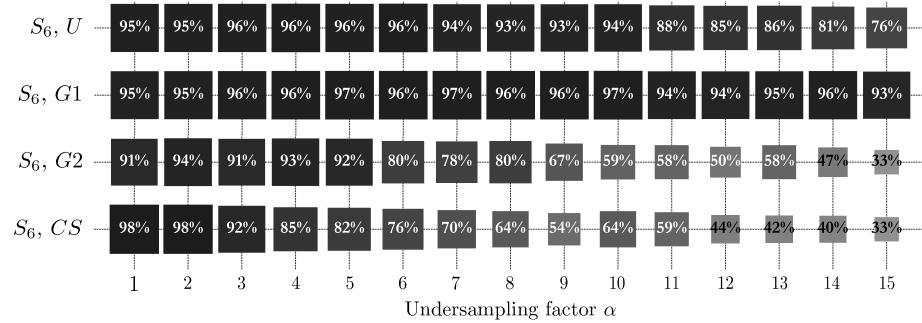


Fig. 4. Identification performance of the BAY-FI algorithm and the conventional CS method as a function of α for sensor S_6 .

successfully using both the CS method (row S_6, CS in Figure 4) and the BAY-FI algorithm under varying prior distributions (rows S_6, U , $S_6, G1$, and $S_6, G2$ in Figure 4) for $\alpha \leq 5$ –6. Beyond this threshold, a general decline in the success rate is observed for the CS method and BAY-FI with misleading Gaussian priors (rows $S_6, G2$ and S_6, CS in Figure 4). Noteworthy, BAY-FI demonstrates a consistently high success rate, even for significantly higher undersampling factors, with relevant performance sustained to $\alpha = 15$. Similarly, BAY-FI with a uniform prior maintains robust identification capabilities, although its performance gradually decreases at higher values of α .

Lastly, Figure 5 shows the impact of sensor position on frequency detection capability. Specifically, for a fair comparison with the CS algorithm, we assumed a uniform prior for the BAY-FI, i.e., to have no prior knowledge in the main frequency of the system. As the location shifts away from the position of maximum information – identified here at the middle span of the beam – correct frequency identification becomes increasingly less consistent. In particular, at positions still close to the middle-span, such as for sensor S_4 , the BAY-FI algorithm demonstrates robust performance up to $\alpha = 10$. In contrast, the CS method is capable of successful identification only for $\alpha \leq 3$. Instead, close to the support location (position S_2), the CS method generally fails to achieve reliable identification. However, BAY-FI, even using a uniform prior, achieves good accuracy for $\alpha \leq 3$. These findings underscore the superior robustness of BAY-FI, particularly in scenarios where information availability is limited.

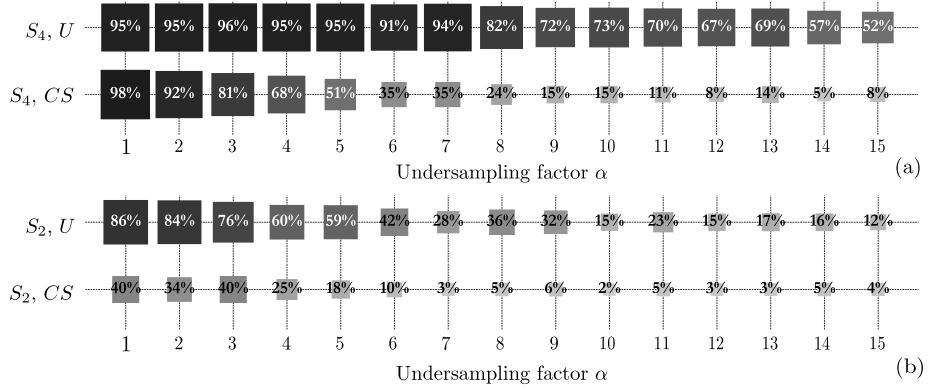


Fig. 5. Impact of the sensor position on the frequency detection performance comparing BAY-FI (uniform prior, U) and the standard CS. (a) Sensor S_4 , close to one fourth of the beam; (b) Sensor S_2 , close to the left edge of the beam.

4 Conclusions and Future Perspectives

In this study, we introduce BAY-FI, a novel Bayesian algorithm for identifying the main frequency in SHM systems. Specifically, the methodology leverages the sparsity of vibration signals in the frequency domain and enhances frequency identification even in significant samples under sampling conditions. It builds on optimized single-sinusoidal fitting method, driven by SNR-based metric, thus eliminating the need for tuning as in other state-of-the-art competitors.

A simple-supported beam is deployed as a benchmark case study to test the limit and robustness of the algorithm. Besides, validation against direct comparison with the classical CS method, considered the state-of-the-art competitor, is carried out. In addition, the influence of sensor position on frequency identification is evaluated, demonstrating the advantages of deploying BAY-FI over traditional CS approaches.

Future research should focus on evaluating the algorithm's performance across different acquisition time windows and its application to more complex case studies, thereby refining its robustness and extending its applicability to a broader range of dynamic systems.

Acknowledgement

This research is supported by MUR for the project DICAM-EXC (grant L232/2016), Autostrada del Brennero Spa (project DIGITAL TWINS), the EU project HORUS (CUP E53D23003560006) and the PRIN project PREVENT (CUP E53D23004440006, for the first author until 31/12/2024). Besides, the first author would like to acknowledge the Marie-Sklodowska Curie program and the REACTIS project, GA no. 101147351. Views and opinions expressed are however those of the authors only and do not necessarily reflect those of the European Union, or REA or any sponsor. Neither the European Union nor the granting authority can be held responsible for them.

References

1. C. R. Farrar and K. Worden, "An introduction to structural health monitoring," *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, vol. 365, no. 1851, pp. 303–315, Feb. 2007.
2. X. Yu, Y. Fu, J. Li, J. Mao, T. Hoang, and H. Wang, "Recent advances in wireless sensor networks for structural health monitoring of civil infrastructure," *Journal of Infrastructure Intelligence and Resilience*, vol. 3, no. 1, p. 100066, 2024.
3. F. Bin Zahid, O. Chao, and S. Khoo, "A review of operational modal analysis techniques for in-service modal identification," *Journal of the Brazilian Society of Mechanical Sciences and Engineering*, vol. 42, 08 2020.
4. Y. Bao, Z. Tang, and H. Li, "Compressive-sensing data reconstruction for structural health monitoring: a machine-learning approach," *Structural Health Monitoring*, vol. 19, 2019.
5. H.-P. Wan, G.-S. Dong, Y. Luo, and Y.-Q. Ni, "An improved complex multi-task bayesian compressive sensing approach for compression and reconstruction of shm data," *Mechanical Systems and Signal Processing*, vol. 167, p. 108531, 2022.
6. K. Gkoktsi, A. Giaralis, R. Klis, V. Dertimanis, and E. Chatzi, "Output-only vibration-based monitoring of civil infrastructure via sub-nyquist/compressive measurements supporting reduced wireless data transmission," *Frontiers in Built Environment*, 09 2019.
7. G. Huang, S. Deng, A. Ni, Y. Zhang, W. Lu, and J. Wang, "Distributed sub-nyquist sampling and parameters measurement of common frequency support meds signals," *Digital Signal Processing*, vol. 142, p. 104223, 2023.
8. Y. Bao, J. Beck, and H. Li, "Compressive sampling for accelerometer signals in structural health monitoring," *Structural Health Monitoring*, vol. 10, p. 235–246, 2011.
9. Y. Bao, Y. Yu, H. Li, X. Mao, W. Jiao, Z. Zou, and J. Ou, "Compressive sensing-based lost data recovery of fast-moving wireless sensing for structural health monitoring," *Structural Control and Health Monitoring*, vol. 22, 07 2014.
10. Y. Yang and S. Nagarajaiah, "Output-only modal identification by compressed sensing: Non-uniform low-rate random sampling," *Mechanical Systems and Signal Processing*, vol. 56, pp. 15–34, 2015.
11. K. Gkoktsi and A. Giaralis, "A multi-sensor sub-nyquist power spectrum blind sampling approach for low-power wireless sensors in operational modal analysis applications," *Mechanical Systems and Signal Processing*, vol. 116, pp. 879–899, 2019.
12. M. Vetterli and J. Kovačević, *Wavelets and subband coding*, ser. Prentice Hall signal processing series. Upper Saddle River, NJ: Prentice Hall, 1995.
13. J. Kennedy and R. Eberhart, "Particle swarm optimization," in *Proceedings of ICNN'95 - International Conference on Neural Networks*, vol. 4, 1995, pp. 1942–1948 vol.4.
14. N. Testoni, C. Aguzzi, V. Arditì, F. Zonzini, L. De Marchi, A. Marzani, and T. S. Cinotti, "A sensor network with embedded data processing and data-to-cloud capabilities for vibration-based real-time shm," *Journal of Sensors*, vol. 2018, no. 1, p. 2107679, 2018.
15. L. J. Miranda, "Pyswarms: a research toolkit for particle swarm optimization in python," *Journal of Open Source Software*, vol. 3, no. 21, p. 433, 2018.