

Article

An Autoencoder-Based Deep Learning Approach for Load Identification in Structural Dynamics

Luca Rosafalco ^{1,*} , Andrea Manzoni ² , Stefano Mariani ¹  and Alberto Corigliano ¹ 

¹ Dipartimento di Ingegneria Civile e Ambientale, Politecnico di Milano, Piazza L. da Vinci 32, 20133 Milano, Italy; stefano.mariani@polimi.it (S.M.); alberto.corigliano@polimi.it (A.C.)

² MOX, Dipartimento di Matematica, Politecnico di Milano, Piazza L. da Vinci 32, 20133 Milano, Italy; andrea1.manzoni@polimi.it

* Correspondence: luca.rosafalco@polimi.it; Tel.: +39-02-2399-4273

Abstract: In civil engineering, different machine learning algorithms have been adopted to process the huge amount of data continuously acquired through sensor networks and solve inverse problems. Challenging issues linked to structural health monitoring or load identification are currently related to big data, consisting of structural vibration recordings shaped as a multivariate time series. Any algorithm should therefore allow an effective dimensionality reduction, retaining the informative content of data and inferring correlations within and across the time series. Within this framework, we propose a time series AutoEncoder (AE) employing inception modules and residual learning for the encoding and the decoding parts, and an extremely reduced latent representation specifically tailored to tackle load identification tasks. We discuss the choice of the dimensionality of this latent representation, considering the sources of variability in the recordings and the inverse-forward nature of the AE. To help setting the aforementioned dimensionality, the false nearest neighbor heuristics is also exploited. The reported numerical results, related to shear buildings excited by dynamic loadings, highlight the signal reconstruction capacity of the proposed AE, and the capability to accomplish the load identification task.

Keywords: load/system identification; deep learning; structural dynamics; autoencoder; false nearest neighbor



Citation: Rosafalco, L.; Manzoni, A.; Mariani, S.; Corigliano, A. An Autoencoder-Based Deep Learning Approach for Load Identification in Structural Dynamics. *Sensors* **2021**, *21*, 4207. <https://doi.org/10.3390/s21124207>

Academic Editor: Hoon Sohn

Received: 20 May 2021
Accepted: 16 June 2021
Published: 19 June 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Modern sensor technology makes it possible to continuously monitor structures and infrastructures and, thank to the use of deep learning (DL) techniques, also an effective health management and structural prognostics [1]. The analysis of the vibrational response of buildings, i.e., of displacement and/or acceleration recordings shaped as time series (TS), is a popular approach to structural health monitoring (SHM) aiming to assess the damage state of the structures [2]. The said vibrational response depends both on the structural properties (mass and stiffness distribution, sources of damping, possible damage pattern), and on the environmental and operational conditions [3]. The identification of the loading conditions can improve the effectiveness of a SHM system allowing, e.g., the setting of physics-based models matching the structural frequencies to be monitored and not the harmonic components of the loading [4,5]. The importance of the topic has led, in the past, to a series of approaches addressing modal identification for output-only systems [6,7]. Due to the ill-posedness of the load identification problem, regularization methods based on singular value decomposition were proposed in [8]: they require a minimum tuning work to be applied, possibly within a multi-objective optimization framework [9]. Deterministic regularization methods have been recently adopted to quantify uncertainty as interval numbers [10], keeping the computational burden limited with respect to probabilistic approaches.

Similarly to what done by singular value decomposition-based methods, we exploit a reduced representation of vibrational data to deal with the ill-posedness of the load identification problem [11]. An AutoEncoder (AE) has been designed to perform the dimensionality reduction; such an AE represents the feature extraction task of the SHM approach [12]. An AE is a neural network (NN)-based architecture capable of performing nonlinear dimensionality reduction of its input [13]. Due to the difficulty in reconstructing the TS frequency content and peaks, AEs are normally employed after a pre-processing stage. In the SHM field, an AE was employed in [14] to reduce the dimensionality of measurements of vibrations induced by the vehicle-bridge interaction, preliminary transformed into the frequency domain, and exploited to quantify and localise damage. In [15], a sparse AE was used as regression model to identify the damage location in a seven-storey steel frame structure, performing dimensionality reduction through modal identification. Other paradigms were exploited for fault detection via signal reconstruction, e.g., through auto-associative extreme learning machines [16].

The advantages of employing DL to construct a regression model, after a preliminary data dimensionality reduction, were explored in [17,18]. In [19], contractive AEs were used to extract robust features and perform fault detection for a rotating machinery. Here, we also aim at exploiting the knowledge of the dynamic behaviour of the structure and its interaction with the loading conditions to highlight the connection between the reconstruction capacity of the AE and the properties of the structure under study.

Compared to classic regularization techniques, an AE and, more generally speaking, NNs require a training stage wherein the network is trained by means of a suitable dataset (training dataset). Such training of the AE allows us to handle large uncertainty levels, identifying hidden dependencies between the handled signals and the loading conditions. Dimensionality reduction is therefore intended to perform a data mining task [20]. Instead of applying a pre-processing stage before data dimensionality reduction, we directly operate on multivariate time series (MTS) to achieve at least two advantages: avoiding any feature engineering; obtaining a data reduced representation exploitable for different tasks.

A further aspect addressed in this paper is related to the optimal setting of the dimension of the latent representation; both the reconstruction capacity of the AE and the load identification task are indeed affected by the number of latent variables. As this number looks hard to establish a-priori, due to the highly nonlinear response of deep NNs, the false nearest neighbour (FNN) technique has been exploited. We show how this critical feature can be set a-posteriori, by looking at the activation of the latent variables.

Compared to the existing literature, the contributions of our work are manifold, and deal with: (i) the application of a DL framework for data dimensionality reduction and regression in view of load identification; (ii) the analysis of the impact of structural dynamics on the statistical content of the signals, thereby helping to understand the outcome of the AE training; (iii) the establishment of a criterion to set a suitably reduced number of latent variables.

The remainder of the paper is organized as follows. In Section 2, the AE architecture is described, with an eye on its use in the load identification task for structural dynamics. In Section 3, the a-posteriori choice of the dimensionality of the reduced data representation is addressed, showing how the FNN heuristics can be exploited (see also [21]). In Section 4, extensive numerical results are presented, focusing on shear buildings: first, a deep investigation is reported for a two-storey case; next, the 39-storey Pirelli Tower in Milan is considered to quantitatively assess the performance and the computational burden of the proposed methodology. Finally, Section 5 gathers some concluding remarks, and suggests future developments to ensure the applicability of the approach to other structural systems.

2. Autoencoders for Input (Load) Identification

2.1. Autoencoder Paradigm

In a NN architecture, weights and biases, collected in Ω , are set during the training stage by minimizing a loss function $c(\bar{\mathbf{U}}, \mathbf{U})$, where: $\bar{\mathbf{U}} \in \mathbb{R}^{L \times N^{out}}$ is the desired output, which is known within a supervised learning framework; $\mathbf{U} = \mathbf{U}(\mathbf{V}, \Omega) \in \mathbb{R}^{L \times N^{out}}$ is the NN output; \mathbf{V} is the NN input. The definition of $c(\mathbf{U}, \bar{\mathbf{U}})$ depends on the task to be accomplished by the NN. In this work, as we aim at load identification, the loss function is given by the mean squared error (MSE):

$$c(\bar{\mathbf{U}}, \mathbf{U}) = \sum_{l=1}^L \sum_{n=1}^N (\bar{u}_{ln} - u_{ln})^2, \quad (1)$$

where u_{ln} is the (l, n) -th entry of \mathbf{U} , and \bar{u}_{ln} is the corresponding entry of $\bar{\mathbf{U}}$. Loss minimization during training usually relies on the back-propagation of the error $c(\bar{\mathbf{U}}, \mathbf{U})$ (see, e.g., [22]); to achieve it, Adam [23], namely a first-order stochastic gradient descent algorithm, has been adopted.

Tailored to the mentioned load identification task, AEs are exploited to perform dimensionality reduction of the structural vibrations recorded by a sensor network. Within a time window $(0, T)$, each sensor provides a univariate TS, while the entire collection of the recordings by the whole sensor network provides an MTS, that is a collection of synchronized TS. From now on, we denote by $\mathbf{V} = [\mathbf{v}_1, \dots, \mathbf{v}_N] \in \mathbb{R}^{L \times N}$ the generic output of a monitoring system employing N sensors, each of them consisting of L samples within the considered time window.

Convolutional layers allow to detect both local correlations within a TS, and correlations among different TS. This latter aspect is very important for the SHM of civil infrastructures, as it provides a means to implicitly recognize the shape of the vibration modes [24,25]. By stacking convolutional layers, more complex correlation problems in time can be handled, allowing to detect the aforementioned modes of the structure, still keeping the number of the weights in Ω rather limited.

The designed AE stacks a first sequence of convolutional layers, which represent the encoder function $enc : \mathbf{V} \rightarrow \mathbf{z}$, where $\mathbf{z} \in \mathbb{R}^P$ and $P \ll (L \times N)$. The vector \mathbf{z} , which is the latent representation of \mathbf{V} , does not usually have a clear physical meaning, and requires a supervised procedure to interpret its content. This procedure, generally consisting of the recovery of the generative factors of the input dataset, can be accomplished only if few of the generative factors are known in advance [26]; in the following Section 3, the generative factors will be discussed in details.

The architecture of the adopted enc is depicted in Figure 1. It features two main branches: the first one stacking a sequence of three inception modules, operating dimensionality reduction [27]; the second one stacking nine one-dimensional convolutional layers, featuring a scaled exponential linear unit (SELU) activation function [28]. The two branches are preceded by a one-dimensional convolutional layer, and are followed by a concatenation layer, a global average pooling (GAP) layer [27] and, finally, two fully connected (FC) layers. Regarding these two final FC layers, the former employs a SELU activation function while the latter does not exploit any activation. All the convolutional layers employ a stride $S = 1$. The inception modules have been proved beneficial to improve the reconstruction capacity of the AE, at variance with the residual learning paradigm (ResNet) [29] or the squeeze-and-excitation modules (SENet) [30]. The described architecture has been obtained after having explored several designs suited to pattern recognition.

We highlight that the reason behind the use of inception modules is the attempt of approximating the sparse deep NN structure for the optimal representation (intended as the capacity of inferring correlations) of the input dataset, see [31] for a theoretical discussion of this topic. Despite the lack of a rigorous mathematical proof connecting the peculiar design of the inception modules with the aforementioned theoretical evidences, inception modules have been proven successful in many applications. The same idea of

approximating a sparse NN architecture by exploiting dense operators available in libraries like Tensorflow [32], here used with the Keras [33] API, has inspired the design of the employed two-branch architecture.

Through the GAP, synthetic information collected in a one-dimensional array is determined by computing an average value for each input channel. The last two FC layers are finally used to give extra flexibility to the encoder, to combine the synthetic descriptions of the channel contents extracted by the GAP and, as a result, to obtain \mathbf{z} .

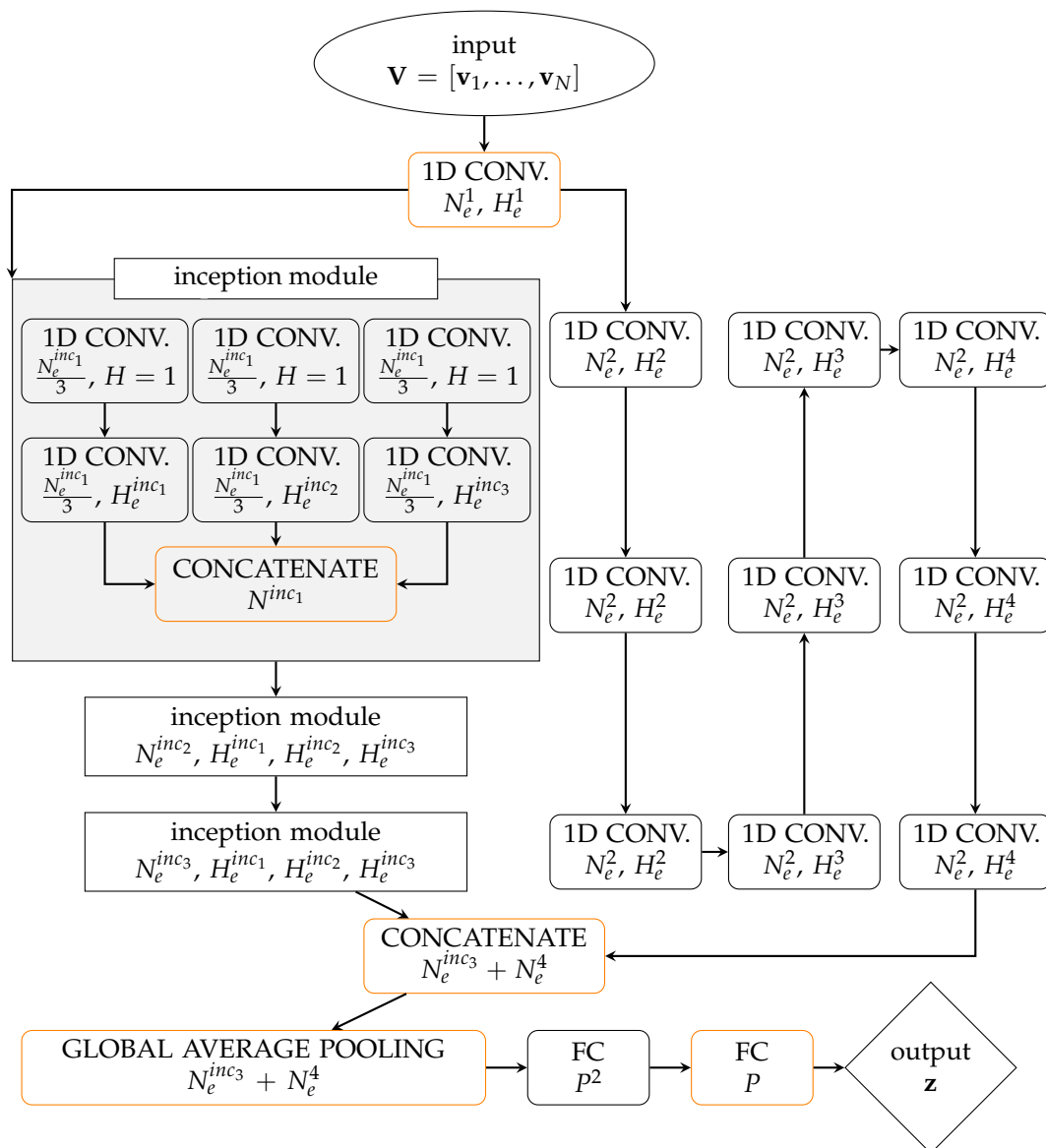


Figure 1. Architecture of the encoder. Layers are schematically represented by boxes: rectangular boxes with rounded corners depict single layer operations; rectangular boxes with sharp corners depict inception modules with dimensionality reduction; the elliptic box depicts the encoder input, while the diamond box depicts the encoder output. Orange edges are used for the layers that do not apply any activation function; for the other layers, the ReLU is used as activation function in the inception module, while the SELU is employed elsewhere. Each inception module assembles three one-dimensional convolutional (1D CONV.) layers and one concatenation layer, as shown in the grey box. In each box, the text in the first line specifies the layer type; the text in the second and (possibly) the third lines specifies instead the number of channels of the layer output and, for the convolutional layers, the kernel dimension.

A second sequence of layers provides the subsequent decoder part, whose corresponding function is denoted by $dec : \mathbf{z} \rightarrow \bar{\mathbf{U}}$. Convolutional layers are also exploited for the

decoder. The designed NN architecture of *dec* is depicted in Figure 2: first, a FC layer, employing no activation functions, is used to expand \mathbf{z} ; second, the one-dimensional array obtained with the FC layer, is reshaped as a two-dimensional array fashioned as an MTS; next, three convolutional sets are used, each of which stacking seven dilated convolutional layers; finally, three FC layers, the first two using a SELU activation function and the third without any activation, are employed to provide \mathbf{U} . A dilated convolutional layer [34], with H setting the kernel of the convolution, operates by spacing its input with a dilation rate J . By setting $J = 1$, a dilated convolution yields a standard convolution. The seven dilated convolutional layers feature $H = 2$ and J doubled every layer up to $J = 128$.

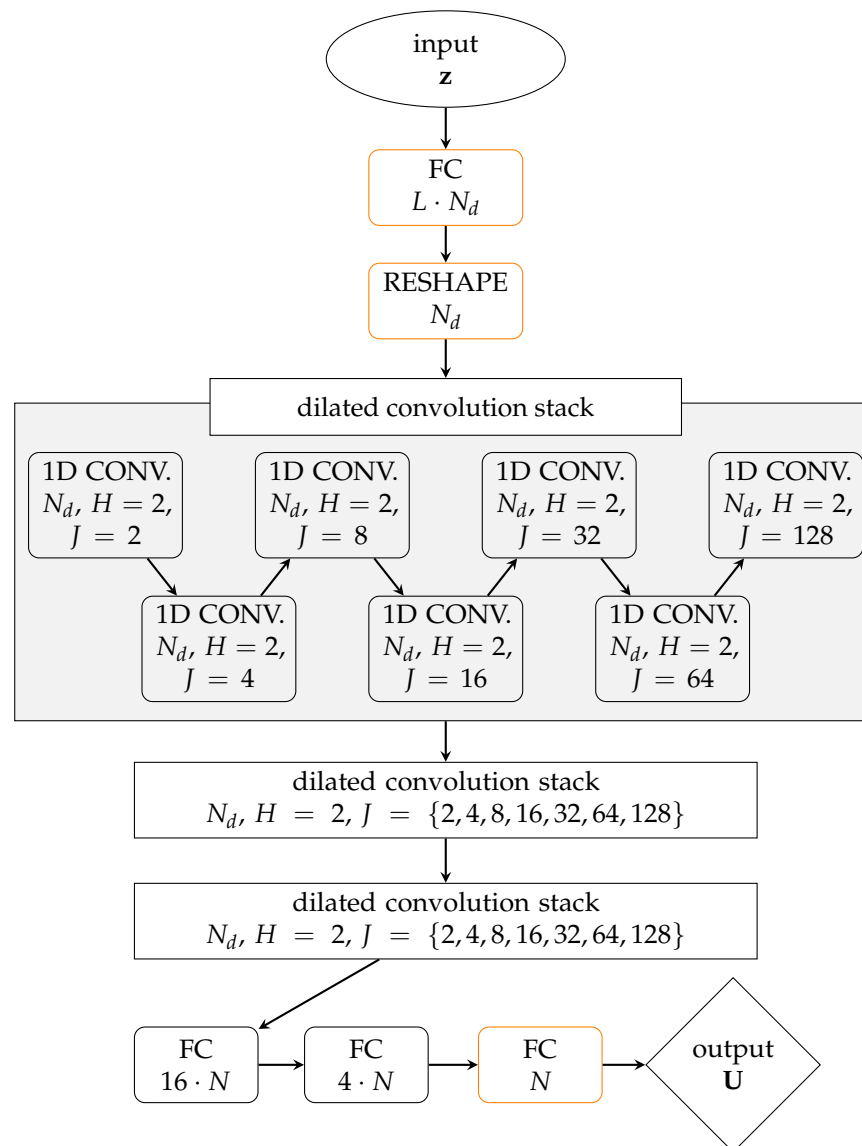


Figure 2. Architecture of the decoder. Layers are schematically represented by boxes, with the notation as detailed in the caption of Figure 1. Orange edges are used for the layers that do not apply any activation function, while the SELU is employed elsewhere. Each stack of dilated convolutions is formed by seven dilated convolutional layer, as shown in the grey box.

The use of three dilated convolution stacks have been inspired by [35]: there, it was argued that a stack of dilated convolutions, with J doubled every layer, is equivalent to apply a unique convolutional layer (with $H = 256$), but in a by far more efficient way and involving much less parameters to tune. The intuition behind this architecture is that convolutions featuring smaller values of J are responsible for reconstructing short-

term patterns, while those featuring larger J values allow the reconstruction of long-term patterns. Compared to [35], we have not employed causal connections, originally proposed to avoid the violation of the ordering of the layer input data, given the stationary framework of the present study. The last FC layers are used for combining the (possibly redundant) information contained in the output of the convolution stacks, reducing its dimensionality along the channel dimension.

If $\bar{\mathbf{U}}$ has the same dimensions of \mathbf{V} , the AE can accomplish the task of reconstructing the input \mathbf{V} by setting $\bar{\mathbf{U}} = \mathbf{V}$ in Equation (1): the resulting working principle of the AE is sketched in Figure 3. The sought reduced representation \mathbf{z} of the input \mathbf{V} is thus obtained by composing the encoder and the decoder as $dec \circ enc : \mathbf{V} \rightarrow \mathbf{U}$, and setting $\mathbf{U} \approx \mathbf{V}$.

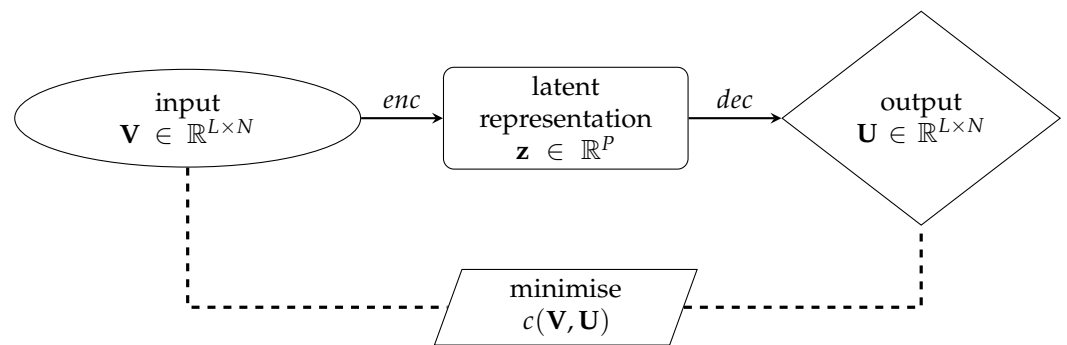


Figure 3. Schematic representation of the AutoEncoder. The loss function $c(\mathbf{V}, \mathbf{U})$ is computed as in Equation (1). For the enc and dec architectures, see Figures 1 and 2, respectively.

2.2. Solving Regression Problems

The latent representation obtained with the proposed AE can be exploited for regression tasks [17], like load identification. Starting from the MTS \mathbf{V} , assumed to be recorded by the deployed sensors, our goal is to determine the loadings applied to the structural system within the time window of interest. To this aim, we describe the loading conditions through a vector of parameters $\boldsymbol{\eta} \in \mathbb{R}^Q$; suitable probability density functions (pdfs) \mathcal{P}_q must be then associated to the entries η_q , $q = 1, \dots, Q$. A similar approach was adopted in [36], where random variables, referred to as probabilistic input data, were used to describe the geometric properties and the parameters governing the carbonation process of a reinforced concrete beam.

In the resulting problem, the independent regression variables are collected in \mathbf{V} , while the dependent ones are collected in $\boldsymbol{\eta}$. Using \mathbf{V} as regression input allows to establish the dependence between $\boldsymbol{\eta}$ and the $L \times N$ variables in \mathbf{V} or, in other words, to infer some relevant parameters ruling thousands of measurements of the structural response, without digging into the temporal correlation within each TS and the correlations among different TS. In this latter case, the use of a NN to accomplish the inference task, would require a large number of parameters in Ω to model $\boldsymbol{\eta}$ as $\boldsymbol{\eta} = \boldsymbol{\eta}(\mathbf{V})$. For this reason, it would be way better to use \mathbf{z} as regression input, being \mathbf{z} a synthetic and exhaustive description of \mathbf{V} , with P entries only. Figure 4 shows how the regression model $r : \mathbf{z} \rightarrow \boldsymbol{\eta}$ is integrated within the AE architecture.

The designed NN architecture of the regression model r is depicted in Figure 5: first, six FC layers are stacked, each one featuring a SELU activation function and empowered by skip connections [29] and batch normalization (BN) [37]; then, a FC layer with no activation function provides the output $\boldsymbol{\eta}_r$. Compared to the encoder and the decoder architectures, only one-dimensional arrays are involved in the layer definitions. Each FC layer outputs N_r channels, and N_r is therefore the only hyperparameter to tune in the proposed architecture.

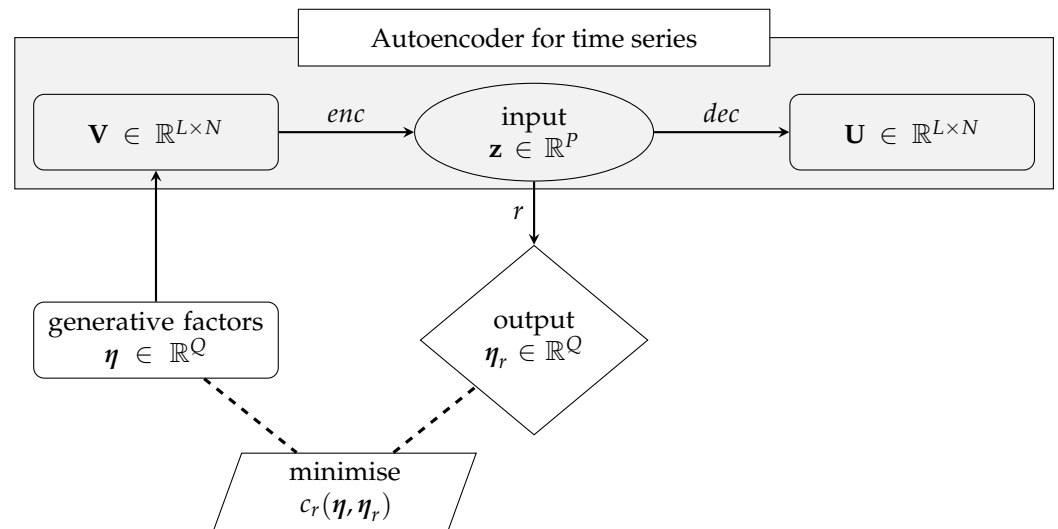


Figure 4. Schematic representation of the regression between z and η , wherein the loss function $c_r(\eta, \eta_r)$ is computed via Equation (2). For the enc and dec architectures, see again Figures 1 and 2. For the r architecture, see Figure 5.

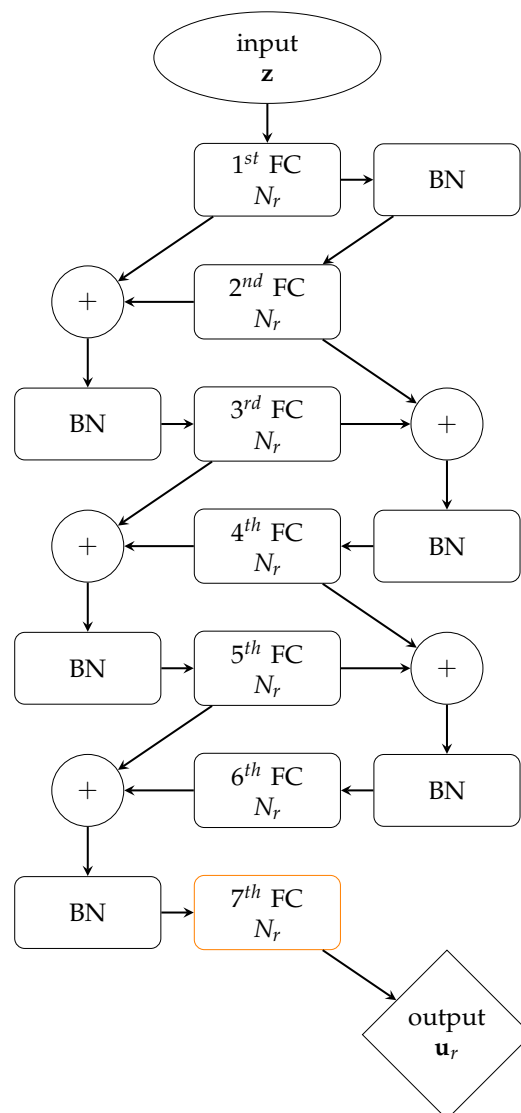


Figure 5. Architecture of the regression model; as for the notation, see the caption of Figure 1.

Such a design has been inspired by the general trend to employ deeper and deeper NNs to minimize the loss $c_r(\boldsymbol{\eta}, \boldsymbol{\eta}_r)$, defined as:

$$c_r(\boldsymbol{\eta}, \boldsymbol{\eta}_r) = \sum_{q=1}^Q (\eta_q - \eta_{rq})^2, \quad (2)$$

where $\boldsymbol{\eta}$ is the target output and $\boldsymbol{\eta}_r \in \mathbb{R}^Q$ is the output of the regression model.

Skip connections are employed to address the degradation problem usually suffered by deep NNs [38]. Denoting as $w : \mathbf{y}_{i-1} \rightarrow \mathbf{y}_i$ the transformation operated by the i -th FC layer, skip connections allow to obtain $\tilde{w} : \mathbf{y}_{i-1} \rightarrow (\mathbf{y}_i + \mathbf{y}_{i-1})$, by summing \mathbf{y}_{i-1} (the i -th layer input) to \mathbf{y}_i (the i -th layer output). As discussed in [29], this usually enhances the training procedure, since it acts as a useful preconditioning of the mapping, so that the error can be more easily back-propagated through the whole NN. BN finally addresses the additional issue of the vanishing/exploding gradient problem, by zero-centering and normalizing the input.

We underline that the two training stages featured by this architecture are run separately: first, $dec \circ enc$ is trained by minimizing $c(\mathbf{V}, \mathbf{U})$; then, r is trained by minimizing the relevant cost function $c_r(\boldsymbol{\eta}, \boldsymbol{\eta}_r)$. A combined minimization of the two loss functions would make the training more complex, without enhancing the performance of the load identification task.

3. Choice of the Latent Dimension

3.1. Generative Factors

For the present study, if the mass and stiffness properties of the structural systems are assumed to be known, the structural frequencies and vibration modes are also set. The latent variables of the AE have to describe the time variability of the applied loads and, as a result, the structural response; the dimension P of \mathbf{z} thus seems to be linked to $\boldsymbol{\eta}$ only, as this vector is responsible for the variability of the output and collects the generative factors. However, as it will be shown in the results Section, setting P according to this rationale does not prove to be a satisfactory option: to minimize the reconstruction error of the AE and to effectively solve the regression problem starting from the latent representation \mathbf{z} , P must be increased. To understand how to set P , we first try to analyse separately how enc and dec work. Compared to other dimensionality reduction techniques, like principal component analysis, the AE implicitly requires to solve an inverse problem. More precisely, enc solves an inverse problem (find the latent factors \mathbf{z} enabling the generation of the input \mathbf{V}), while dec solves the specular forward problem, (reconstruct \mathbf{V} from \mathbf{z}). As seen previously, the inverse and forward problems are jointly solved by minimising the loss function $c(\mathbf{V}, \mathbf{U})$; it is thus necessary to train the NN in an unsupervised manner, without knowing a-priori \mathbf{z} [17]. The solution of this problem may justify the need of adopting P greater than the least possible number of generative factors that one could assume just by considering the input dataset.

Indeed, as inverse problems are often ill-posed, a one-to-one correspondence between inputs in \mathbf{V} and outputs in \mathbf{z} usually does not exist.

3.2. False Nearest Neighbour Heuristics

From the discussion above, it seems hard to set a-priori the optimal number of latent variables, even when the number of generative factors for the dataset at hand is known. On one side, a too small value of P could have an impact on the decoder capability of reproducing \mathbf{V} ; on the other side, a too large value of P would introduce redundancy in the latent representation, that could ultimately spoil the minimization of $c(\mathbf{V}, \mathbf{U})$, making the understanding of the correlation between the latent variables and the model response even harder. It would be therefore useful to design a method to automatically set the optimal value of P .

Promising results in this regard come from the use of the FNN heuristics. Initially adopted for setting the appropriate embedding dimension of dynamic systems with the method of lags [39], the FNN heuristics has been recently proposed in [21] as activity regulariser for any NN employing hidden layers. The idea is to add a regularization term $q(\mathbf{z})$ to the AE loss function $c(\mathbf{V}, \mathbf{U})$, according to:

$$c_{FNN}(\mathbf{V}, \mathbf{U}, \mathbf{z}) = c(\mathbf{V}, \mathbf{U}) + \gamma Lq(\mathbf{z}), \quad (3)$$

so that unnecessary activations of the latent variables are penalized. In Equation (3), $\gamma > 0$ is an hyperparameter that sets the strength of the regularization term, while L is again the number of samples in each input TS.

If the variance of the activation values assumed by a latent variable becomes small, that variable will marginally affect the signal reconstruction, as if it were partially turned off. Therefore, once the AE is trained, the optimal number of latent variables is determined by looking at the latent variables featuring large activation variances for the dataset at hand. The crux is moved to the setting of a proper value for γ , an issue discussed in Section 4 and that we aim to further address in the future.

Coming to the regularization term, $q(\mathbf{z})$ is computed by counting the false nearest neighbors observed for different dimensions $P \leq P_{max}$ of the latent space, where P_{max} is the maximum allowed dimension. The aforementioned counting is carried out by looking at the latent representation of each MTS \mathbf{V} as a point in the space \mathbb{R}^P , and evaluating if overlapping points in \mathbb{R}^P are well separated in \mathbb{R}^{P+1} . If this is the case, these points are false neighbors in $\text{span}\{\mathbf{z}\} \in \mathbb{R}^P$, and thus increase $q(\mathbf{z})$. A variational formulation of the FNN algorithm, originally proposed in [39], was developed in [21] to be compatible with the training of the NN via gradient descent-based techniques, and has been used in this work.

To clearly report the entire methodology here proposed, Figure 6 provides a recap of what discussed so far, addressing also the offline part of the procedure, namely the steps that must be taken before the use as a monitoring system.

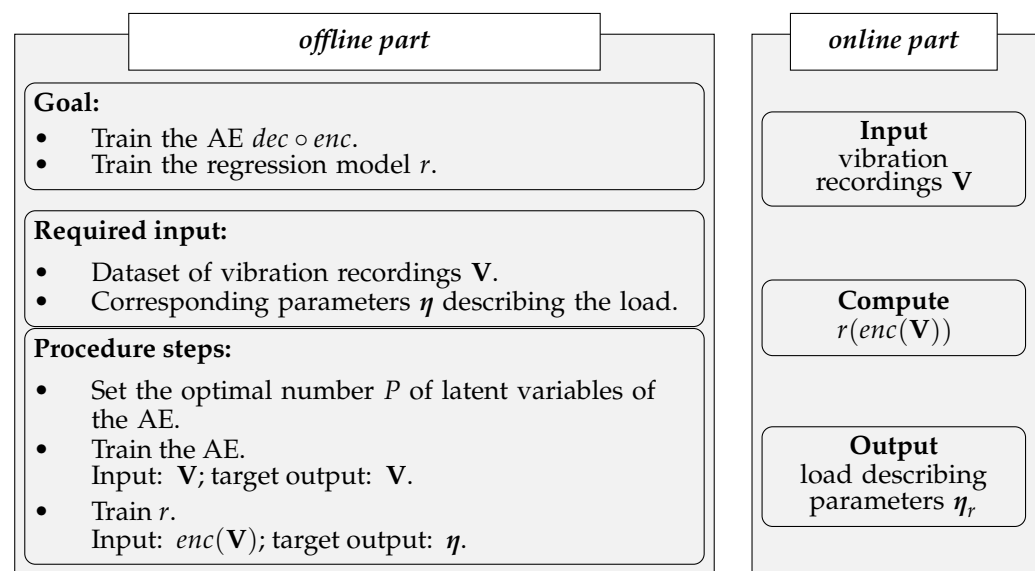


Figure 6. Proposed load identification strategy.

4. Numerical Results

Two case studies are considered in the following, with the purpose of assessing the capability of the proposed approach to identify the parameters ruling the space and time variability of the loading applied to a structural system.

The first one deals with a two-storey shear building with lateral loads acting at each storey; it can be considered as the simplest possible structure of this type. Due to the limited

need of computational resources to run the relevant structural model, a deep investigation is reported regarding the effect of the latent representation on the accuracy of the inverse problem solution, allowing also for the FNN heuristics.

The second one is the Pirelli tower in Milano, already studied, e.g., in [40–42], a cast-in-place reinforced concrete building featuring 39 storeys for a total height of 130 m. Through this numerical case study, we show the feasibility of the proposed approach to deal with the difficulties arising from a real life high-rise building, still considering horizontal excitations only. The identification of the parameters ruling load amplitude and time variability has been performed in this case by exploiting the insights gained from the first case study.

4.1. Two-Storey Shear Building

4.1.1. Shear Building Model

Shear building models are widely used in civil engineering for the vibration analysis of structures subject to lateral loads. Despite their simplicity, these models are very effective in characterizing the dynamic behavior of buildings whose floors have large out-of-plane stiffness. Their use in seismic analysis is encouraged by design codes like Eurocode 8 [43], that allow their adoption whenever requirements related to the aforementioned out-of-plane stiffness are satisfied, e.g., in terms of minimum slab thickness. To further simplify the analysis, we assume that the distribution of masses and stiffnesses are such that torsional effects can be neglected. Accordingly, the effects of lateral forces are decoupled along the two in-plan directions and the structural response of the building can be obtained by running two separate analysis, one for each lateral direction, by employing just one degree-of-freedom (dof) per floor. Damping effects have been disregarded, since they are usually not relevant in the identification of continuously excited systems [5,44]. A schematic representation of the considered two-storeys shear building model is reported in Figure 7.

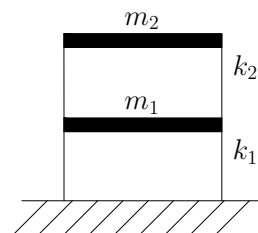


Figure 7. Two-storey shear building model. The horizontal, or lateral displacements are assumed to be recorded by the SHM system.

The time dependent lateral loads $\mathbf{F}(t) \in \mathbb{R}^{L \times N}$ are applied to the floors, with a linearly increasing amplitude along the height of the building and a sinusoidal variation in time, namely

$$F_n(t) = \frac{n}{N} \alpha \sin(2\pi\phi t) \quad n = 1, \dots, N, \quad (4)$$

where t is time; N is the number of floors (here $N = 2$); α is the load amplitude factor, and ϕ is the load frequency. The resulting parameter vector is thus $\boldsymbol{\eta} = \{\alpha, \phi\}$, with $Q = 2$ to fully describe each loading condition. To define the load, we have taken inspiration from the lateral force method described in [43], which can be adopted whenever the structural response is mainly related to the first mode of vibration of the building. The lateral force method can be adopted to define the peak actions in structural members through an equivalent static analysis under earthquake excitation; in our dynamic analyses, we have used the same variation of the load amplitude along the vertical direction. The sinusoidal variation in time has been instead adopted by considering that any dynamic loading can be decomposed into the sum of sinusoidal components through a Fourier series.

By assuming that the loading parameters α and ϕ , the floor masses and the inter-storey stiffnesses are constant in time, the building response is stationary within the time window

of interest. If within each MTS the building always features the same mass and stiffness properties, long-term degradation effects like concrete carbonation and rebars rusting, chloride and sulphate attacks for reinforced concrete are disregarded. A discussion on how to cope with the effects of the mentioned degradation processes within a SHM procedure like the one here proposed, is beyond the scope of this work. Anyhow, it is worth stressing that our procedure will be, in principle, able to address this issue by including in the generative factors the stiffness reduction of the structural members, hence by modelling the effect of damage on the building response [24,25,45].

To identify the loads applied on a two-storey shear building, the dynamic response of the structure is numerically simulated and pseudo-experimental sensor recordings, shaped as MTS to form the input dataset, are represented by the storey lateral displacements.

4.1.2. Signal Reconstruction

The AE has been first employed to reduce the dimensionality of the pseudo-experimental vibrational recordings. Each MTS $\mathbf{V} \in \mathbb{R}^{250 \times 2}$ consists of two TS reporting the sampled time evolution of the lateral floor displacements for $T = 5$ s, with a sampling rate of $\Delta t = 0.02$ s. Each \mathbf{V} is associated to a loading condition, whose governing parameters α and ϕ are sampled from two uniform pdfs $\mathcal{U}_\alpha(0.625 \times 10^3, 6.25 \times 10^3)$ N and $\mathcal{U}_\phi(1, 15)$ Hz. Two different building configurations, termed A and B and characterized by different vibration frequencies, have been considered to assess the effects of the interaction between the mentioned structural vibration frequencies and the accuracy of the results of the AE-based identification procedure. In Table 1, the data related to the two configurations are collected; note that the chosen sampling frequency for the pseudo-experimental measurements avoids signal aliasing to occur.

Table 1. Two-storey shear building model: data relevant to configurations A and B, see Figure 7.

		Configuration A	Configuration B
m_1, m_2	(ton)	625, 625	625, 1250
k_1, k_2	$\left(\frac{\text{kN}}{\text{m}}\right)$	$10^6, 10^6$	$10^6, 3 \times 10^6$
f_1, f_2	(Hz)	3.93, 10.3	3.41, 14.5

To train the AE, 16,000 MTS have been generated for both the configurations: 75% of these samples have been used to train the AE and back-propagate the error; the remaining 25% have been employed as validation set. Based on the loss $c(\mathbf{V}, \mathbf{U})$ computed on the validation set, an early-stopping strategy may be exploited. The training of the AE has been started from scratch for both configurations A and B, hence without relying on transfer learning.

For testing the trained AE, a further test set of 512 MTS has been generated. The reconstruction capacity of the AE is assessed not only qualitatively but also quantitatively, by computing for each MTS the two error measures reported in Table 2, and the relevant mean values and scattering around it over the whole training, validation and test sets.

Table 2. Adopted error measures, where $\sigma(\mathbf{v}_n)$ is the standard deviation of the TS \mathbf{v}_n .

standardized L^2 norm	$\frac{\ \mathbf{u}_n - \mathbf{v}_n\ _2}{\sigma(\mathbf{v}_n)}$
standardized L^∞ norm	$\frac{\ \mathbf{u}_n - \mathbf{v}_n\ _\infty}{\sigma(\mathbf{v}_n)}$

Results in the following refer to AE hyperparameters gathered in Table 3: training has been repeated for different sets of hyperparameters, and the one adopted led to the minimum loss $c(\mathbf{V}, \mathbf{U})$ and has been finally selected for the analyses. As for the kernels of the convolutional layers, their dimensions have been set to cope with the fundamental period of structural vibrations, so as to exploit the capacity of convolutional layers to detect a correlation within a TS. Alternatives to this trial-and-error procedure are represented

by the Bayesian methods [46] or by multi-objective optimization [47], which are anyway computationally infeasible in practical situations of interest for our study. The drawbacks of both approaches may be mitigated by their combination, as suggested in [48].

Table 3. Two-storey shear building model: AE hyperparameters.

Encoder		Decoder	
$[N_e^{inc1}, N_e^{inc2}, N_e^{inc3}]$	$[3N, 3N, 6N]$	N_d	$48N$
$[H_e^{inc1}, H_e^{inc2}, H_e^{inc3}]$	$[13, 8, 5]$		
N_e^2	$6N$		
$[H_e^1, H_e^2, H_e^3]$	$[8, 5, 3]$		

Regarding the impact of P on the reconstruction capacity of the AE, it has been assessed by means of its effects on mimicking the structural response in the frequency domain, starting from the reduced representation. As previously discussed, a lower bound on P can be assumed to be equal to 2, i.e., equal to the number of generative factors in η . Given the ill-posedness of the inverse problem, by adopting $P > 2$ we expect to have beneficial effects on the AE performance; therefore, we have tested the cases $P = \{2, 3, 4, 5, 6\}$. As shown in Figures 8–11, the reconstruction error gets progressively reduced by increasing P for both the considered configurations, but unevenly and showing different correlations with the load frequency ϕ . This behavior is due to the stochastic nature of the training algorithm and also to the strong nonlinearity of $c(\mathbf{V}, \mathbf{U})$. More specifically: Figures 8 and 9 report the error, via the standardized L^2 norm, in reconstructing the displacement \mathbf{v}_1 of the first floor as a function of ϕ , when the input signals, respectively, belong to the training and validation sets; Figures 10 and 11 report instead the error, via the standardized L^∞ norm, in reconstructing the displacement \mathbf{v}_2 of the second floor, when the input signals belong to the test set only. For comparison with these plots, Figures 12 and 13 further provide, for configurations A and B, a sketch of the reconstruction capacity for the training and validation sets via the standardized L^∞ norm, and a sketch of the reconstruction capacity for the test set via the standardized L^2 norm, both for $P = 4$. Similar results have been obtained for the other values of P , but are not reported here for the sake of brevity. Such results are shown since, if the reconstruction capacity for the training and validation sets were greater than the one related to the test set, overfitting would have probably spoiled the AE performance: the NN would not acquire any generalization capacity, being limited to reproduce the instances seen during the training.

The investigated reconstruction capacity is less affected by the load amplitude α , as shown in Figure 14, due to the linearity of the structural behavior. Indeed, when the standardized L^∞ norm is considered, as it measures the inaccuracy in the peak reconstruction, larger errors are found for values of α smaller than 2000 N. In spite of the data normalization procedure preceding training, the structural displacements under excitations featuring small values of α have small peaks too, and their incorrect reconstruction results less penalized during the training.

The link between the reconstruction error and the load frequency ϕ varies with P , and depends on the adopted error measure. Figures 8 and 9 have shown that the standardized L^2 error is larger when ϕ gets closer to the structural vibration frequencies f_1 and f_2 , that is when the load induces a resonant response of the structure. This outcome is somehow expected, as the relevant beats in the displacement recordings are signal characteristics hard to catch by the AE. The larger error found for $\phi \approx f_1$ results as a consequence of the *dec* difficulty of reproducing the long-range temporal correlation characterizing the first vibration mode.

Figures 10 and 11 have shown instead that the standardized L^∞ error is still large when ϕ gets close to the second structural vibration frequency, while becomes rather small, roughly by ten times, for $\phi \approx f_1$. An analysis of the dynamics of the two configurations suggests the reason behind this result. During training, the loss function allows modifying

more largely the weights Ω , when the AE fails to reconstruct the vibration mode that has a larger impact on the dynamic response of the structure. The excitation frequency ϕ is sampled from \mathcal{U}_ϕ and the mentioned modes can have a different impact for different instances. To compute the impact that the vibration modes $\psi_s, s = 1, 2$, have on the solution, we first solve a non-standard eigenvalue problem in the form [49]:

$$\left[\mathbf{K} - (2\pi f_s)^2 \mathbf{M} \right] \psi_s = \mathbf{0} \quad (5)$$

the stiffness and mass matrices of the structure being:

$$\mathbf{K} = \begin{bmatrix} k_1 + k_2 & -k_2 \\ -k_2 & k_2 \end{bmatrix} \quad \mathbf{M} = \begin{bmatrix} m_1 & 0 \\ 0 & m_2 \end{bmatrix},$$

and enforce $\psi_s^T \mathbf{M} \psi_s = 1$ as normalization rule.

The equations governing the dynamics of the structure read:

$$\mathbf{M} \ddot{\mathbf{v}}(t) + \mathbf{K} \mathbf{v}(t) = \mathbf{F}(t), \quad (6)$$

where: $\ddot{\mathbf{v}}(t) \in \mathbb{R}^2$ and $\mathbf{v}(t) \in \mathbb{R}^2$ are the vectors of storey accelerations and displacements, respectively; $\mathbf{F}(t)$ is the vector of the external loads. For each load case, by sampling $\mathbf{v}(t)$ at the two floors, we obtain $\mathbf{v}_1, \mathbf{v}_2 \in \mathbb{R}^{250}$ and the instance $\mathbf{V} = [\mathbf{v}_1, \mathbf{v}_2]$.

Due to the linear behavior of the structure, through modal superposition Equation (6) is decoupled as follows:

$$\psi_s^T \mathbf{M} \psi_s \ddot{x}_s(t) + \psi_s^T \mathbf{K} \psi_s x_s(t) = \psi_s^T \mathbf{F}(t), \quad (7)$$

with:

$$\mathbf{v}(t) = \psi_1 x_1(t) + \psi_2 x_2(t). \quad (8)$$

Since $\psi_s^T \mathbf{M} \psi_s = 1$, $\psi_s^T \mathbf{K} \psi_s = (2\pi f_s)^2$ and we obtain:

$$\ddot{x}_s(t) + (2\pi f_s)^2 x_s(t) = \psi_s^T \mathbf{F}(t). \quad (9)$$

If the structure is initially at rest and if the entries of the load vector \mathbf{F} are defined according to Equation (4), the time history of $x_s(t)$ is given by:

$$x_s(t) = \frac{\alpha}{(2\pi f_s)^2 - (2\pi\phi)^2} \left(-\frac{\phi}{f_s} \sin(2\pi f_s t) + \sin(2\pi\phi t) \right) \Gamma_s, \quad (10)$$

where:

$$\Gamma_s = \psi_s^T \begin{bmatrix} 0.5 \\ 1 \end{bmatrix} \quad (11)$$

actually depends on the structural dynamics (through ψ_s) and on the spatial distribution of loads.

At a specific time instant \bar{t} , the modal response becomes $x_s(\bar{t}) = \bar{x}_s$, whose expected value $E[\bar{\mathcal{X}}_s]$ can be computed as:

$$E[\bar{\mathcal{X}}_s] = \int_{\alpha_m}^{\alpha_M} \int_{\phi_m}^{\phi_M} \left\{ \frac{\alpha}{(2\pi f_s)^2 - (2\pi\phi)^2} \left(-\frac{\phi}{f_s} \sin(2\pi f_s \bar{t}) + \sin(2\pi\phi \bar{t}) \right) \Gamma_s \mathcal{U}_\alpha \mathcal{U}_\phi \right\} d\alpha d\phi. \quad (12)$$

where we have accounted for that α and ϕ , respectively, vary in the ranges (α_m, α_M) and (ϕ_m, ϕ_M) . Computing the integrals, we obtain:

$$E[\bar{\mathcal{X}}_s] = \left\{ \frac{\alpha_M - \alpha_m}{2} \left(\frac{\sin(2\pi f_s \bar{t})}{8\pi^2 f_s} \ln \frac{f_s^2 - \phi_M^2}{f_s^2 - \phi_m^2} + I \right) \right\} \Gamma_s, \quad (13)$$

where:

$$I = \int_{\phi_m}^{\phi_M} \frac{\sin(2\pi\phi\bar{t})}{4\pi^2(f_s^2 - \phi^2)} d\phi.$$

The term within curly brackets in Equation (13) provides the dependence of $E[\bar{\mathcal{X}}_s]$ on α and ϕ .

At the same time instant, the expected value of the storey displacement $\bar{v}_n = v_n(\bar{t})$, $n = 1, 2$, thought of as sampled from the corresponding pdf $\bar{\mathcal{V}}_n$, is obtained by exploiting Equation (8) and the linearity of the expectation rule [50]:

$$E[\bar{\mathcal{V}}_n] = \psi_{1n}E[\bar{\mathcal{X}}_1] + \psi_{2n}E[\bar{\mathcal{X}}_2]. \quad (14)$$

The contribution to $E[\bar{\mathcal{V}}_n]$ of each mode depends linearly on $E[\bar{\mathcal{X}}_1]$ and $E[\bar{\mathcal{X}}_2]$, and therefore on Γ_1 and Γ_2 . For the case at hand, the ratio between Γ_1 and Γ_2 is equal to 9.67 for configuration A, and to 7.33 for configuration B. Accordingly, the error provided by the AE in reconstructing the contribution of the first vibration mode is, on average, roughly ten times larger than the error linked to the second vibration mode. The loss function $c(\mathbf{U}, \mathbf{V})$ leads to the setting of the NN weights in the same way. Due to this rationale, the AE is driven to learn better the first vibration mode.

In such a discussion, we have disregarded the temporal dependence of $\mathbf{v}(t)$; this has an impact on the AE capacity of accounting for each mode of vibration. In the comment to Figures 8 and 9, we have already addressed that the tendency to learn better the first mode of vibration is counterbalanced by the long range temporal correlation featured by the first mode. The adopted error measures have been introduced with the purpose of investigating these issues, and seem to adequately address them.

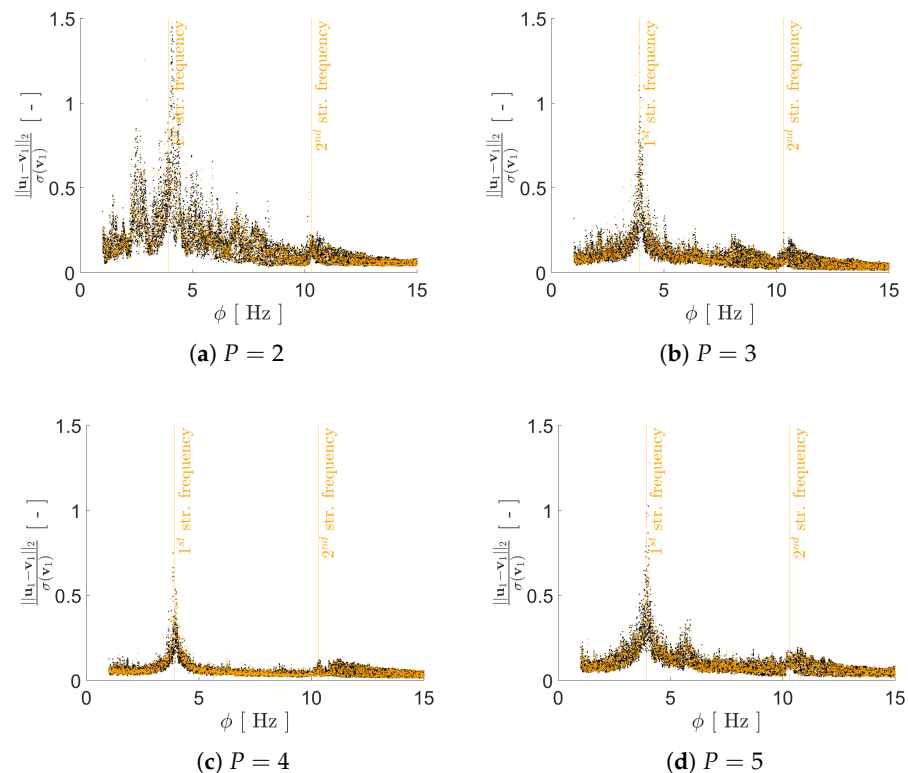


Figure 8. Cont.

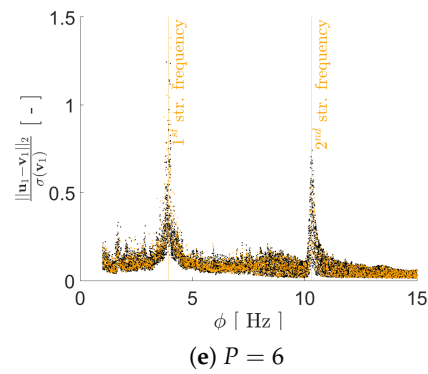


Figure 8. Two-storey shear building, configuration A. Reconstruction error for floor displacement v_1 via the standardized L^2 norm, as a function of the load frequency ϕ and for a varying value of P . In the charts, black dots refer to the training set, while orange dots refer to the validation set.

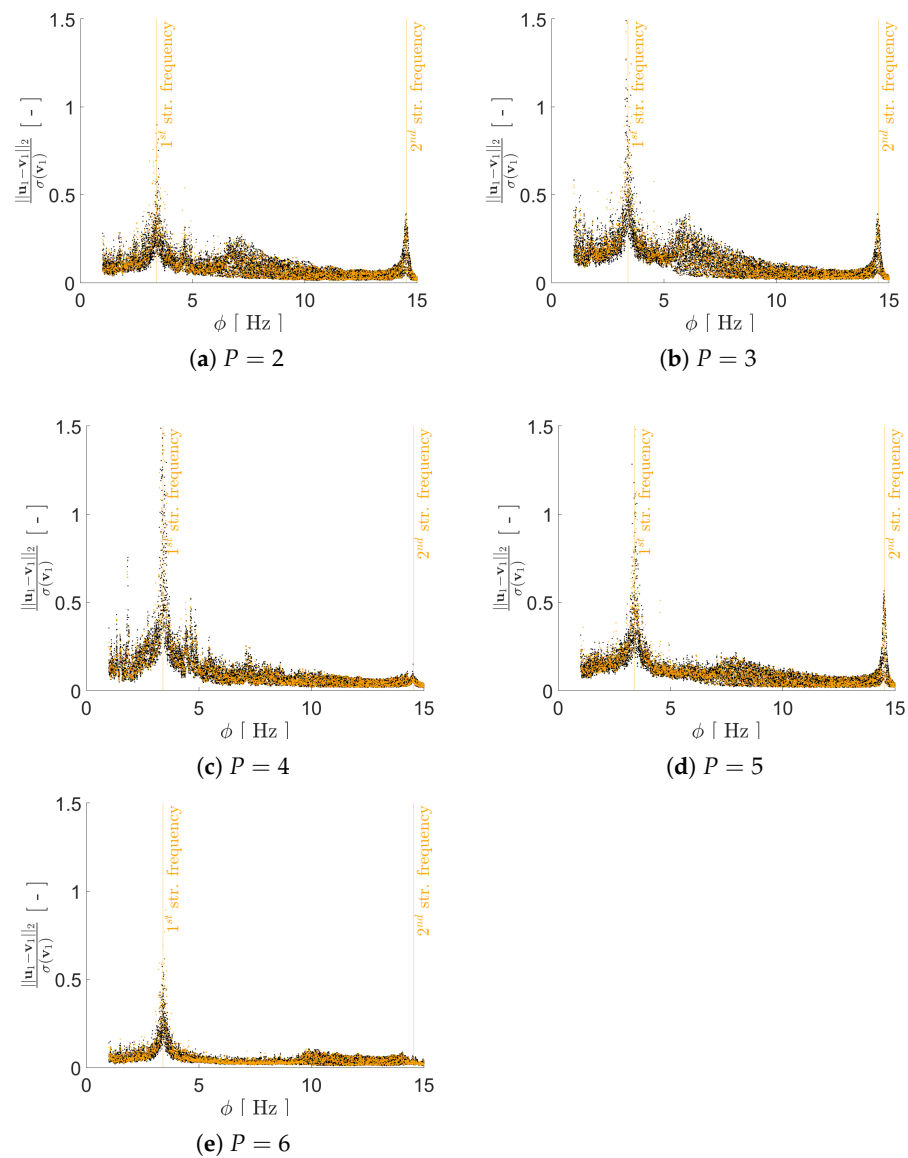


Figure 9. Two-storey shear building, configuration B. Reconstruction error for floor displacement v_1 via the standardized L^2 norm, as a function of the load frequency ϕ and for a varying value of P . In the charts, black dots refer to the training set, while orange dots refer to the validation set.

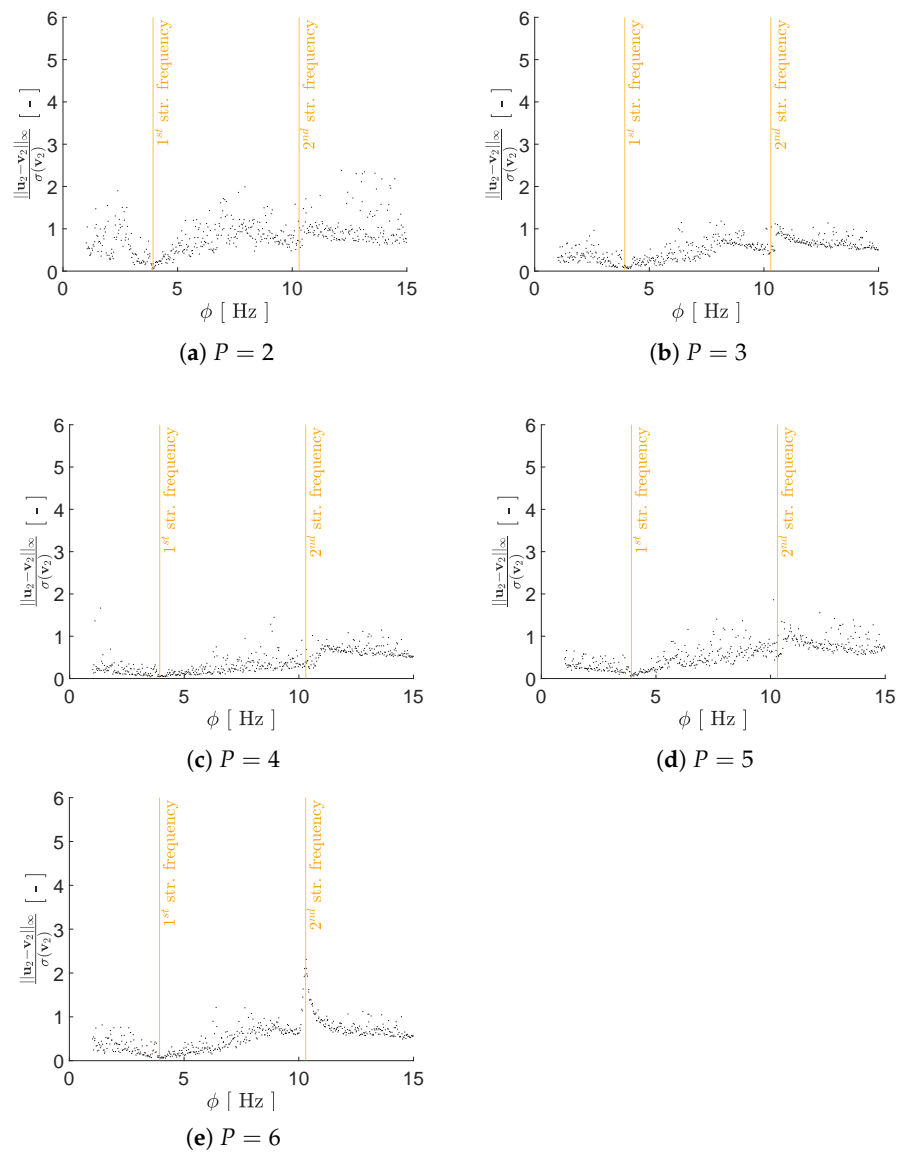


Figure 10. Two-storey shear building, configuration A. Reconstruction error for floor displacement v_2 via the standardized L^∞ norm, as a function of the load frequency ϕ and for a varying value of P . In the charts, the dots refer to the test set.

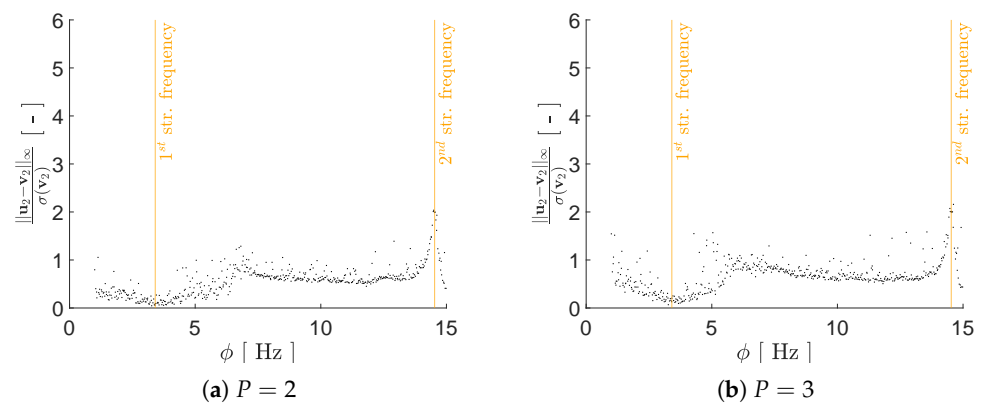


Figure 11. Cont.

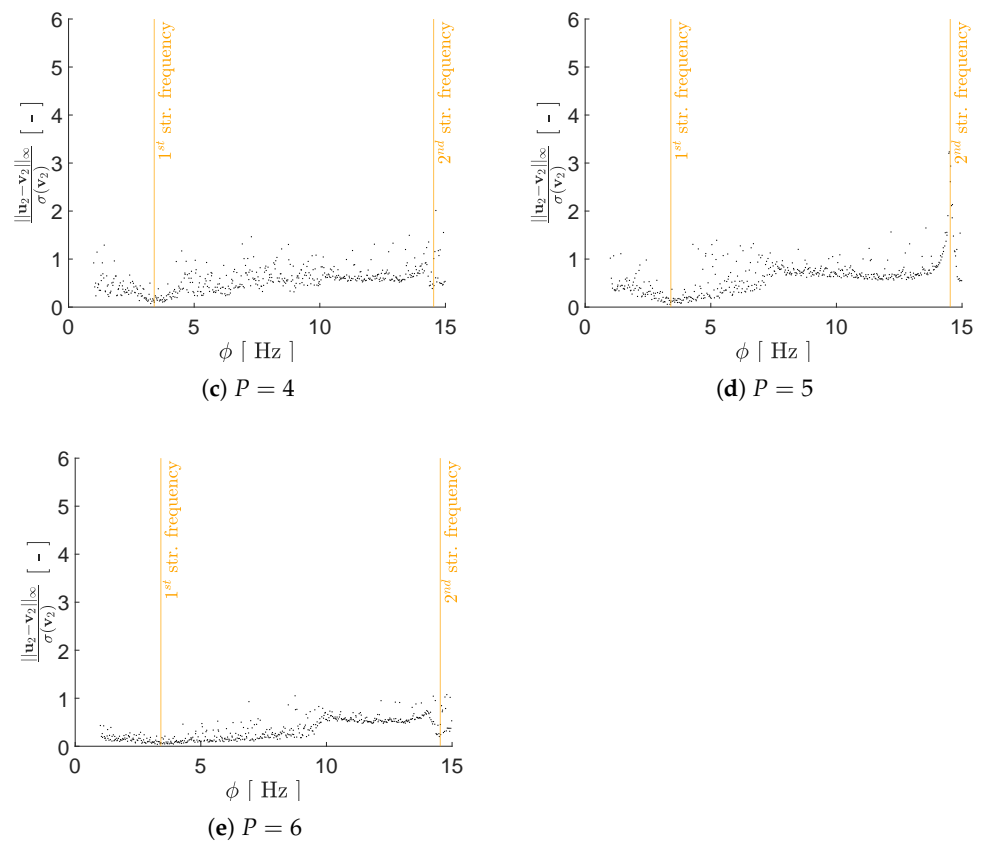


Figure 11. Two-storey shear building, configuration B. Reconstruction error for floor displacement v_2 via the standardized L^∞ norm, as a function of the load frequency ϕ and for a varying value of P . In the charts, the dots refer to the test set.

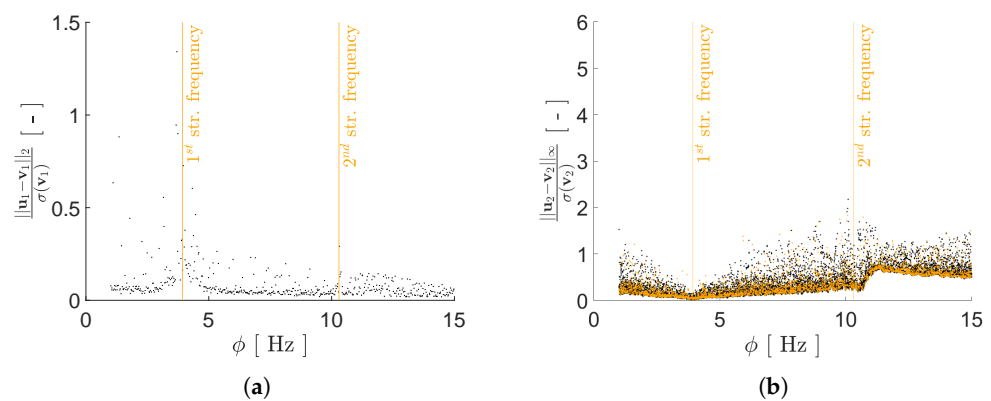


Figure 12. Two-storey shear building, configuration A, $P = 4$. Reconstruction errors as a function of the load frequency ϕ : (a) standardized L^2 norm relevant to the floor displacement v_1 for the test set; (b) standardized L^∞ norm relevant to the floor displacement v_2 for the training (black dots) and validation (orange dots) sets.

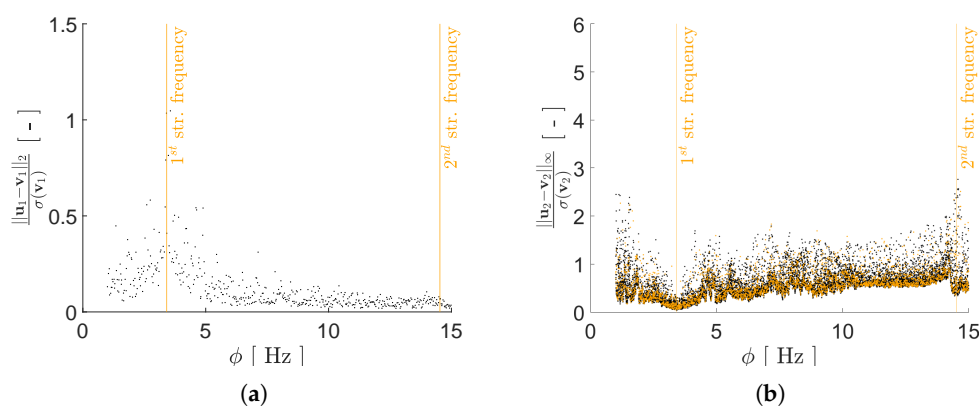


Figure 13. Two-storey shear building, configuration B, $P = 4$. Reconstruction errors as a function of the load frequency ϕ : (a) standardized L^2 norm relevant to the floor displacement \mathbf{v}_1 for the test set; (b) standardized L^∞ norm relevant to the floor displacement \mathbf{v}_2 for the training (black dots) and validation (orange dots) sets.

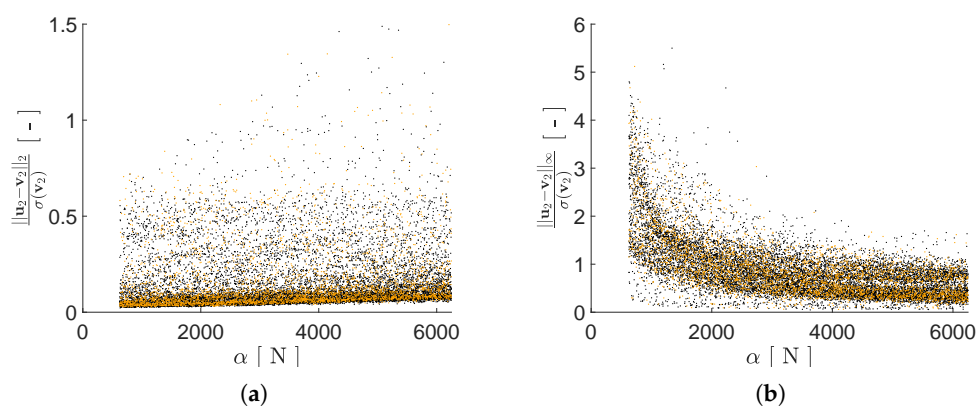


Figure 14. Two-storey shear building, configuration A, $P = 4$. Reconstruction errors as a function of the load amplitude α , relevant to the floor displacement \mathbf{v}_2 for the training (black dots) and validation (orange dots) sets: (a) standardized L^2 norm; (b) standardized L^∞ norm.

Figure 15 shows a comparison of the reconstructed \mathbf{u}_1 and the input \mathbf{v}_1 signals taken from the validation set, either for $\phi \approx f_1$ or $\phi \approx f_2$ of configuration B and for $P = 6$, to further get insights into what the two error norms provide. In spite of the rather large reconstruction error measured by the standardized L^2 norm and shown by Figure 9e for the first resonant frequency, \mathbf{u}_1 and of \mathbf{v}_1 in Figure 15a are almost perfectly superposed. This comparison confirms that both error measures bring meaningful information, with the standardized L^2 norm measuring inaccuracies in the reproduction of the frequency content of the input signal, while the standardized L^∞ norm highlighting the inability to catch peaks in the same input signal.

To better assess the impact of P on the reconstruction capacity of the AE, box plots depicting the mean and the scattering around it for the two adopted error norms are reported in Figures 16 and 17 for configurations A and B, respectively. In the charts, errors are given for both the training and test sets, to evaluate the generalization capacity of the AE. As a general rule, the values of the load configuration-dependent reconstruction error evaluated for the test set is more scattered than the one evaluated for the training set, while the relevant median values are quite similar; the said difference is larger if measured through the standardized L^2 norm.

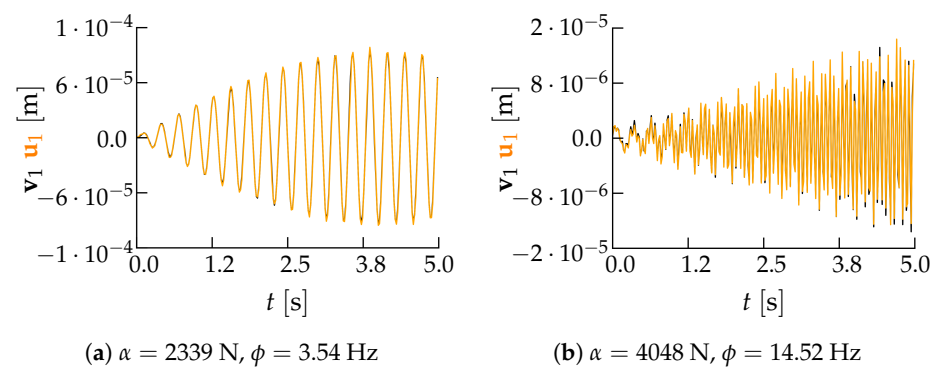


Figure 15. Two-storey shear building, configuration B, $P = 6$. Comparison between input (black) and reconstructed (orange) time histories of the first floor displacement for two cases belonging to the validation set and for load frequencies (a) $\phi \approx f_1$ and (b) $\phi \approx f_2$.

According to Figure 16, the optimal number of latent variables for configuration A results to be $P = 4$ when looking at the standardized L^2 norm, and $P = 3$ when looking at the standardized L^∞ norm if outliers are also allowed for. By increasing P and, therefore, the redundancy in the latent representation, an improvement of the AE reconstruction capacity is not achieved. As shown in Figure 17, also for configuration B an increase of the value P does not lead to a monotonic reduction of the reconstruction error. Even if the best AE accuracy has been obtained for $P = 6$, good performances have been attained with $P = 4$ too, with a slight deterioration for $P = 5$.

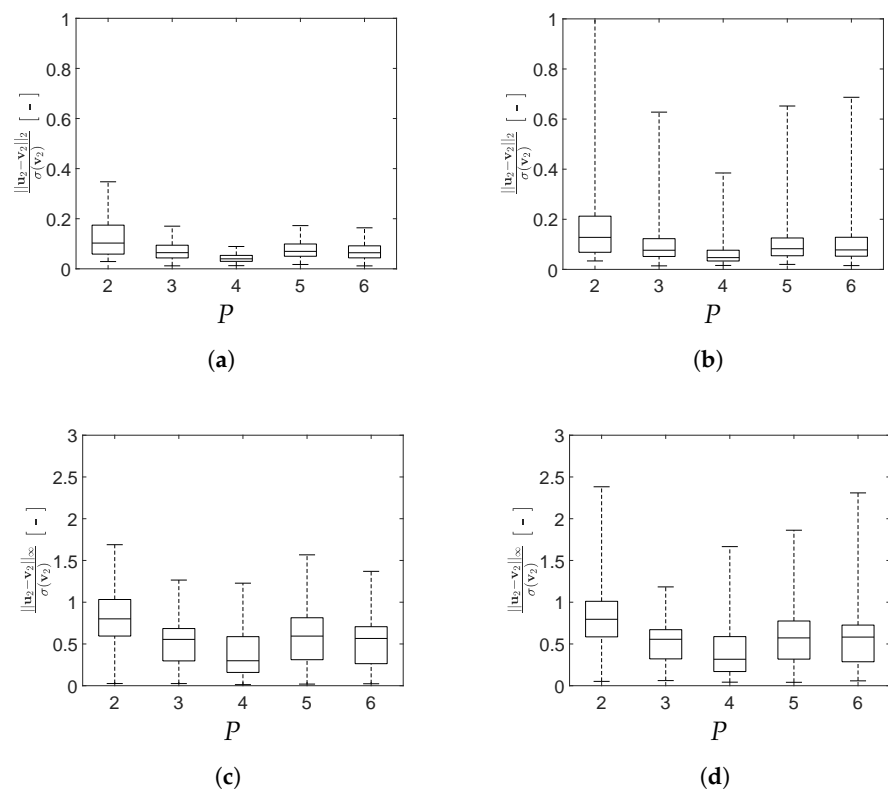


Figure 16. Two-storey shear building, configuration A. Effect of P on the statistics of the reconstruction errors relevant to the floor displacement v_2 , represented through box plots: (top row) standardized L^2 norm, and (bottom row) standardized L^∞ norm; data taken from (left column) training set, and (right column) test set.

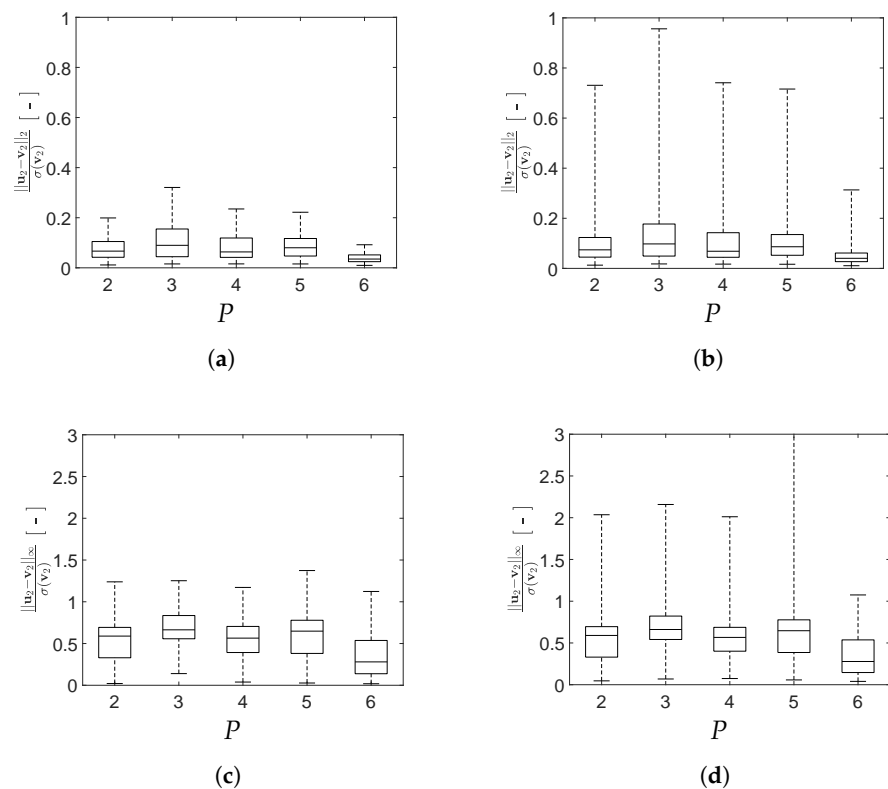


Figure 17. Two-storey shear building, configuration B. Effect of P on the statistics of the reconstruction errors relevant to the floor displacement \mathbf{v}_2 , represented through box plots: **(top row)** standardized L^2 norm, and **(bottom row)** standardized L^∞ norm; data taken from **(left column)** training set, and **(right column)** test set.

Moving deeper into the assessment of the AE performances, a comparison is reported in Figure 18 between the reconstruction errors for both configurations A and B. For the standardized L^2 norm, the variation relevant to the error values for configuration B is slightly smaller than that relevant to configuration A. A similar trend can be recognized also for the standardized L^∞ norm. This outcome can be linked to the smaller gap between the resonance frequencies f_1 and f_2 featured by configuration A; it is worth mentioning that a similar difficulty was already observed with methods like independent component analysis or second order blind identification, when the identification of closely spaced modes is involved [51].

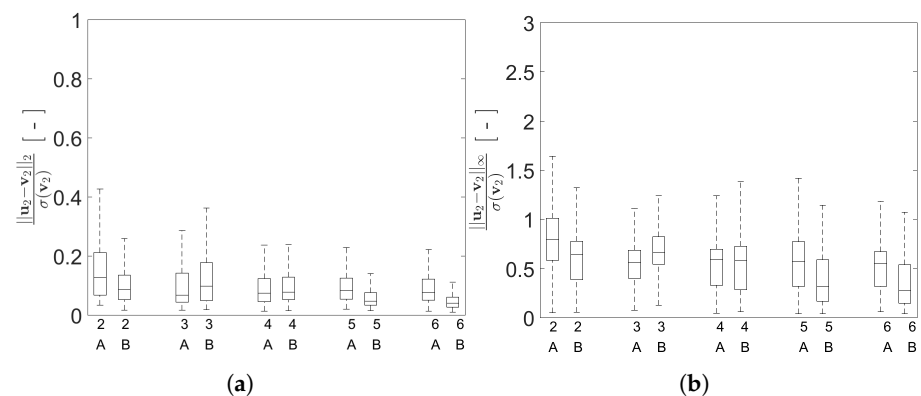


Figure 18. Two-storey shear building, effect of P on the statistics of the reconstruction errors relevant to the floor displacement \mathbf{v}_2 , represented through box plots, for data taken from the test set and both structural configurations A and B: **(a)** standardized L^2 norm; **(b)** standardized L^∞ norm.

4.1.3. False Nearest Neighbour Heuristics

If the FNN heuristics is included in the AE loss function formulation, a regularization term $\varrho(\mathbf{z})$ is added to $c(\mathbf{V}, \mathbf{U})$ in accordance with Equation (3). As an outcome, Figure 19 reports the variance $\sigma(z_p)$, with $p = 1, \dots, P_{max} = 6$, of each latent variable for the training set of configuration A, at varying value of the regularization parameter γ in the added term $\varrho(\mathbf{z})$. Except the case $\gamma = 10^{-3}$, for which the regularization term appears to be too small, $c_{FNN}(\mathbf{V}, \mathbf{U}, \mathbf{z})$ allows to automatically turn off some of the latent variables. Increasing values of γ are not associated to a clear trend in the number of deactivated latent variables. Indeed, the way in which the plain AE loss function $c(\mathbf{V}, \mathbf{U})$ and the regularization term $\gamma\varrho(\mathbf{z})$ affect the solution is made practically unpredictable by the strong nonlinear behavior of $c_{FNN}(\mathbf{V}, \mathbf{U}, \mathbf{z})$.

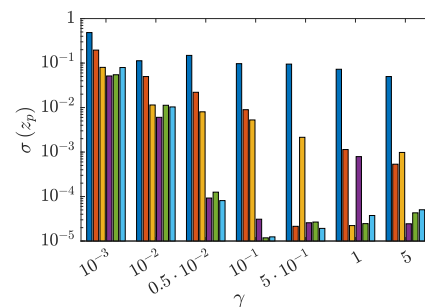


Figure 19. Two-storey shear building, configuration A. Effect of the regularization parameter γ on the value of the variance of each latent variable in \mathbf{z} , evaluated on both the training and test sets for $P_{max} = 6$.

Even the reconstruction capacity of the AE does not show a clear trend at varying γ , as highlighted in Figure 20 in terms of the results obtained for the training and test sets relevant to configuration A. Though a non-monotonic variation of the AE performance is obtained, the FNN heuristics can be exploited to set the value of γ for which the mean value of the reconstruction error and the variation around it, are minimized. For configuration A, the minimum is obtained for $\gamma \approx 10^{-1}$, irrespective of the handled standardized norm and the dataset. For cases featuring such values of γ , the number of active latent variables is $P = 3$, the same found as optimal using the standardized L^∞ error norm. Even though they are not shown here for the sake of brevity, results relevant to configuration B are characterized by a slightly more regular effect of γ on the number of active variables in \mathbf{z} , with a sub-optimal solution attained with the same range of values for the regularization term, endowed with $P = 4$. In this case, the FNN heuristics has been able to attain a sub-optimal solution, but not the optimal one, featuring instead $P = 6$.

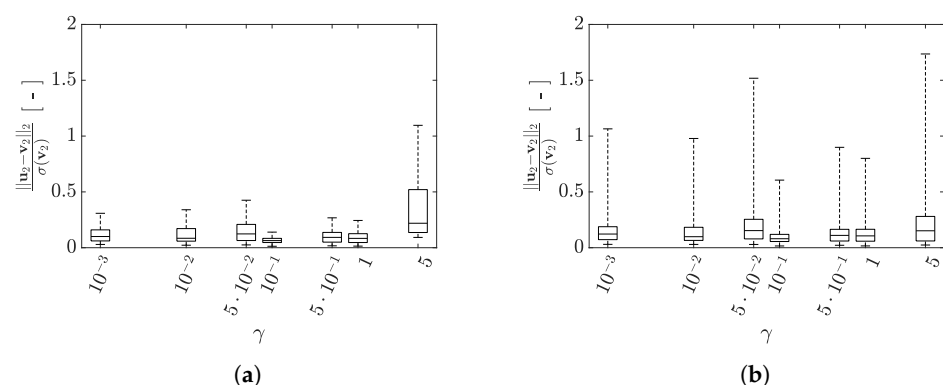


Figure 20. Cont.

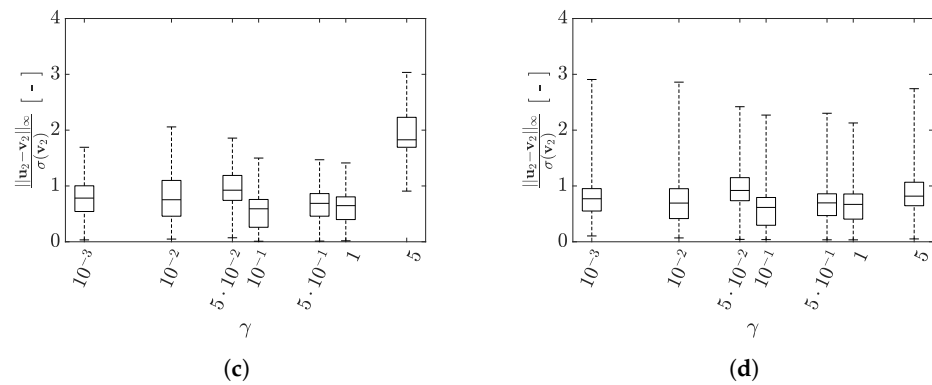


Figure 20. Two-storey shear building, configuration A. Effect of γ on the statistics of the reconstruction errors relevant to the floor displacement \mathbf{v}_2 , represented through box plots: **(top row)** standardized L^2 norm, and **(bottom row)** standardized L^∞ norm; data taken from **(left column)** training set, and **(right column)** test set.

The regularization of the loss function based on the FNN heuristics thus allows to basically achieve the same optimal AE settings already found. The strength of the regularization term seems to have a non negligible impact on the results. For this reason, the use of $c_{FNN}(\mathbf{V}, \mathbf{U}, \mathbf{z})$ leads to marginal advantages for the proposed cases study, given that the tuning of P is substituted by the tuning of γ . On the other hand, this approach can be considered useful if there is no a clear understanding on the number of generative factors.

4.1.4. Load Identification

As already pointed out, the latent representation \mathbf{z} can be exploited to reduce the computational cost of solving inverse problems. For the present case, the identification of the loading parameters $\boldsymbol{\eta}$ on the basis of the relevant observations in \mathbf{V} , has been approached by first exploiting the encoder function enc , to obtain \mathbf{z} , and then by applying the regression model r . Main results are collected in Figure 21, where the effects of P on the identification of α and ϕ are reported for configuration A. In the graphs, the predicted values are reported along the vertical axis and are displayed against the corresponding ground-truth data reported along the horizontal axis: the perfectly identified values in each chart are therefore those aligned with the line $\alpha_r = \alpha$ or $\phi_r = \phi$. The accuracy of the regression model r is further assessed through Table 4, reporting the root mean square error (RMSE) between the predicted and the actual values, and the Pearson correlation coefficient R^2 : when $R^2 = 1$, an exact linear correlation between predicted and ground-truth data is found, indicating the perfect alignment with the line $\alpha_r = \alpha$ or $\phi_r = \phi$.

The reconstruction accuracy of the AE has a direct impact on the regression accuracy, which is accordingly higher for $P = 4$ and lower for $P = 2$. A non monotonic dependence of the AE accuracy on the latent space size P is shown by the values of the RMSE (the lower, the better) and R^2 (the higher, the better) when both α and ϕ are targeted. Both indicators improve moving from $P = 2$ to $P = 4$, deteriorate for $P = 5$, get better for $P = 6$ but still without reaching the performance for $P = 4$. These results are in agreement with the box plots in Figure 16 and, partially, with the indications of FNN heuristics, pointing out $P = 4$ as the optimal dimension for the latent representation. Regarding the load frequency ϕ , the regressor r faces some difficulties in the frequency range within which also the reconstruction capacity of the AE has been shown to be detrimentally affected. Indeed, by comparing Figures 10e and 21f, the scattering in the prediction for $\phi > f_2$ corresponds to the high error measured via the standardized L^∞ norm.

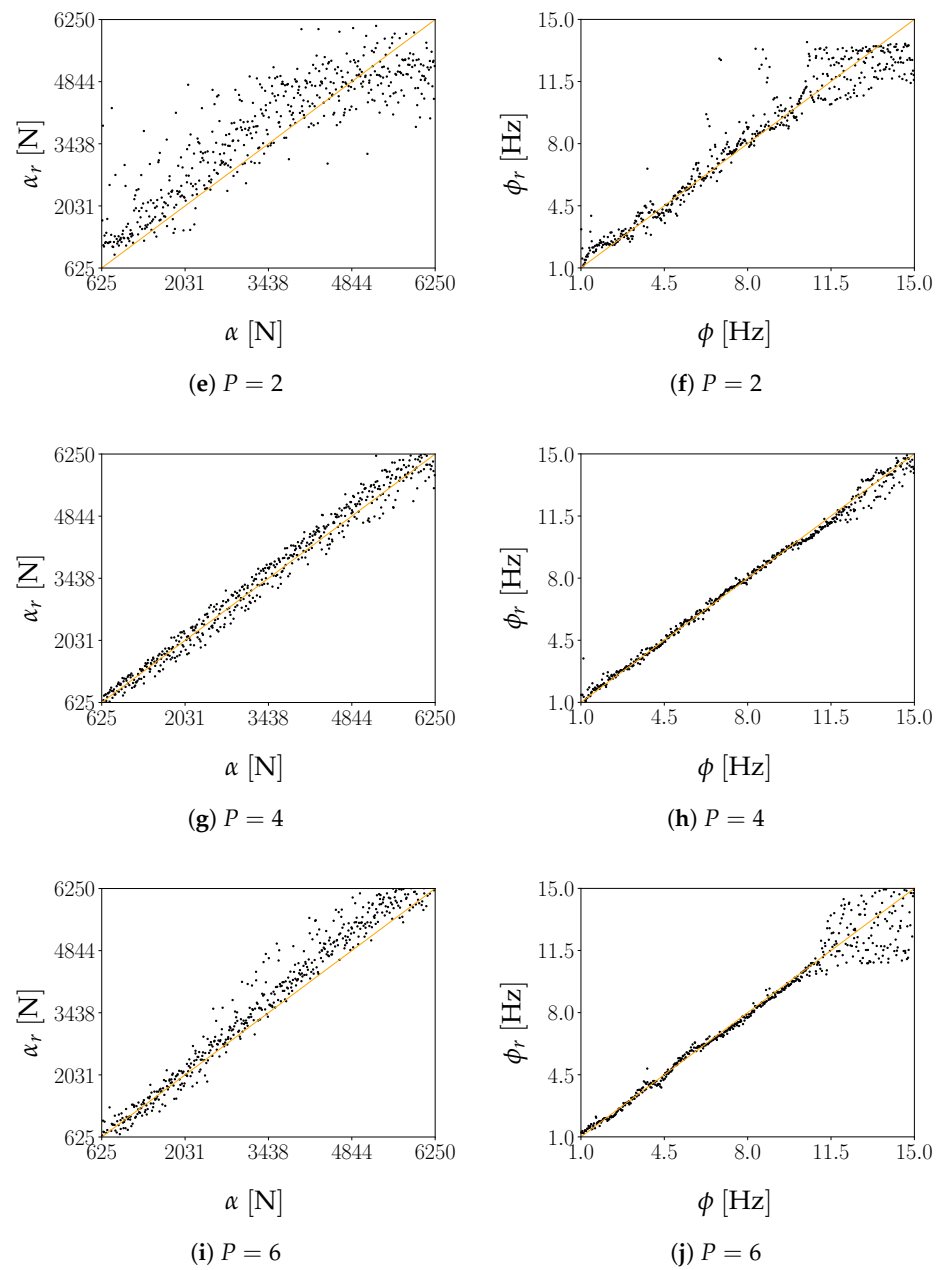


Figure 21. Two-storey shear building, configuration A, test set. Parity plots showing the regression outcomes for the load amplitude α (left column) and the load frequency ϕ (right column), at varying dimension P of the latent representation, against the ground-truth data.

Table 4. Two storey shear building, configuration A, test set. RMSE and correlation coefficient R^2 of the regression outcomes for the load amplitude α and the load frequency ϕ , at varying dimension P .

P	α		ϕ	
	RMSE [N]	R^2 [-]	RMSE [Hz]	R^2 [-]
2	1263	0.654	1.74	0.912
3	1243	0.717	1.11	0.963
4	552	0.941	0.69	0.987
5	1004	0.803	1.16	0.960
6	769	0.897	1.28	0.966

4.2. Pirelli Tower

A real-life case is now considered, to get more insights regarding the capability of the proposed approach: the Pirelli Tower in Milan (see Figure 22a). This tower is a 39 storey building, 35 of which out of ground, having a total height of 130 m. A schematic representation of the standard floor, whose dimensions are approximately 70×20 m, is reported in Figure 22b. Due to its constant plan geometry, a shear building model was proven to describe well the dynamic response of the tower under horizontal actions: the floors behave like rigid diaphragms, connected through compliant columns modelling the lateral load resisting system. The soil-structure interaction at its basement has been modeled via a lumped-parameter approach: additional details were thoroughly reported in [40].

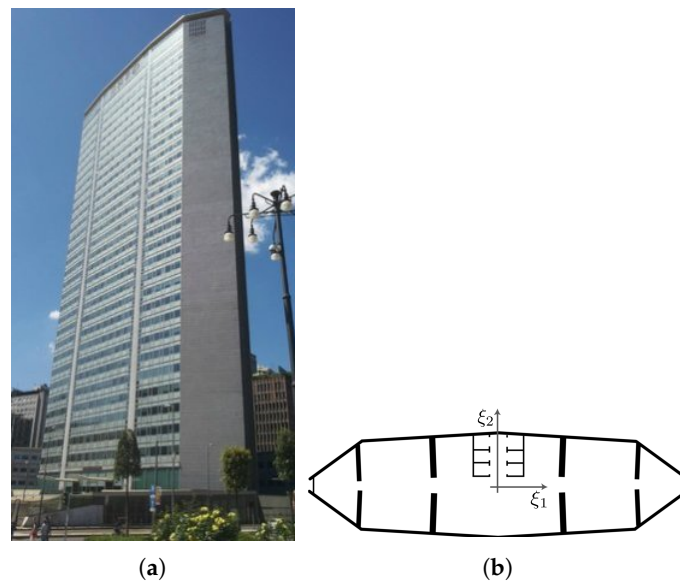


Figure 22. The Pirelli Tower in Milan. (a) Picture taken from Piazza Duca D'Aosta, and (b) schematic plan of a standard floor.

To slightly simplify the analysis, torsional effects due to the small eccentricity of the center of stiffness with respect to the center of mass along the ξ_2 axis, as induced by the asymmetric central core, have been disregarded. The lateral loads given by Equation (4) have been applied along the ξ_1 direction; the amplitude α has been sampled from the uniform pdf $\mathcal{U}_\alpha(2 \times 10^3, 20 \times 10^3)$ N; the load frequency ϕ has been instead sampled from three uniform pdfs $\mathcal{U}_\phi^1(1, 6)$ Hz, $\mathcal{U}_\phi^2(1, 9)$ Hz, and $\mathcal{U}_\phi^3(1, 15)$ Hz, respectively, defined as load cases 1, 2 and 3. The load case 3 features the same frequency range used for the two-storey building; load cases 1 and 2 have been instead designed aiming to ease the load identification task, as can be ascertained by comparing the relevant frequency ranges with the fundamental vibration ones of the building, gathered in Table 5.

A pseudo-experimental monitoring frame has been adopted, assuming that the lateral displacements at the 20th and at the 39th floors are recorded for $T = 5$ s, with a sampling frequency of 50 Hz to avoid aliasing in relation to the first 13 vibration modes of the structure. An optimal placement of the sensors to monitor the health of this building was discussed in [42], by exploiting a Bayesian experimental design that requires a detailed description of the parameter uncertainties. An optimization of sensor placement based on non-probabilistic interval analysis may be instead preferable when such a description of the uncertainties can not be fully provided [47]. To train the AE 32,000 MTS have been generated, still keeping 75% of the samples for training and 25% of them for validation. The number of instances have been doubled with respect to the two-storey case, due to the greater complexity of the present model. Regarding computational costs and resources, an RTX 2080 Ti (GPU) has been exploited for both the training and the testing stages.

For the case at hand, the training of the AE has lasted 1 h and 50 min, while the training of the regressor r has lasted about 10 min. The testing has resulted instead almost inexpensive, since each instance is processed in about 0.05 s.

Table 5. Fundamental vibration frequencies of the Pirelli tower.

Vibration Mode	Frequency f (Hz)
1	0.25
2	1.08
3	2.60
4	4.71
5	7.06
6	8.79
7	9.56
8	9.91
9	11.38
10	13.36
11	14.64
12	18.30
13	22.14

To qualitatively assess the reconstruction capacity of the AE when test instances are considered, results are reported in Figure 23 in terms of the histories of the 20th floor displacement for a couple of solutions related to each load case. For each case, the reconstructed signal has resulted close to the input one for low excitation frequencies, but the accuracy seems to get reduced by larger values of ϕ . This outcome is again related to what discussed in Section 4.1.2 regarding the mode factors Γ_s , computed similarly to Equation (11) for $s = 1, \dots, 39$, which are larger for the low-order vibration modes.

Regarding load identification, outcomes are reported in Figure 24 and in Table 6. As shown for the two-storey shear building, the reconstruction accuracy of the AE has an impact on the performance of the regressor r . For $\phi > 5$ Hz, the regression outcomes of the load cases 1 and 2 get slightly deteriorated, with an increased scattering of the values of ϕ_r for load case 2 as assessed by the relevant values of the RMSE (passing from 0.144 Hz to 0.417 Hz) and of the Pearson correlation coefficient R^2 (moving down from 0.998 to 0.984). For those two load cases, the load amplitude has been always well predicted, with errors bounded independently of α . Hence, the AE reconstruction capacity is only marginally affected by the load amplitude.

A different type of outcomes has been obtained for the load case 3. Focusing on ϕ_r , the regression task fails if $\phi > 9$ Hz. The Pearson correlation coefficient R^2 falls from almost unitary values for the load case 1 and 2 to 0.808 for the load case 3 when the identification of α is addressed, and from 0.984 for load case 2 to 0.679 when ϕ is targeted. A way to explain this behavior is again linked to the small values of Γ_s associated to higher vibration modes; accordingly, the critical part of the procedure is still assumed to be due to signal reconstruction. Moreover, in comparison to the load cases 1 and 2, the regression performance gets spoiled in the frequency range $5.5 \text{ Hz} < \phi < 9 \text{ Hz}$. The predicted values exhibit a sort of cut-off around 8 Hz, above which it becomes impossible to obtain reasonable predictions. The complexity of the AE loss $c(\mathbf{V}, \mathbf{U})$ hence forces the training algorithm to converge to a local minimum that worsens the regression performance, not only within the previously mentioned range but also for all the values larger than 6 Hz. A reduced performance has been also shown regarding the load amplitude α , even if a cut-off has not been reported. Overall, the regressor tends to systematically underestimate α , except for the instances featuring $\phi < 5$ Hz which are instead caught correctly.

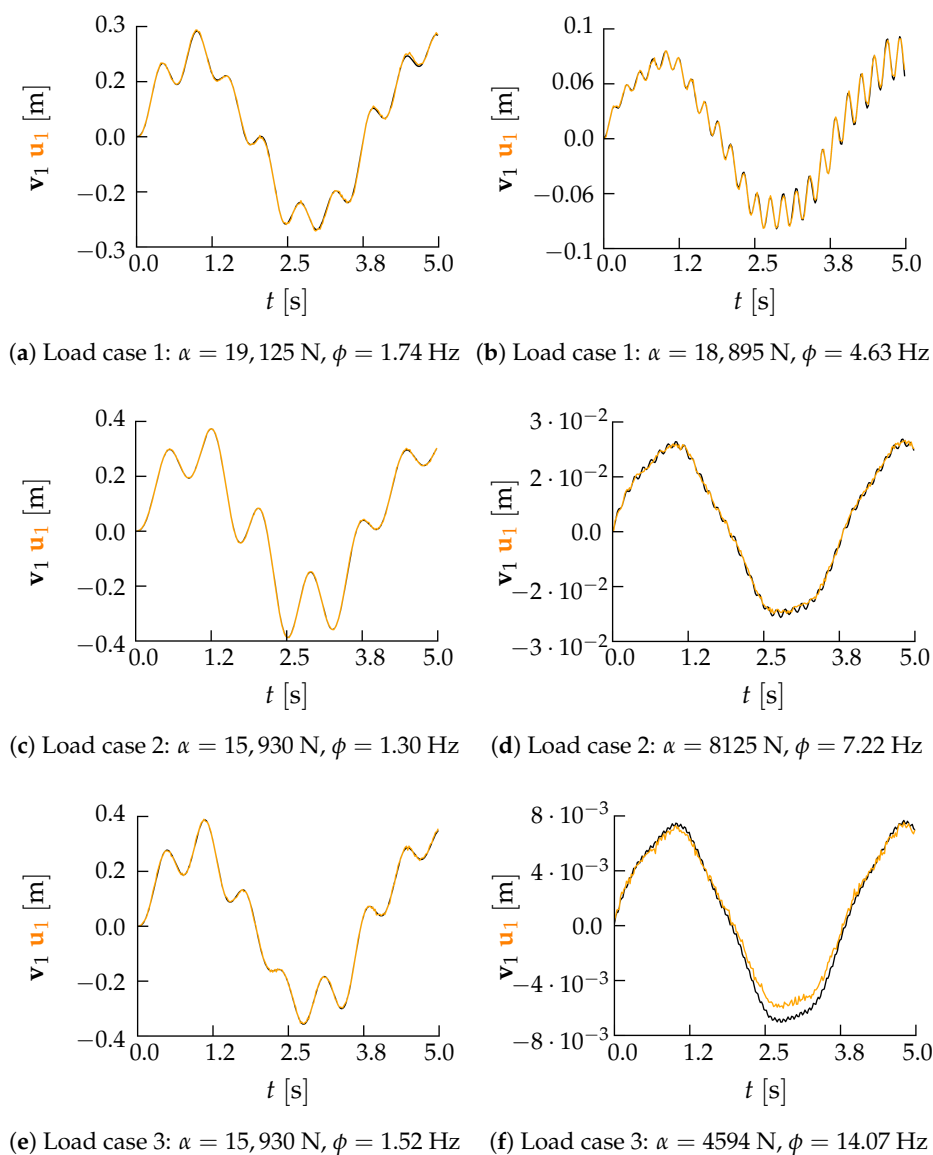


Figure 23. Pirelli Tower. Comparison between input (black lines) and reconstructed (orange lines) time histories of the 20-th floor displacement for six cases belonging to the test set.

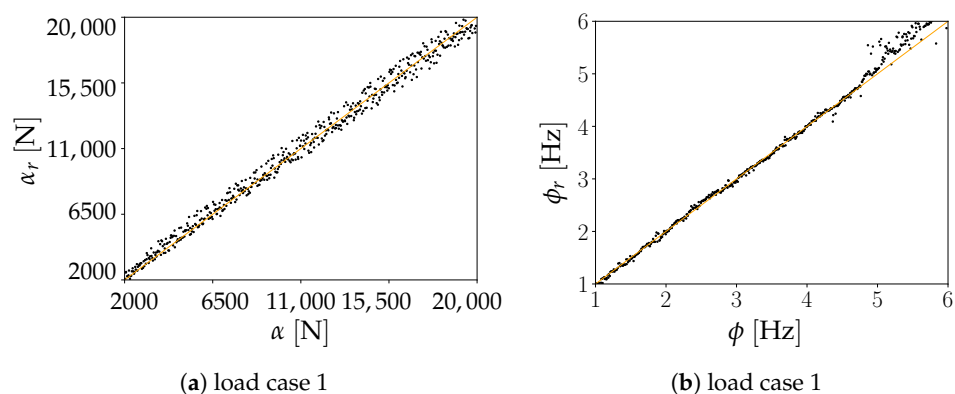


Figure 24. Cont.

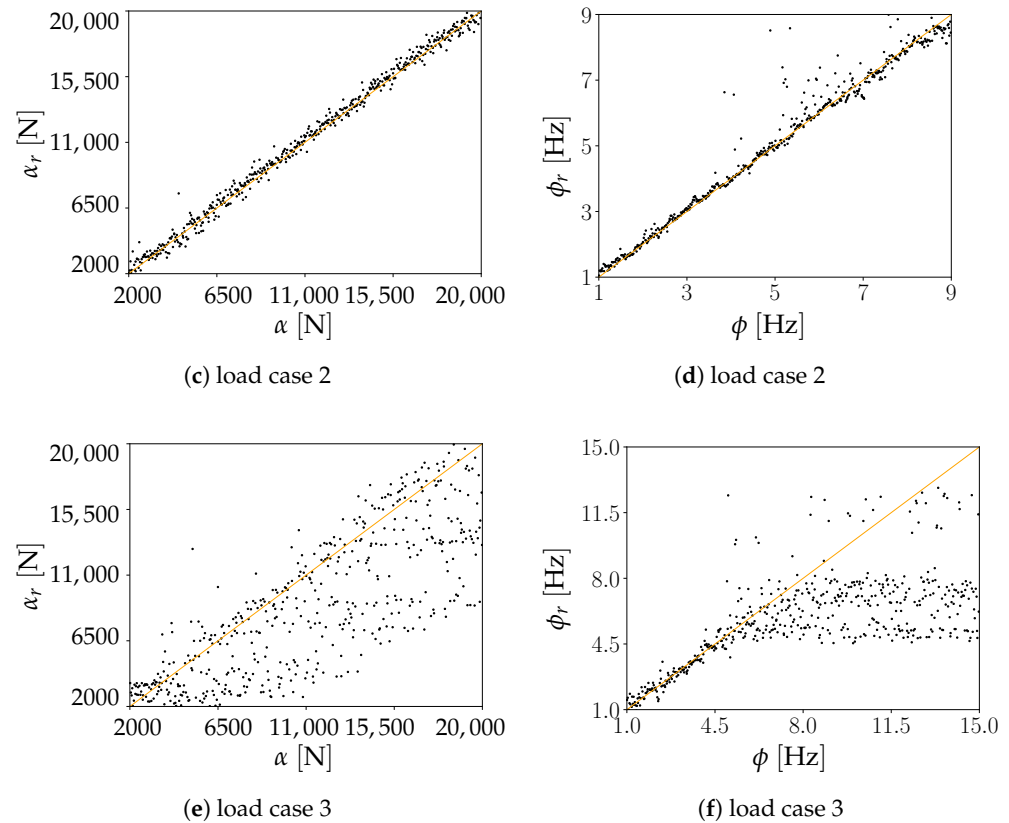


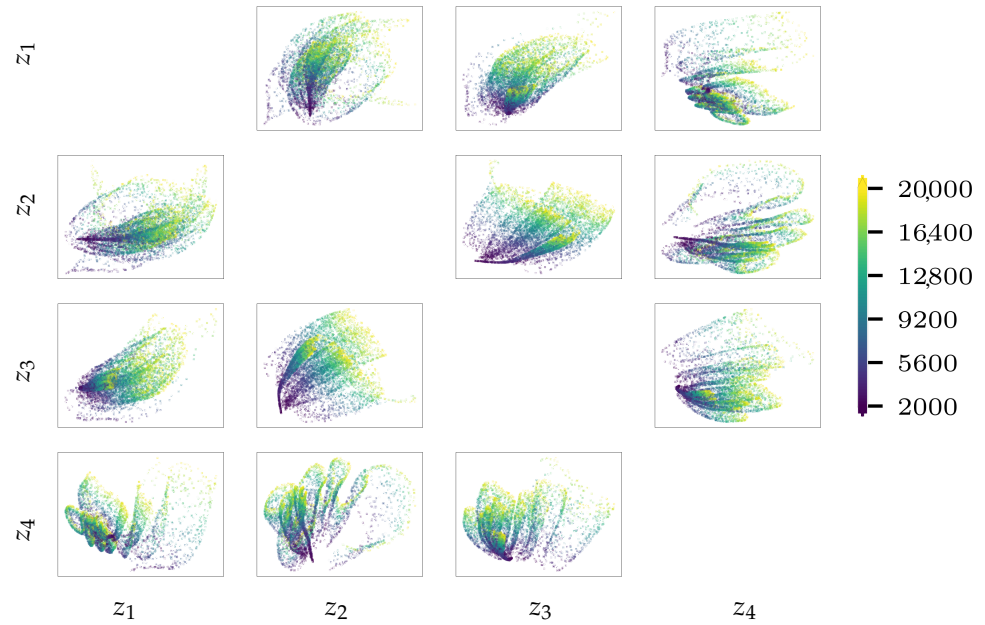
Figure 24. Pirelli Tower. Parity plots showing the regression outcomes for the load amplitude α (left column) and the load frequency ϕ (right column) for the three considered load cases.

Table 6. Pirelli tower. RMSE and correlation coefficient R^2 of the regression outcomes for the load amplitude α and the load frequency ϕ for the three considered load cases.

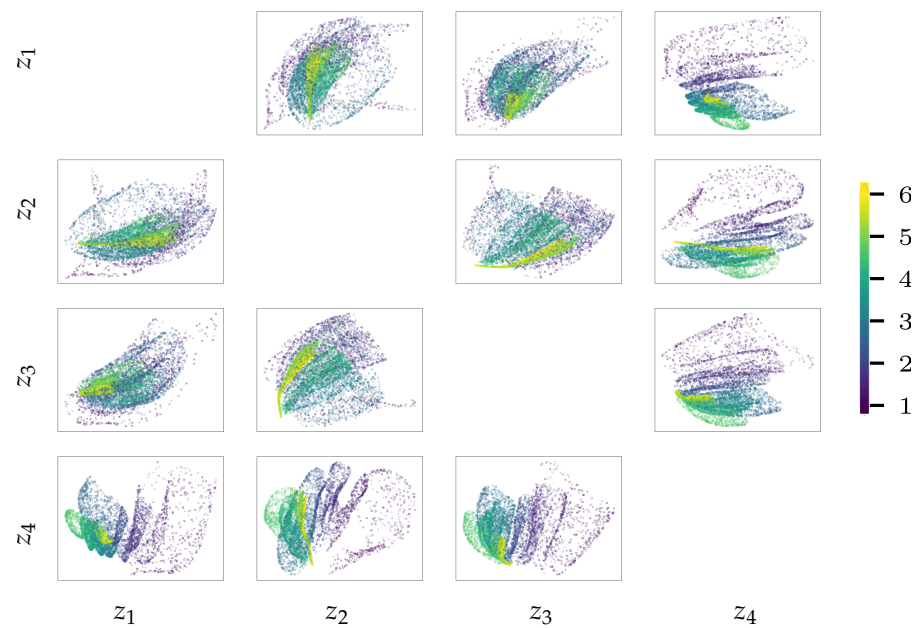
Load Case	α		ϕ	
	RMSE [N]	R^2 [-]	RMSE [Hz]	R^2 [-]
1	469	0.996	0.144	0.998
2	439	0.997	0.417	0.984
3	3852	0.808	3.758	0.679

To address the above discussed issues, we report in Figures 25 and 26 the latent variables \mathbf{z} for the validation sets of load cases 1 and 3. The focus is only on the out-of-diagonal dispersion plots in this matrix-like representation of the results, wherein each chart shows the values taken by pairs of latent variables. A color code is used in all the plots, to visualize how the generative factors α and ϕ affect the encoding distribution in the latent space. These results confirm that the load identification issues are due to the encoding procedure: looking at Figure 26b, it is clearly shown that when ϕ is larger than the cut-off-value discussed in relation to Figure 24f, the encoded values are not spread anymore in the sub-space spanned by the said pairs of latent variables. For a frequency in the range $5 \text{ Hz} < \phi < 8 \text{ Hz}$, the encoded states start to show a tendency towards collapsing into lower-dimensional loci, though a small scattering is still present and the regressor can predict load frequencies in this range. We can thus state that the greater the scattering of the latent variables, the more effective the encoded representation in terms of signal reconstruction and load identification. This is confirmed by looking at the results related to load case 1: the best reconstructed instances are obtained for the most scattered encodings, characterized by small values of ϕ . When the encodings become less scattered, they also become less informative for the signal reconstruction. By looking at Figure 25, it can be

also noticed that the instances featuring both larger values of α (reported in yellow in the plots) and smaller values of ϕ (reported in violet in the plots) are those farthest from the collapsed loci.



(a)



(b)

Figure 25. Pirelli Tower, load case 1. Scattered latent representation z determined by *enc* for the validation set, with coding set by (a) the load amplitude α or by (b) the load frequency ϕ .

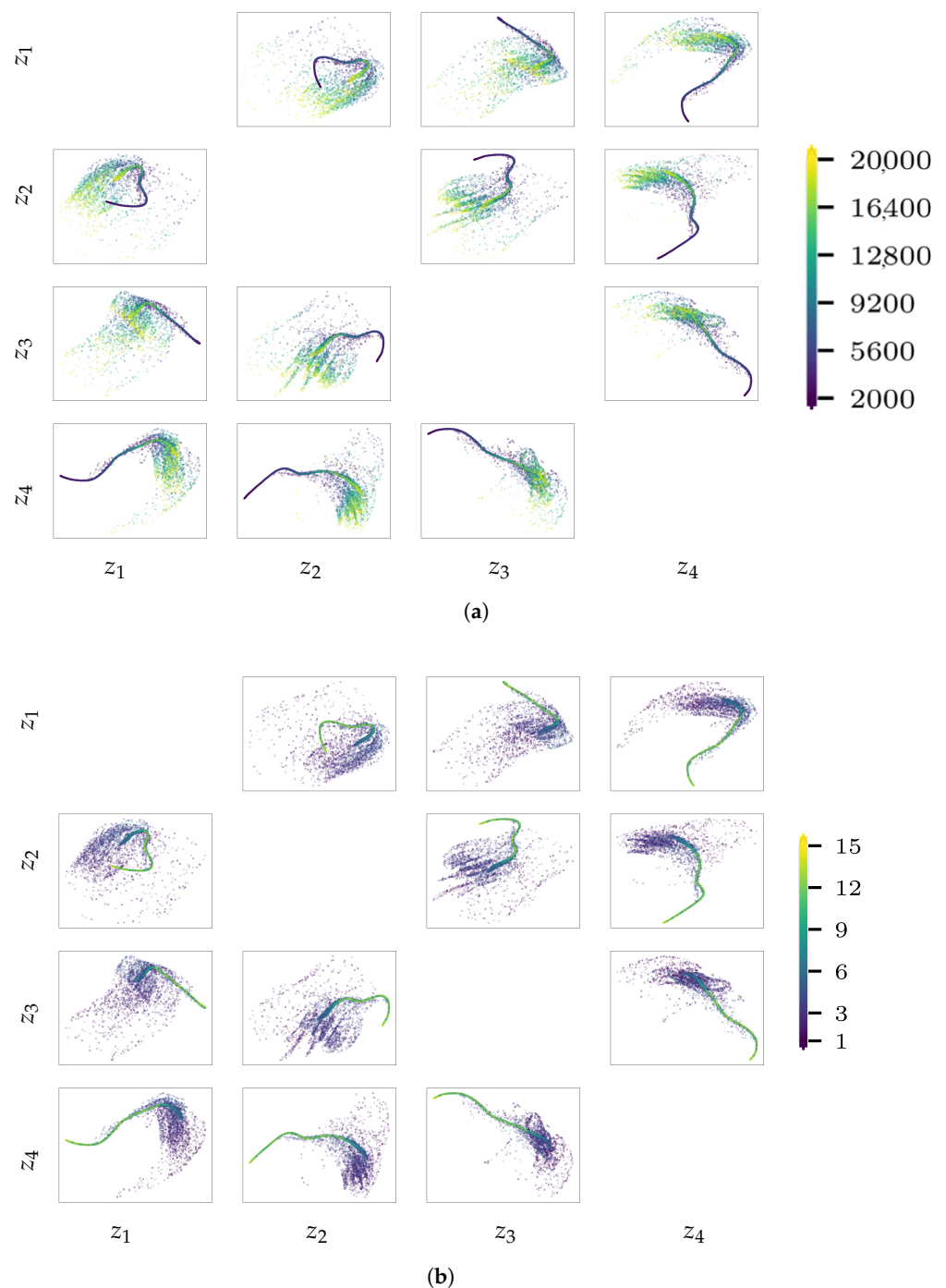


Figure 26. Pirelli Tower, load case 3. Scattered latent representation \mathbf{z} determined by *enc* for the validation set, with coding set by (a) the load amplitude α or by (b) the load frequency ϕ .

5. Conclusions

A time series AE has been designed and adopted to solve a regression task linked to load identification, which plays a crucial role in the assessment of the health of a structure if operational variability is allowed for, or in case of an output-only architecture of the monitoring system. The role of the latent representation provided by the AE has been extensively discussed, allowing for the sources of output variability of the tackled problem and the inverse-forward nature beyond the autoencoding paradigm.

For a two-storey shear building model, the reconstruction capacity of the AE has been quantitatively assessed by using two error norms, and by investigating the effect on it of the structural vibration frequencies and modes. Physically-sound links with the dynamics

of the system have been shown, pinpointing how the excited structural vibrations can play a key role in setting the AE outcomes. The false nearest neighbor heuristics has been also allowed for to automatically set the dimension of the latent representation. The results obtained for the two-storey shear building have been then taken into account to approach load identification for a digital twin of a high-rise building, the Pirelli Tower in Milan. Promising outcomes have been reported, to foresee future applications of the proposed methodology to real-life situations.

The next steps will be to further empower the setting procedure for the AE hyperparameters, by combining the here adopted trial and error procedure with Bayesian methods [48]. Optimization methods for the deployment of the sensors of the monitoring system will be also proposed, in order to maximize the information content of the latent representation and, at the same time, minimize the number of sensors to deploy and avoid issues linked to big data (see, e.g., [52]).

Author Contributions: Conceptualization, L.R., A.M., S.M. and A.C.; methodology, L.R. and S.M.; software, L.R.; validation, L.R.; formal analysis, L.R., A.M. and S.M.; investigation, L.R. and S.M.; writing—original draft preparation, L.R.; writing—review and editing, A.M., S.M. and A.C.; supervision, A.C. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The numerical cases of study have been exhaustively described. The reader may generate the same data used in this paper.

Acknowledgments: The authors are indebted to Matteo Torzoni for the valuable discussions concerning several aspects of this research, and to Gianluca Barbella and Federico Perotti for having provided the numerical model of the Pirelli Tower.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Fink, O.; Wang, Q.; Svensén, M.; Dersin, P.; Lee, W.J.; Ducoffe, M. Potential, challenges and future directions for deep learning in prognostics and health management applications. *Eng. Appl. Artif. Intell.* **2020**, *92*, 103678. [[CrossRef](#)]
2. Doebling, S.W.; Farrar, C.; Prime, M. A Summary Review of Vibration-Based Damage Identification Methods. *Shock Vib. Dig.* **1998**, *30*, 91–105. [[CrossRef](#)]
3. Sohn, H.; Worden, K.; Farrar, C.R. Statistical Damage Classification Under Changing Environmental and Operational Conditions. *J. Intell. Mater. Syst. Struct.* **2002**, *13*, 561–574. [[CrossRef](#)]
4. Azam, S.E.; Mariani, S.; Attari, N.K. Online damage detection via a synergy of proper orthogonal decomposition and recursive Bayesian filters. *Nonlinear Dyn.* **2017**, *89*, 1489–1511. [[CrossRef](#)]
5. Azam, S.E.; Mariani, S. Online damage detection in structural systems via dynamic inverse analysis: A recursive Bayesian approach. *Eng. Struct.* **2018**, *159*, 28–45. [[CrossRef](#)]
6. Brincker, R.; Zhang, L.; Andersen, P. Modal identification of output-only systems using frequency domain decomposition. *Smart Mater. Struct.* **2001**, *10*, 441–445. [[CrossRef](#)]
7. Azam, S.E.; Chatzi, E.; Papadimitriou, C. A dual Kalman filter approach for state estimation via output-only acceleration measurements. *Mech. Syst. Signal Process.* **2015**, *60–61*, 866–886. [[CrossRef](#)]
8. Hansen, P.C.; Sekii, T.; Shibahashi, H. The Modified Truncated SVD Method for Regularization in General Form. *SIAM J. Sci. Stat. Comput.* **1992**, *13*, 1142–1150. [[CrossRef](#)]
9. Yang, C.; Hou, X.; Chang, S. A synchronous placement and size-based multi-objective optimization method for heat dissipation design on antenna module of space solar power satellite. *Sustain. Energy Technol. Assess.* **2021**, *45*, 101183. [[CrossRef](#)]
10. Yang, C. A novel uncertainty-oriented regularization method for load identification. *Mech. Syst. Signal Process.* **2021**, *158*, 107774. [[CrossRef](#)]
11. Seo, J.K.; Kim, K.C.; Jargal, A.; Lee, K.; Harrach, B. A Learning-Based Method for Solving Ill-Posed Nonlinear Inverse Problems: A Simulation Study of Lung EIT. *SIAM J. Imaging Sci.* **2019**, *12*, 1275–1295. [[CrossRef](#)]
12. Farrar, C.; Worden, K. *Structural Health Monitoring A Machine Learning Perspective*; Wiley: Hoboken, NJ, USA, 2013. [[CrossRef](#)]
13. Wang, Y.; Yao, H.; Zhao, S. Auto-encoder based dimensionality reduction. *Neurocomputing* **2016**, *184*, 232–242. [[CrossRef](#)]
14. Liu, J.; Chen, S.; Bergés, M.; Bielak, J.; Garrett, J.H.; Kovačević, J.; Noh, H.Y. Diagnosis algorithms for indirect structural health monitoring of a bridge model via dimensionality reduction. *Mech. Syst. Signal Process.* **2020**, *136*, 106454. [[CrossRef](#)]

15. Wang, R.; Li, L.; Li, J. A Novel Parallel Auto-Encoder Framework for Multi-Scale Data in Civil Structural Health Monitoring. *Algorithms* **2018**, *11*, 112. [[CrossRef](#)]
16. Hu, Y.; Palmé, T.; Fink, O. Fault detection based on signal reconstruction with Auto-Associative Extreme Learning Machines. *Eng. Appl. Artif. Intell.* **2017**, *57*, 105–117. [[CrossRef](#)]
17. Bengio, Y.; Courville, A.; Vincent, P. Representation Learning: A Review and New Perspectives. *IEEE Trans. Pattern Anal. Mach. Intell.* **2013**, *35*, 1798–1828. [[CrossRef](#)] [[PubMed](#)]
18. Zamir, A.R.; Sax, A.; Shen, W.; Guibas, L.J.; Malik, J.; Savarese, S. Taskonomy: Disentangling Task Transfer Learning. In Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–23 June 2018; pp. 3712–3722. [[CrossRef](#)]
19. Shen, C.; Qi, Y.; Wang, J.; Cai, G.; Zhu, Z. An automatic and robust features learning method for rotating machinery fault diagnosis based on contractive autoencoder. *Eng. Appl. Artif. Intell.* **2018**, *76*, 170–184. [[CrossRef](#)]
20. Fu, T. A review on time series data mining. *Eng. Appl. Artif. Intell.* **2011**, *24*, 164–181. [[CrossRef](#)]
21. Gilpin, W. Deep reconstruction of strange attractors from time series. In Proceedings of the 34th Conference on Neural Information Processing Systems (NeurIPS 2020), Virtual Conference, 6–14 December 2020.
22. Haykin, S. *Neural Networks and Learning Machines*; Prentice Hall: Upper Saddle River, NJ, USA, 2009.
23. Kingma, D.; Ba, J. Adam: A Method for Stochastic Optimization. In Proceedings of the International Conference on Learning Representations (ICLR), San Diego, CA, USA, 7–9 May 2015; pp. 1–13.
24. Rosafalco, L.; Manzoni, A.; Mariani, S.; Corigliano, A. Fully convolutional networks for structural health monitoring through multivariate time series classification. *Adv. Model. Simul. Eng. Sci.* **2020**, *7*, 38. [[CrossRef](#)]
25. Rosafalco, L.; Torzoni, M.; Manzoni, A.; Mariani, S.; Corigliano, A. Online structural health monitoring by model order reduction and deep learning algorithms. *arXiv* **2021**, arXiv:2103.14328.
26. Locatello, F.; Tschannen, M.; Bauer, S.; Rätsch, G.; Schölkopf, B.; Bachem, O. Disentangling Factors of Variations Using Few Labels. In Proceedings of the 8th International Conference on Learning Representations (ICLR), Vancouver, BC, Canada, 26–30 April 2020.
27. Szegedy, C.; Liu, W.; Jia, Y.; Sermanet, P.; Reed, S.; Anguelov, D.; Erhan, D.; Vanhoucke, V.; Rabinovich, A. Going deeper with convolutions. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition CVPR, Las Vegas, NV, USA, 26 June–1 July 2015; pp. 1–9. [[CrossRef](#)]
28. Klambauer, G.; Unterthiner, T.; Mayr, A.; Hochreiter, S. Self-Normalizing Neural Networks. In *Advances in Neural Information Processing Systems 30*; Guyon, I., Luxburg, U.V., Bengio, S., Wallach, H., Fergus, R., Vishwanathan, S., Garnett, R., Eds.; Curran Associates, Inc.: Red Hook, NY, USA, 2017; pp. 971–980.
29. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep Residual Learning for Image Recognition. In Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition CVPR, Las Vegas, NV, USA, 27–30 June 2016; pp. 770–778. [[CrossRef](#)]
30. Hu, J.; Shen, L.; Sun, G. Squeeze-and-Excitation Networks. In Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–23 June 2018; pp. 7132–7141. [[CrossRef](#)]
31. Arora, S.; Bhaskara, A.; Ge, R.; Ma, T. Provable Bounds for Learning Some Deep Representations. In Proceedings of the 31st International Conference on Machine Learning, Beijing, China, 22–24 June 2014; PMLR: Beijing, China, 2014; Volume 32, pp. 584–592.
32. Abadi, M.; Agarwal, A.; Barham, P.; Brevdo, E.; Chen, Z.; Citro, C.; Corrado, G.S.; Davis, A.; Dean, J.; Devin, M.; et al. TensorFlow: Large-Scale Machine Learning on Heterogeneous Systems. 2015. Available online: tensorflow.org (accessed on 28 March 2020).
33. Chollet, F. Keras. 2015. Available online: <https://keras.io> (accessed on 4 April 2020).
34. Dutilleul, P. *An Implementation of the “Algorithme à Trous” to Compute the Wavelet Transform*; Wavelets; Springer: Berlin/Heidelberg, Germany, 1990; pp. 298–304. [[CrossRef](#)]
35. van den Oord, A.; Dieleman, S.; Zen, H.; Simonyan, K.; Vinyals, O.; Graves, A.; Kalchbrenner, N.; Senior, A.; Kavukcuoglu, K. WaveNet: A Generative Model for Raw Audio. *arXiv* **2016**, arXiv:1609.03499.
36. Sudret, B.; Defaux, G.; Pendola, M. Stochastic evaluation of the damage length in RC beams submitted to corrosion of reinforcing steel. *Civ. Eng. Environ. Syst.* **2007**, *24*, 165–178. [[CrossRef](#)]
37. Ioffe, S.; Szegedy, C. Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift. In Proceedings of the 32nd International Conference on Machine Learning, ICML, Lille, France, 6–11 July 2015; Volume 37, pp. 448–456.
38. He, K.; Sun, J. Convolutional neural networks at constrained time cost. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition CVPR, Las Vegas, NV, USA, 26 June–1 July 2015; pp. 5353–5360. [[CrossRef](#)]
39. Kennel, M.B.; Brown, R.; Abarbanel, H.D.I. Determining embedding dimension for phase-space reconstruction using a geometrical construction. *Phys. Rev. A* **1992**, *45*, 3403–3411. [[CrossRef](#)] [[PubMed](#)]
40. Barbella, G.; Perotti, F.; Simoncini, V. Block Krylov subspace methods for the computation of structural response to turbulent wind. *Comput. Methods Appl. Mech. Eng.* **2011**, *200*, 2067–2082. [[CrossRef](#)]
41. Eftekhar Azam, S.; Mariani, S. Investigation of computational and accuracy issues in POD-based reduced order modeling of dynamic structural systems. *Eng. Struct.* **2013**, *54*, 150–167. [[CrossRef](#)]
42. Capellari, G.; Chatzi, E.; Mariani, S. Structural Health Monitoring Sensor Network Optimization through Bayesian Experimental Design. *ASCE-ASME J. Risk Uncertain. Eng. Syst. Part A Civ. Eng.* **2018**, *4*, 04018016. [[CrossRef](#)]

43. European Committee for Standardization. *Eurocode 8: Design of Structures for Earthquake Resistance—Part 1: General Rules, Seismic Actions and Rules for Buildings*; CEN: Bruxelles, Belgium, 2003; pp. 66–74.
44. Corigliano, A.; Mariani, S. Parameter identification in explicit structural dynamics: Performance of the extended Kalman filter. *Comput. Methods Appl. Mech. Eng.* **2004**, *193*, 3807–3835. [[CrossRef](#)]
45. Pandey, A.; Biswas, M. Damage Detection in Structures Using Changes in Flexibility. *J. Sound Vib.* **1994**, *169*, 3–17. [[CrossRef](#)]
46. Shahriari, B.; Swersky, K.; Wang, Z.; Adams, R.P.; de Freitas, N. Taking the Human Out of the Loop: A Review of Bayesian Optimization. *Proc. IEEE* **2016**, *104*, 148–175. [[CrossRef](#)]
47. Yang, C.; Liang, K.; Zhang, X. Strategy for sensor number determination and placement optimization with incomplete information based on interval possibility model and clustering avoidance distribution index. *Comput. Methods Appl. Mech. Eng.* **2020**, *366*, 113042. [[CrossRef](#)]
48. Falkner, S.; Klein, A.; Hutter, F. BOHB: Robust and Efficient Hyperparameter Optimization at Scale. In Proceedings of the 35th International Conference on Machine Learning, PMLR, Stockholm, Sweden, 10–15 July 2018; Volume 80, pp. 1437–1446.
49. Clough, R.W.; Penzien, J.; Griffin, D. *Dynamics of Structures*; American Society of Mechanical Engineers Digital Collection. 1977. Available online: <https://asmedigitalcollection.asme.org/> (accessed on 28 March 2020).
50. Severini, T.A. *Elements of Distribution Theory*; Cambridge Series in Statistical and Probabilistic Mathematics; Cambridge University Press: Cambridge, UK, 2005. [[CrossRef](#)]
51. Hazra, B.; Sadhu, A.; Roffel, A.J.; Narasimhan, S. Hybrid Time-Frequency Blind Source Separation Towards Ambient System Identification of Structures. *Comput. Aided Civ. Infrastruct. Eng.* **2012**, *27*, 314–332. [[CrossRef](#)]
52. Entezami, A.; Sarmadi, H.; Behkamal, B.; Mariani, S. Big Data Analytics and Structural Health Monitoring: A Statistical Pattern Recognition-Based Approach. *Sensors* **2020**, *20*, 2328. [[CrossRef](#)] [[PubMed](#)]