

Temperature Effects Removal from Non-Stationary Bridge–Vehicle Interaction Signals for ML Damage Detection

*Original*

Temperature Effects Removal from Non-Stationary Bridge–Vehicle Interaction Signals for ML Damage Detection / Niyozov, Sardorbek; Domaneschi, Marco; Casas, Joan R.; Delgadillo, Rick M.. - In: SENSORS. - ISSN 1424-8220. - 23:11(2023), pp. 1-26. [10.3390/s23115187]

*Availability:*

This version is available at: 11583/2979115 since: 2023-06-05T16:03:25Z

*Publisher:*

MDPI

*Published*

DOI:10.3390/s23115187

*Terms of use:*

openAccess

This article is made available under terms and conditions as specified in the corresponding bibliographic description in the repository

*Publisher copyright*

(Article begins on next page)

## Article

# Temperature Effects Removal from Non-Stationary Bridge–Vehicle Interaction Signals for ML Damage Detection

Sardorbek Niyozov <sup>1,\*</sup>, Marco Domaneschi <sup>1,\*</sup> , Joan R. Casas <sup>2</sup>  and Rick M. Delgadillo <sup>3</sup> 

<sup>1</sup> Department of Structural, Geotechnical and Building Engineering, Politecnico di Torino, 10129 Turin, Italy; sardorbek.niyozov@studenti.polito.it

<sup>2</sup> Department of Civil and Environmental Engineering, Universitat Politècnica de Catalunya, 08034 Barcelona, Spain; joan.ramon.casas@upc.edu

<sup>3</sup> Department of Civil Engineering, Universidad de Ingeniería y Tecnología—UTEC, Jr. Medrano Silva 165, Barranco, Lima 15063, Peru; rdelgadillo@utec.edu.pe

\* Correspondence: marco.domaneschi@polito.it

**Abstract:** Bridges are vital components of transport infrastructures, and therefore, it is of utmost importance that they operate safely and reliably. This paper proposes and tests a methodology for detecting and localizing damage in bridges under both traffic and environmental variability considering non-stationary vehicle-bridge interaction. In detail, the current study presents an approach to temperature removal in the case of forced vibrations in the bridge using principal component analysis, with detection and localization of damage using an unsupervised machine learning algorithm. Due to the difficulty in obtaining real data on undamaged and later damaged bridges that are simultaneously influenced by traffic and temperature changes, the proposed method is validated using a numerical bridge benchmark. The vertical acceleration response is derived from a time-history analysis with a moving load under different ambient temperatures. The results show how machine learning algorithms applied to bridge damage detection appear to be a promising technique to efficiently solve the problem's complexity when both operational and environmental variability are included in the recorded data. However, the example application still shows some limitations, such as the use of a numerical bridge and not a real bridge due to the lack of vibration data under health and damage conditions, and with varying temperatures; the simple modeling of the vehicle as a moving load; and the crossing of only one vehicle present in the bridge. This will be considered in future studies.

**Keywords:** bridge damage detection; non-stationary; PCA; K-means; operational-environmental variability



**Citation:** Niyozov, S.; Domaneschi, M.; Casas, J.R.; Delgadillo, R.M. Temperature Effects Removal from Non-Stationary Bridge–Vehicle Interaction Signals for ML Damage Detection. *Sensors* **2023**, *23*, 5187. <https://doi.org/10.3390/s23115187>

Academic Editors: Filippo Ubertini and Luca De Marchi

Received: 17 April 2023

Revised: 10 May 2023

Accepted: 25 May 2023

Published: 30 May 2023



**Copyright:** © 2023 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

Aging, material degradation, deterioration, fatigue, and corrosion are some of the reasons structures lose their performance. They can sometimes cause structures to fail unexpectedly. The research in this area [1] has increased as major disasters have occurred worldwide [2], resulting in the death of a considerable number of people and large economic consequences. In this regard, structural health monitoring (SHM) systems are becoming an important tool to prevent structural failures with direct and indirect losses.

A useful classification of SHM was proposed by Rytter [3], who categorized four levels of damage assessment depending on the characteristics of damage that a particular SHM system can achieve.

- Level 1: Indication of damage existence (Detection);
- Level 2: Information of damage position (Localization);
- Level 3: Damage intensity (Assessment);
- Level 4: Prognosis (Remaining life prediction).

With the advancement of machine learning (ML) algorithms, a new level that corresponds to the type of damage can be introduced above [4]. This new level stands between Levels 2 and 3.

Vibration data has typically been used to detect bridge deterioration based on changes in modal parameters brought on by the presence of damage. It is rather easy to obtain dynamic parameters as modal ones and damping properties using traditional time-frequency signal processing techniques such as fast Fourier transform (FFT) and short-time Fourier transform (STFT) [5,6]. Such signal processing techniques, however, are predicated on the idea that the signal is stationary and linear, which may be argued in general to be valid for the duration of the short time window, but is not valid over the entire timespan of the observation. In addition, FFT presents a few intrinsic limitations that can hinder the effectiveness of the system identification and the following damage detection [7]. For instance:

- (i) The structural health state may be lost during the data reduction procedure, which is specifically retrieved by FFT [8].
- (ii) FFT cannot identify the time dependence of the dynamic parameters, and when using signals from naturally excited structures, it is unable to capture the evolutionary traits that can be observed in these signals [9].

The higher frequencies are poorly excited and closely spaced, and they typically represent damages that are local events [10,11]. As a result of the linearity and stationarity assumptions, damage detection methods using FFT show relevant limitations. The signals derived from vehicles crossing a bridge, however, do not exhibit these properties because the frequency spectrum of the response shifts along the crossing due to the vehicle-bridge interaction and because the signal's amplitude is modified by the vehicle's position on the bridge in relation to the location where the response is recorded.

It becomes clear that the recorded vibration's frequency and amplitude are time dependent. As a result, techniques for modal parameter extraction (such as FFT) that rely on the linearity and stationarity of the process are inapplicable for obtaining trustworthy and precise damage indicators.

Wavelet-based methods, addressed in the literature and applied to bridges, allow for valuable insights with respect to modal identification. However, the developed techniques are aimed at the identification of general time-varying systems, while it is known that other critical conditions such as the vehicle-bridge interaction contain closely spaced spectral components due to the vehicle-bridge dynamic coupling, which instead requires alternative approaches [12].

Two major limitations of SHM during continuous monitoring are also present. First, by applying increasingly sophisticated algorithms to monitoring data, the number of features that are sensitive to degradation are estimated rather than being measured directly. Natural frequencies, for instance, are determined through measurements of vibrational reactions, such as accelerations, which are subject to estimation mistakes. In addition to changes brought on by structural degradation, the features discovered through monitoring are also sensitive to operational (traffic load) and environmental (temperature, humidity, and wind speed) conditions. As a result, in order to properly apply SHM, it is necessary to consider the accuracy of the predicted features as well as environmental and operational variability.

In conclusion, since environmental and traffic effects always have an impact on dynamic response in real bridges in use (causing non-linearity and non-stationarity in the recorded signals), there are two main challenges when using damage detection techniques in active bridges: how to eliminate ambient effects and extract precise damage features that are practical for damage detection, location, and quantification from short non-linear and non-stationary vibration produced when vehicles cross the bridge. This paper will demonstrate how relevant damage features derived from vibration records and machine learning techniques can deliver a relevant answer to the query.

## 2. Damage Detection Method

It is generally recognized that a number of external elements, including the environment and operational circumstances (traffic), do have effects on the efficiency of SHM systems when adopted for bridges. Several SHM works from the literature are focused on the use of machine and statistical learning techniques to address this issue. For instance, when the temperature is constant, Santos et al. [13] suggested an online unsupervised detection technique for early damage identification. Additionally, Soo et al. [14] proposed a procedure for distinguishing environmental effects from damage in close to real-time using principal component analysis, although without taking the operational loading into account. They do not, however, offer a complete solution for bridges that are subject to both traffic loads and temperature fluctuations. Tenelema et al. [15] showed that principal component analysis (PCA) is efficient for differentiating environmental effects from selected data when traffic data are used. They used instantaneous phase differences as damage-sensitive features on a steel bridge and demonstrated that the first principal component (PC) is purely related to temperature effects. Even though the temperature effect was clearly shown, the bridge FEM model did not accurately represent a real bridge structure. Moreover, the damage detection was based on the variance changes in the second PC of PCA, which cannot guarantee certainty in real cases because of excessive noise.

Delgadillo and Casas [16] demonstrated on a real steel truss bridge that instantaneous frequencies (IF) are reliable damage-sensitive features by applying the Hilbert-Huang transform and mode decomposition, performing unsupervised cluster-based machine learning (K-means) on symbolic data. As a result, K-means proved to be efficient to detect and localize the damage, but temperature effects were not considered. Therefore, the K-means algorithm is a promising tool to be tested in further developments that include temperature effect removal.

More recent contributions from the literature can also be reported. Wang et al. [17] proposed a hybrid method for damage detection under environmental changes. Specifically, they used a PCA and Gaussian mixture method to cluster the data into damaged and undamaged states.

Huang et al. [18] presented a “train-based performance warning method” for bridge main girders using long-term data, considering temperature, wind, and traffic load. However, their objective is more focused on removing the traffic effects from the analyzed vibration record rather than on using the capability of vibration due to traffic for damage detection.

Meixedo et al. [19,20] considered together different environmental and operational effects on bridges, i.e., temperature, wind, and train load. In particular, they used the portion of the vibration response when the load is on the bridge for damage detection purposes.

Zhang et al. [21] proposed a steady-state data baseline model for bridges by eliminating the non-stationary effects of the temperature by employing PCA directional projection. Strain data at the different points of the bridge were used.

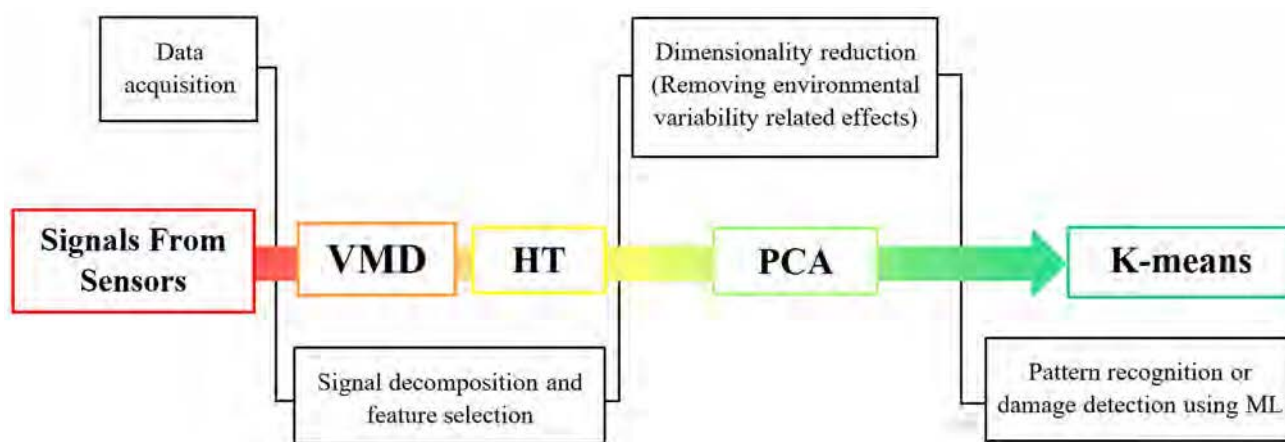
Wang et al. (2022) [22] discussed the displacement model error-based performance warning to detect structural anomalies using PCA to eliminate traffic loads, assuming the first PC only to be related to the traffic loads.

This paper’s goal is to provide an SHM methodology for a bridge under operational (traffic loading) and environmental (temperature) effects. Therefore, in this study, both conditions are considered together, while the contributions by Tenelema et al. and Delgadillo and Casas [15,16] do not deal completely with both effects and do not provide an appropriate threshold to define the presence of damage. With respect to Wang et al. [17], who argued that the main components carry information about the state of the structure, this paper reconstructs the data by removing first the PC, related to temperature effects, and then identifies damage by using an advanced machine learning approach. Conversely, Huang et al. [18] adopted PCA to remove wind and traffic effects, while temperature was eliminated using a canonically correlated method. Zhang et al. [21] applied a machine learning approach to the second PC for damage detection without considering other PCs and environmental effects, while the present work employs K-means for detecting and

localizing damage. Moreover, considering the contribution by Wang et al. (2022) [22], where long-span bridges and static data only are considered, this paper focuses on standard and frequent bridges with short and medium spans and uses dynamic response.

Finally, with respect to Meixedo et al. [19,20], where the vibration due to trains was used to obtain autoregressive (AR) and autoregressive exogenous input (ARX) models as feature extractors and used for damage detection only by fusing the data from different sensors, the main novelty of the present work is the vibration decomposition to deal with the intrinsic non-stationarity of the dynamic response induced by highway traffic to define instantaneous damage indicators in the frequency domain, with the main advantage that they may be used for both detection and localization of damage based on the results of different sensors.

This study treats recorded non-stationary signals of the vehicle-bridge interaction in a comprehensive way, proposing PCA analysis for the removal of temperature effects and instantaneous frequencies as damage features. Furthermore, it includes the use of a damage index, based on a machine learning approach, as a clear threshold for differentiating damaged and undamaged scenarios. Figure 1 shows a flow chart of the proposed methodology, which is summarized as follows:



**Figure 1.** Flow chart of the proposed SHM methodology.

1. Acceleration records under traffic loads and at different ambient temperatures are collected from sensors.
2. Variational mode decomposition (VMD) and Hilbert transform (HT) are applied to recorded data in order to choose a damage-sensitive feature for the following step. This step aims to deconstruct the non-stationary signal into intrinsic mode functions (IMFs). In this study, instantaneous frequency is considered as the relevant damage feature.
3. Principal component analysis is performed to remove the environmental effects (temperature).
4. Symbolic data analysis is performed, and a cluster-based moving window K-means algorithm is applied to selected scenarios for damage detection and localization.
5. In the following subsections, the above-mentioned numerical tools are described in detail.

### 2.1. Variational Mode Decomposition (VMD)

Since Dragomiretskiy and Zosso's 2014 [23] proposal, variational mode decomposition (VMD) has been extensively utilized in a variety of applications. By concurrently extracting modes, they devised a variational mode decomposition model that is completely non-recursive. It was developed to overcome empirical mode decomposition's (EMD) limitations of sensitivity to noise and sampling.

VMD is frequently used in structural engineering for modal identification by means of the recorded dynamic response. For instance, Bagheri et al. [24] used a variety of case

studies (numeric, experimental, and on site) to show the effectiveness of the VMD method. A field study of a pedestrian bridge looked at the structural response to ambient vibrations input, while a laboratory investigation focused on the vibratory response of a multi-story structure. In addition, the shear frame's modal characteristics were calculated using an analytical method and then compared to the modal parameters computed by laboratory tests. As a result, compared to conventional signal decomposition techniques such as EMD, the VMD-based system identification was shown to be robust against noise signals and sampling frequency.

The limitation of this method is that the number of modes  $K$  must be set in advance [25]. If the selected mode number is not accurate, the VMD will cause the loss of important modes. In order to determine the number of intrinsic mode functions (IMFs) different methods have been proposed such as the correlation coefficient method [26], the normalized mutual information method [27], and, the most common, the center frequency observation method [28].

VMD's objective is to discretely divide an input signal with real values,  $f$ , into quasi-orthogonal band sub-signals with band limits,  $u_k$  (modes). Each mode is condensed around the central pulsation  $\omega_k$ , and the bandwidth is calculated using the shifted signal's  $H^1$  Gaussian smoothness [28]. The constrained variational problem form of the VMD is as follows:

$$\min_{\{u_k\}, \{\omega_k\}} \left\{ \sum_{k=1}^K \|\partial_t \left[ \left( \delta(t) + \frac{j}{\pi t} \right) * u_k(t) \right] e^{-j\omega_k t} \|_2^2 \right\} \quad (1)$$

$$\text{s. t. } \sum_{k=1}^K u_k = f \quad (2)$$

where  $u_k$  and  $\omega_k$  are the  $k$ th intrinsic mode function, and it is the center frequency, respectively;  $\partial_t$  is the time derivative;  $\delta(t)$  is Dirac's delta function; and  $j$  is an imaginary number.

$$\begin{aligned} \mathcal{L}(\{u_k\}, \{\omega_k\}, \lambda) = & \alpha \sum_k \|\partial_t \left[ \left( \delta(t) + \frac{j}{\pi t} \right) * u_k(t) \right] e^{-j\omega_k t} \|_2^2 \\ & + \|f(t) - \sum_k u_k(t)\|_2^2 + \langle \lambda(t), f(t) - \sum_k u_k(t) \rangle \end{aligned} \quad (3)$$

where the data fidelity constraint's balancing parameter,  $\alpha$ , is also known as the penalty factor.

## 2.2. Hilbert Transform (HT)

The HT is a linear operator that transforms a real signal  $x(t)$  into another real signal, indicated by  $H[x(t)]$ . It is defined as the convolution of  $x(t)$  with the function  $1/(\pi t)$ :

$$H[x_k(t)] = \frac{1}{\pi} P.V \int_{-\infty}^{+\infty} \frac{x_k(\tau)}{t - \tau} d\tau \quad (4)$$

$P.V$  is Cauchy's principal value of the integral, and  $x_k(t)$  corresponds to the  $k$ th IMF component obtained from the mode decomposition technique. For simplicity purposes, the notation  $x_k(t)$  is used instead of  $IMF_k(t)$  in all equations below. The HT allows defining the complex analytic signal  $z_k(t)$ , from which instantaneous amplitude and phase can be computed. An analytic signal represents rotation in the complex plane with the rotation radius  $a_k(t)$  and the instantaneous function  $\theta_k(t)$  [25]. This implies that the analytic signal becomes:

$$z_k(t) = x_k(t) + i H[x_k(t)] = a_k(t) e^{i\theta_k(t)} \quad (5)$$

where  $H[x_k(t)]$  represents the Hilbert transform of  $IMF_k(t)$ ,  $a_k(t)$  is the instantaneous amplitude, and  $\theta_k(t)$  is the instantaneous phase function.

$$a_k(t) = \sqrt{\{x_k(t)\}^2 + \{H[x_k(t)]\}^2} \quad (6)$$

$$\theta_k(t) = \arctan\left(\frac{H[x_k(t)]}{x_k(t)}\right) \quad (7)$$

The instantaneous amplitude  $a_k(t)$  describes the envelope of the denoised  $IMF_k(t)$ , while  $\theta_k(t)$  describes the number of rotations.

The concept of the frequency and phase carry significant importance when applying IMFs [29]. If the IMFs can be considered local, then the instantaneous angular frequency  $\omega_k(t)$  can be defined as:

$$\omega_k(t) = \frac{d\theta_k(t)}{dt} = 2\pi f_k(t) \quad (8)$$

Cohen [30] stated that each IMF is a mono-component signal with a monotonically increasing phase and a positive instantaneous frequency. Therefore, each selected IMF can be defined as:

$$x_k(t) = \text{Re}(z_k(t)) = \text{Re}(a_k(t)e^{i\theta_k(t)}) = a_k(t) \cos[\theta_k(t)] \quad (9)$$

The sum of the instantaneous phases for the chosen IMFs can be used to determine the overall instantaneous phase:

$$\theta(t) = \sum_k \theta_k(t) = \sum_k \arctan\left(\frac{H[x_k(t)]}{x_k(t)}\right) \quad (10)$$

where  $x_k(t)$  are physically meaningful (i.e., closer to the natural frequencies) IMFs selected for spectral analysis, and  $\theta(t)$  is the total number of rotations in the complex plane, in radians (rad), of a significant part of the original measured signal with  $xk(t)$ .

### 2.3. Clustering Algorithm

The study of grouping or clustering things based on measurable or perceived intrinsic features or similarities is known as cluster analysis. Data clustering (unsupervised learning) differs from classification or discriminant analysis since it lacks category information (supervised learning). Clustering techniques do not require defining reference/training data, in contrast to supervised algorithms. By looking for the smallest and most distinct group of clusters, they may “understand” the structure of a data set [31].

The primary objective of clustering algorithms is to identify a structure or a pattern in the data. It is an effort to increase the dissimilarity between objects assigned to various clusters while simultaneously minimizing the distinction between data objects mathematically associated with the same cluster [31]. There is no univocal belief on the whole definition of clustering, although a traditional one is provided below [32]:

1. Instances, in the same cluster, must be similar as much as possible.
2. Instances, in the different clusters, must be different as much as possible.
3. Measurement for similarity and dissimilarity must be clear and have practical meaning.

K-means is taken into account in this work. It is a partition-based clustering approach. This class of algorithms' fundamental principle is to treat each data point's center as the center of its associated cluster [32].

Let  $P_K = \{C_K, \dots, C_K\}$  be partitioned into  $K$  clusters. Then, the overall within-cluster  $W(P_K)$  dissimilarity can be defined as [31]

$$W(P_K) = \frac{1}{2} \sum_{k=1}^K \sum_{c(i)=k} \sum_{c(j)=k} d_{ij} \quad (11)$$

where  $c(i)$  is a many-to-one allocation function that assigns object  $i$  to cluster  $k$ , with respect to a dissimilarity measure,  $d_{ij}$ , defined between each pair of data objects,  $i$  and  $j$ .

The overall dissimilarity of a data set,  $OD$ , is given below.

$$OD = \frac{1}{2} \sum_{i=1}^N \sum_{j=1}^N d_{ij} \quad (12)$$

where  $N$  is the total number of objects. The between-cluster dissimilarity is obtained by subtracting the previous two equations,  $B(P_K) = OD - W(P_K)$ .

The goal of the K-means is to minimize the overall within-cluster dissimilarity,  $W(P_K)$ , of a given partition,  $P_K$ , by the iterative optimization scheme. The K-means requires that the number of  $K < N$  clusters be initially defined [31] with a randomly defined set of  $K$  cluster prototypes from the same type of data. This step is called initialization. Following the initialization, each iteration starts by assigning objects to clusters by allocating rule,  $c(i)$ . The second step of the K-means algorithms is to find the best prototypes that represent clusters defined before. It is called the representation step. K-means represents the clusters by finding their centroids. The process of allocation and representation is continued until an objective function that is dependent on the compactness and separation of the clusters reaches its global minimum value [31].

The K-means uses squared within-cluster dissimilarity measured across the  $K$  clusters as an objective function [32]. They are usually based on distance metrics. Euclidean distance is the most widely utilized one (square root of the sum-of-squares). However, it causes significant false detection occurrences and increases computing complexity when SHM is applied. The alternate Gowda-Diday distance measure is employed to go through these issues [31].

#### 2.4. Damage Identification

The last step in the process is the definition of a quantitative value (a threshold value or detection index) that clearly defines the limit between the undamaged and damaged situations and minimizes the number of false positives.

The K-means algorithm demands the initial definition of the number of  $K \leq N$  of clusters, as well as a randomly defined set of cluster prototypes, which are objects of the same type as those being clustered. The process of choosing which set of prototypes best depicts the clusters specified during the allocation phase is known as the representation step of each K-means iteration. There is no way to predict in advance if this number will match the number of distinct structural conditions identified on site. Clear conclusions cannot be formed without identifying which of the partitions best capture the data's structure. The global silhouette index ( $SIL$ ) is employed in this study to obtain the number of clusters since it performed better in prior investigations when its formulation was thoroughly discussed [13,33]. The  $i$ th observation is given a set number of clusters  $K$ , with the following value, to create the silhouette statistic [16]:

$$s(x_i) = \frac{b(x_i) - a(x_i)}{\max\{a(x_i), b(x_i)\}} \in [-1, 1] \quad (13)$$

where  $a(x_i)$  is the average distance between the  $i$ th object of cluster  $C$  and the remaining  $j$  objects, and  $b(x_i)$  the distance to the center of the closest neighboring cluster. Below are shown, accordingly, the average silhouette widths for all samples and the silhouette index of clusters, where  $(1 \leq M_k \leq N)$ ,  $N$  is the set of objects,  $K$  is the partition for clustering, and the silhouette coefficient ranges from 0 to 1.

$$s(x) = \frac{1}{M_k} \sum_{i=1}^{M_k} s(x_i) \quad (14)$$

$$SIL = \frac{1}{K} \sum_{k=1}^K s(x) \quad (15)$$



For the purposes of the current study, it is significant to note that the partition that, among the  $K$  evaluated, has the highest  $SIL$  value is the one that best fits the examined data set and should therefore be taken into consideration for SHM.

Cluster analysis can automatically distinguish between structural situations without making any assumptions about the site's past structural condition. However, user input is required to determine whether the output is compact or mixed over time.

It creates several difficulties to implement the method in real-time applications. To overcome this issue, Santos et al. [31] proposed a robust strategy for damage detection using a cluster-based algorithm with moving time windows. It relies on the average difference between clusters rather than relying on the allocation of data objects to cluster over time by defining the following damage index  $DC$  [31].

$$DC = \frac{1}{K(K-1)} \sum_{k=1}^K \sum_{\substack{c=1 \\ c \neq k}}^K d_{ck} \quad (16)$$

$c$  and  $k$  are two of the  $K$  clusters,  $K$  is the number of clusters from the partition with the highest  $SIL$ , and  $d_{ck}$  is the Gowda-Diday dissimilarity calculated between their centroids. The K-means clusters are similar and will produce low values of  $DC$  if there is no damage and the structural behavior is stable. Therefore, the K-means algorithm will produce distinct and independent clusters and high values of the  $DC$  if damage is detected. Santos et al. [31] further contend that  $DC$  is particularly sensitive to early damage and capable of large data fusion because the information gathered from numerous sensors is referred to as a single-value index.

Ideally, TRUE/FALSE binary information should be provided by the detection strategy.  $L$  is the time duration of each data sample,  $S$  is the number of samples per window, and  $SL$  is the fixed length in which the time windows are specified. Santos et al. [31] propose the statistical testing of the  $DC$  values acquired inside each time window's length because the value of  $DC$  is not informative by itself (it does not provide TRUE/FALSE information). The statistical testing of the  $DC$  values results in the definition of the confidence boundary ( $CB$ ), which should be exceeded only in the case the structural system exhibits changes.

The  $CB$  is defined at each time window, by statistically testing  $DC$ s distribution, under the assumption that only the random effects influence the residual errors obtained from unchained structures. Generally, normal statistical distribution is used in SHM works related to damage detection. To describe small samples taken from a Gaussian population, the  $t$ -student distribution is more suited because there are just a few  $DC$  values in each time window [31].

The confidence boundary at each time window is obtained as follows:

$$CB = E[DC] + t_{s-1, \frac{1}{2} + \frac{\beta}{2}} \times E[DC - E[DC]] / \sqrt{S} \quad (17)$$

where  $t_{s-1, \frac{1}{2} + \frac{\beta}{2}}$  is the  $\frac{1}{2} + \frac{\beta}{2}$  percentile of a  $t$ -student distribution with  $S - 1$  degrees of freedom, and  $\beta$  is the confidence level taken at 99.9%.

$E[DC]$  and  $E[DC - E[DC]]$  are expected value and variability estimates of the  $DC$  sample within the analyzed time window, respectively.

Finally, Santos et al. [31] suggested an original detection index,  $DI$ , based on  $DC$  and  $CB$  values. It is defined to have the following properties: (i) it is dimensionless, (ii) it has an unsupervised and window-wise character by using only data pertaining to a single time window, and (iii) it has positive values that indicate damage detection ("TRUE") and negative or null values that stand for unchanged structural response ("FALSE"). As defined by the following,  $DI$  (detection index) is defined as the ratio between the median

DC (damage index) value observed within a time interval and the largest DC deviation observed above the CB (confidence boundary).

$$DI = \frac{\max_i(DC_i - CB)}{\text{med}_i(DC_i)}; i = 1, \dots, S \quad (18)$$

### 3. Case Study

#### 3.1. Description of the Concrete Bridge Model

The proposed methodology is applied to a numerical model of a two-span continuous beam bridge's superstructure component under changing operating and environmental conditions. The model was developed starting from an original benchmark steel bridge presented by Tassis and Chatzi [34] via open-source Python scripts, which were made available through GitHub (Table 1), with the objective of representing better actual bridge structures.

**Table 1.** Geometry and mechanical properties of the numerical bridge superstructure.

Geometry (Units)	Symbol	Value
Left-span length (m)	$L_1$	10
Right-span length (m)	$L_2$	10
Total bridge length (m)	$L$	20
Bridge height (m)	$h$	0.6
Bridge width (m)	$t$	10
Material properties at $T = 20 \text{ }^\circ\text{C}$		
Mass density ( $\text{kg}/\text{m}^3$ )	$\rho$	2400
Young's modulus (GPa)	$E$	35
Poisson's ratio	$\nu$	0.2
Shear modulus (GPa)	$G$	14.6

The bridge superstructure (Figure 2) consists of a two-span bridge with equal length ( $L_1$  and  $L_2 = 10 \text{ m}$ ). The cross-section of the beam is rectangular with a constant width of 10 m and a height of 0.6 m.



**Figure 2.** Geometry of the two-span bridge in the longitudinal direction with elastic boundaries.

The bridge is assumed to be made from reinforced concrete with  $E = 35 \text{ GPa}$ . It is a linear elastic material with density  $\rho = 2400 \text{ kg}/\text{m}^3$  at ambient temperature  $T = 20 \text{ }^\circ\text{C}$  and Poisson's ratio  $\nu = 0.2$ . Table 1 summarizes the geometry and the mechanical properties.

The bridge is supported by nine elastomeric rubber bearings that are modeled as three evenly spaced elastic supports at each abutment and intermediate pier and put on the bottommost edge of the bridge superstructure.

Table 2 summarizes the global horizontal stiffness,  $k_x$ , and vertical stiffness,  $k_y$ , adopted for the three spring supports that consider the stiffness of the corresponding elastomeric bearings designed according to the loads acting on the bridge (self-weight, permanent load, traffic load, temperature, etc.).

**Table 2.** Horizontal stiffness,  $k_x$ , and vertical stiffness,  $k_y$ , adopted for the three spring supports.

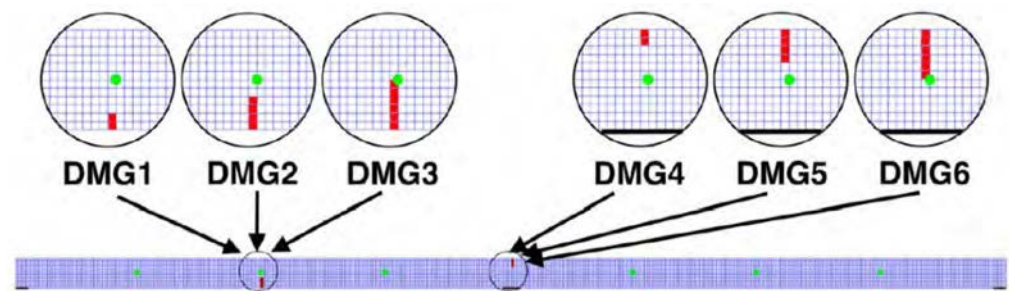
Stiffness (Units)	Left Support	Mid Support	Right Support
$k_x$ (N/m)	10,714,200	19,285,500	10,714,200
$k_y$ (N/m)	$10^{15}$	$10^{20}$	$10^{15}$

### 3.2. Damage Scenarios and Sensors

In order to provide information about the nodal variables in both the  $x$ - and  $y$ - directions, six sensing points, or “sensors,” are thought of (i.e., displacements, velocities, accelerations, strains, etc.). Table 3 describes where these sensors are located. Figure 3 displays the six chosen sensors as green points. Furthermore, six damage scenarios (DMG1 to DMG6) are modeled in the bridge. Damages are introduced by assigning lower stiffness characteristics to selected series of elements. In the following, they are termed “damage-induced mesh elements” (red elements in Figure 3).

**Table 3.** Sensor locations along the bridge.

Sensor	Characteristics	Position on the Beam Neutral Axis ( $y = 0.3$ m)
S-01	$\frac{1}{4} L_1$ from left-hand support	$x = 2.5$ m
S-02	$\frac{1}{2} L_1$ from left-hand support	$x = 5.0$ m
S-03	$\frac{3}{4} L_1$ from left-hand support	$x = 7.5$ m
S-04	$\frac{3}{4} L_2$ from right-hand support	$x = 12.5$ m
S-05	$\frac{1}{2} L_2$ from right-hand support	$x = 15.0$ m
S-06	$\frac{1}{4} L_2$ from right-hand support	$x = 17.5$ m



**Figure 3.** Sensors (in green) and damage locations (in red) in bridge structure.

The first damage region is found in the center of the left span (DMG1, DMG2, DMG3), starting from the bottommost edge of the beam cross-section, while the second damage region is found in the intermediate support section (DMG4, DMG5, DMG6), starting from the uppermost edge of the beam. Moreover, the number of damage-induced mesh elements also varies, as shown in Figure 3. For instance, DMG1 and DMG4 cover an area of two damaged elements, DMG2 and DMG5 a zone of four damaged elements, while DMG2 and DMG6 a zone of six damaged elements. As a result, the damaged elements are 0.05 m wide and range in height from 0.1 to 0.3 m. Table 4 provides a summary of the descriptions of these six damage situations.

**Table 4.** Damage scenarios provided by the benchmark.

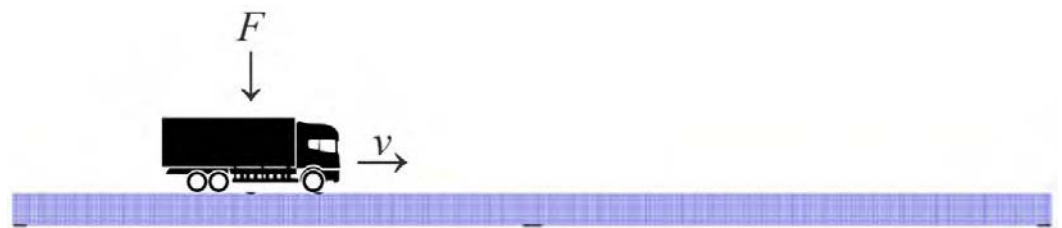
Damage Scenarios	Damaged Elements	Location of Damage
Undamaged Scenario (UND)	0	
Damage Scenario 1 (DMG1)	2	At $\frac{1}{2} L_1$ from left-hand support, starting from the bottommost edge
Damage Scenario 2 (DMG2)	4	
Damage Scenario 3 (DMG3)	6	
Damage Scenario 4 (DMG4)	2	At $L_1$ from left-hand support, starting from the uppermost edge
Damage Scenario 5 (DMG5)	4	
Damage Scenario 6 (DMG6)	6	

In addition to the damage scenarios, the model allows controlling the severity of the damage by setting the reduction of Young’s modulus.

Stiffness reduction is noted by D, followed by the amount of reduction. For example, D50% means 50% reduction in Young's modulus at a given element and damage scenario.

### 3.3. Loading

To perform a time-history analysis simulating traffic flow, a deterministic moving load is used. Data processing methods such as the Hilbert Huang transform can detect these singularities in the signal caused by vibrations created by the moving load and is the appropriate tool to analyze the non-stationary accelerations produced by the crossing load along the bridge [29,30]. The passage of a vehicle is modeled as a moving load  $F$  with constant speed  $v$  in this case study, as shown in Figure 4. The weight of a standard truck of about 30 tons is considered.



**Figure 4.** Loading in the form of a moving vertical force  $F$  with a constant speed  $v$ .

In summary, the following parameters are considered for the time-history analysis:

- The first, second, and third bending modes are selected;
- The Rayleigh damping coefficients are  $\alpha = 0.1654$  and  $\beta = 5.4333 \times 10^{-6}$ ;
- The vehicle velocity is 10 m/s;
- A sampling frequency of  $f_s = 400$  Hz in the virtual sensors, a time step in the time domain analysis of  $\Delta t = 0.001$  s, a final time step  $T_f = 2$  s (time to cross the bridge), and a total number of time samples of 800 are considered.

### Simulation of Temperature

To check the ability of the machine learning methodology proposed to overcome the problems related to the temperature effects in the recorded signals, different responses are obtained under different ambient temperatures. This is introduced in the numerical model by taking into account the relationship between temperature and Young's modulus of concrete. The work by Jiao et al. [35] was assumed, where a linear relationship between the modulus of elasticity and temperature was given:

$$E_c = -0.125T + 29.13 \quad (19)$$

$$R^2 = 0.9852 \quad (20)$$

For the present study  $E_c = 35$  GPa is taken at a reference temperature of 20 °C, and the relationship is modified accordingly.

$$E_c = -0.125T + 37.5 \quad (21)$$

### 3.4. Results and Discussion

As stated earlier, the main objective of this paper is to propose and test the machine-learning-based methodology for damage detection in bridges that will be able to distinguish the real damage from the environmental changes and using the forced vibrations due to the traffic crossing. To simulate those conditions, it was decided to model a passing of the truck at two different temperatures and in both damaged and undamaged states.

Therefore, the following extreme scenarios for time-history response due to traffic action were preliminarily chosen in the present work:

1. UND,  $T = -15\text{ }^{\circ}\text{C}$ , D0%;
2. UND,  $T = 20\text{ }^{\circ}\text{C}$ , D0%;
3. DMG3,  $T = 20\text{ }^{\circ}\text{C}$ , D90%;
4. DMG3,  $T = -15\text{ }^{\circ}\text{C}$ , D90%.

In all cases, the damage is introduced at the mid-span of the left span.

Figure 5 shows the obtained accelerations at Sensor S01 for the four considered scenarios.

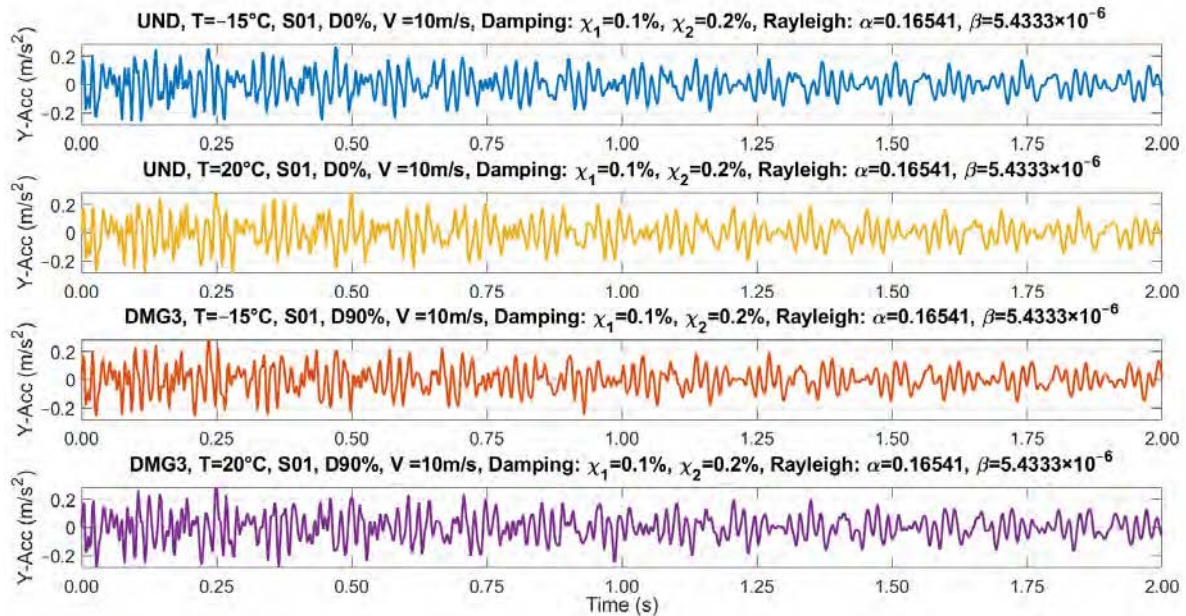


Figure 5. Recorded Y-accelerations at S01.

### 3.4.1. VMD

VMD is a signal decomposition technique with its own experimental parameters. Since the IMFs reveal the frequency information contained in the original signal, the peaks in the Fourier spectrum of this signal give a rough indication of the number of IMFs to be extracted from a particular signal decomposition technique. Therefore, the single-sided amplitude spectrum of the decomposed IMFs is also discussed.

There are six parameters that must be set to apply VMD  $\varepsilon_a$ ,  $\varepsilon_r$ ,  $O$ ,  $\alpha$ ,  $\tau$  and  $K$ .  $\varepsilon_a$  and  $\varepsilon_r$  are absolute and relative tolerance, respectively. In this case study, the absolute tolerance is more restrictive than the relative tolerance, as in the implementation of MATLAB. Therefore,  $\varepsilon_a = 0.1$  and  $\varepsilon_r = 10^{-5}$ . A high number of iterations  $O = 100,000$  is set to stop the VMD only when relative tolerance is met.

The efficiency of the VMD method has been demonstrated. The algorithm always seeks to reconstruct the time series effectively with minimum error. The VMD method ensures no mixture of modes, as well as the orthogonality of the transformation. Although this section refers to the signal decomposition for the undamaged state at  $T = 20\text{ }^{\circ}\text{C}$  in Sensor S-01 (Figure 6), the same procedure is performed for all sensors and compared to the damage scenarios. This study revealed that the signal decomposition is similar for every damage scenario using the same VMD parameters as in Sensor S01:  $\varepsilon_a = 0.1$ ,  $\varepsilon_r = 10^{-5}$ ,  $\tau = 0.1$ ,  $\varepsilon_r = 10^{-5}$  and  $O = 100,000$ , with the exception of the number of modes  $K$ , which varies from one sensor to another, as shown in Table 5.

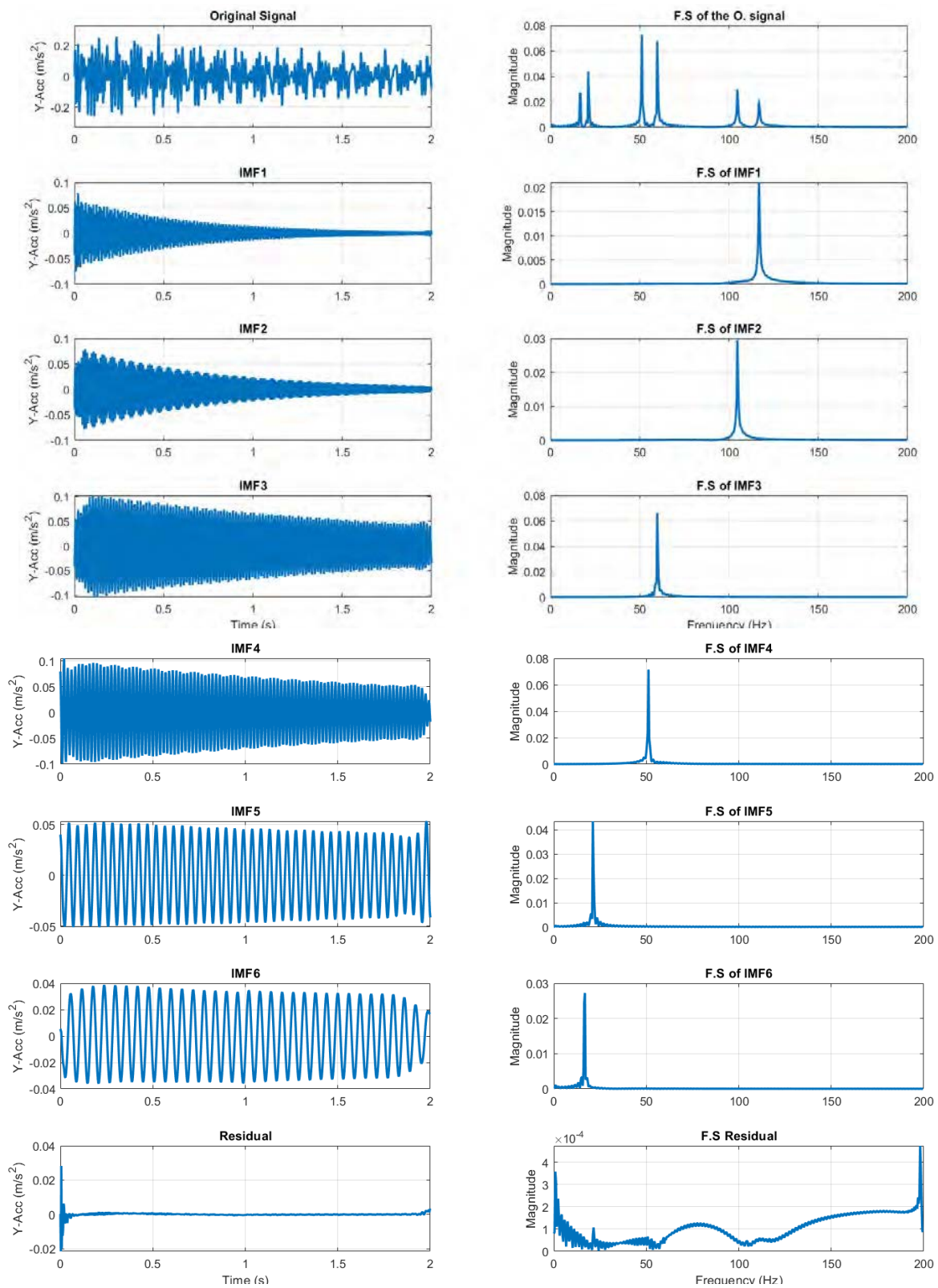
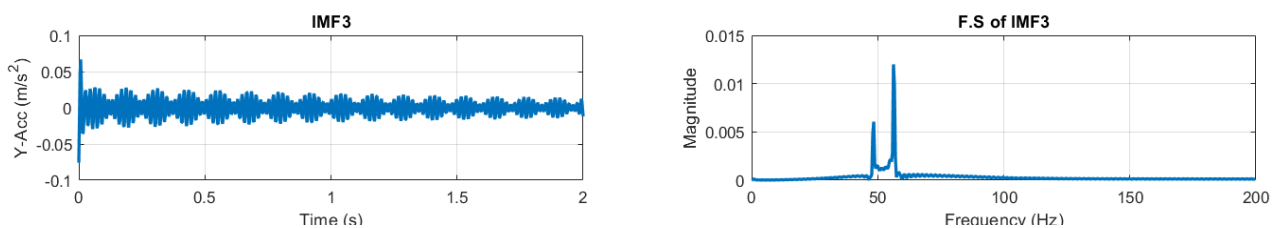


Figure 6. Y-acceleration signal of undamaged state from S01 at  $T = 20\text{ }^{\circ}\text{C}$  and its VMD.

**Table 5.** VMD parameters for each sensor.

Parameters	S01	S02	S03	S04	S05	S06
Number of modes, $K$	6	5	6	6	5	6
Relative tolerance, $\varepsilon_r$	$10^{-5}$	$10^{-5}$	$10^{-5}$	$10^{-5}$	$10^{-5}$	$10^{-5}$
Penalty factor, $\alpha$	1000	1000	1000	1000	1000	1000
Fidelity coefficient, $\tau$	0.1	0.1	0.1	0.1	0.1	0.1
Absolute tolerance, $\varepsilon_a$	0.1	0.1	0.1	0.1	0.1	0.1
Number of iterations, $O$	100,000	100,000	100,000	100,000	100,000	100,000

$K$  is found to be 6 for Sensors S01, S03, S04, and S06, which are located at  $\frac{1}{4}$  from the spring supports, whereas  $K$  is established as 5 for Sensors S02 and S05, which are placed in the middle of the left span and right span, respectively. In these sensor signal amplitudes, the second bending mode is far less than in the rest of the sensors, which results in a mix of the asymmetric and symmetric modes, as shown in IMF3 in Figure 7.

**Figure 7.** IMF 3 of Y-acceleration signal of undamaged state from S02 at  $T = 20$  °C.

### 3.4.2. Hilbert Transform

The Hilbert transform is applied to study the characteristics of different time-varying parameters obtained from scenarios and to obtain damage-sensitive features. In particular, the goal is to examine each IMF and obtain instantaneous frequency as a function of time. It can be achieved by applying the Hilbert transform (HT) to each IMF extracted from the application of the VMD method.

The instantaneous frequencies  $f_k(t)$  are obtained by applying the Hilbert transform to the  $k$  physically meaningful IMFs, e.g., Figure 8 for Sensor S-01. It can be noticed how boundary conditions affect the response: when the load is at the extremities, over the abutments (time instant of 0 s and 2 s), larger oscillations are generated in the model because of the load concentration and the elastic properties (no dissipative effects) of the numerical model. Therefore, it was decided to concatenate the analysis on the instantaneous frequencies in the interval of 0.1–1.5 s as the appropriate time signal for the analysis (Figure 9).

### 3.4.3. Application of PCA (Eliminating Environmental Effects)

When PCA [14,36–38] is applied to the data set, the first principal component (PC1) represents the factor that creates the greatest variance within the data set, the second principal component (PC2) represents the factor that creates the second greatest variance affecting the data, and so on.

According to Soo et al. [37], the data set needs to come from two extreme and opposing temperature situations in order to accurately depict the impacts of temperature fluctuations in PC1. As a result, if the data set is standardized, the extreme cases will be displayed on the opposing sides of the PC1 graph, indicating negative and positive values of the variance. Therefore, additional minor elements affecting the data set, such as structural damages, will be represented by the other primary components.

As shown in the flowchart in Figure 10, the PCA methodology to filter the temperature effects impacting the vibration parameters may be broken down into eight primary steps.

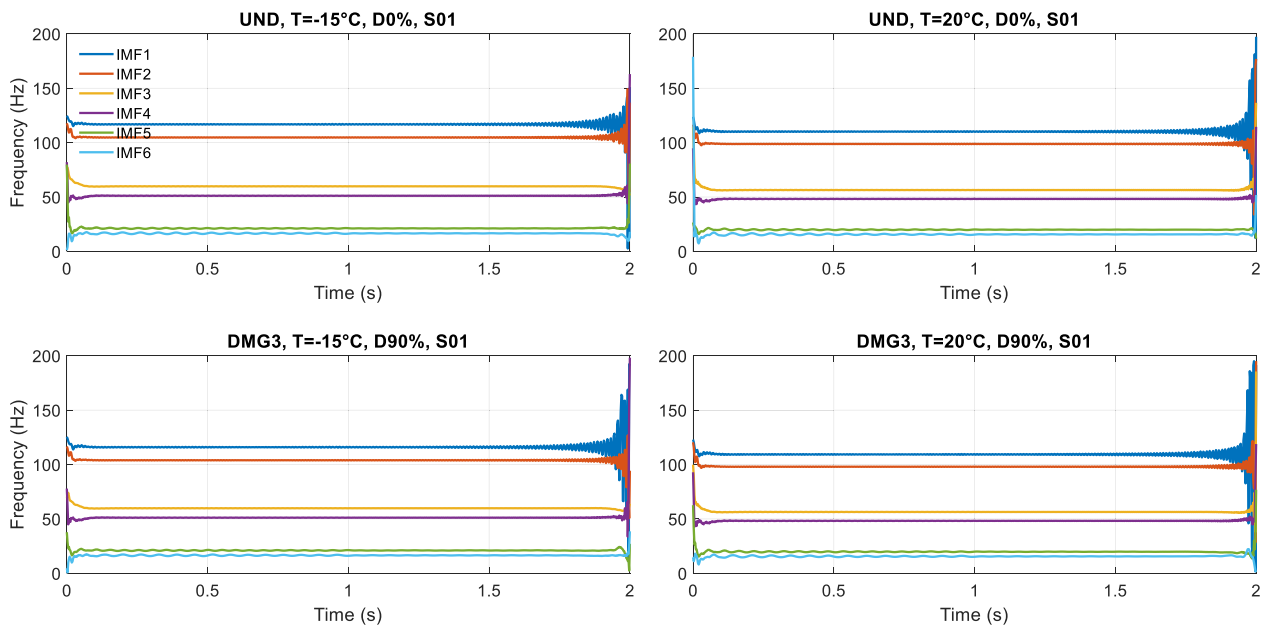


Figure 8. Instantaneous frequency for each scenario in Sensor S-01.

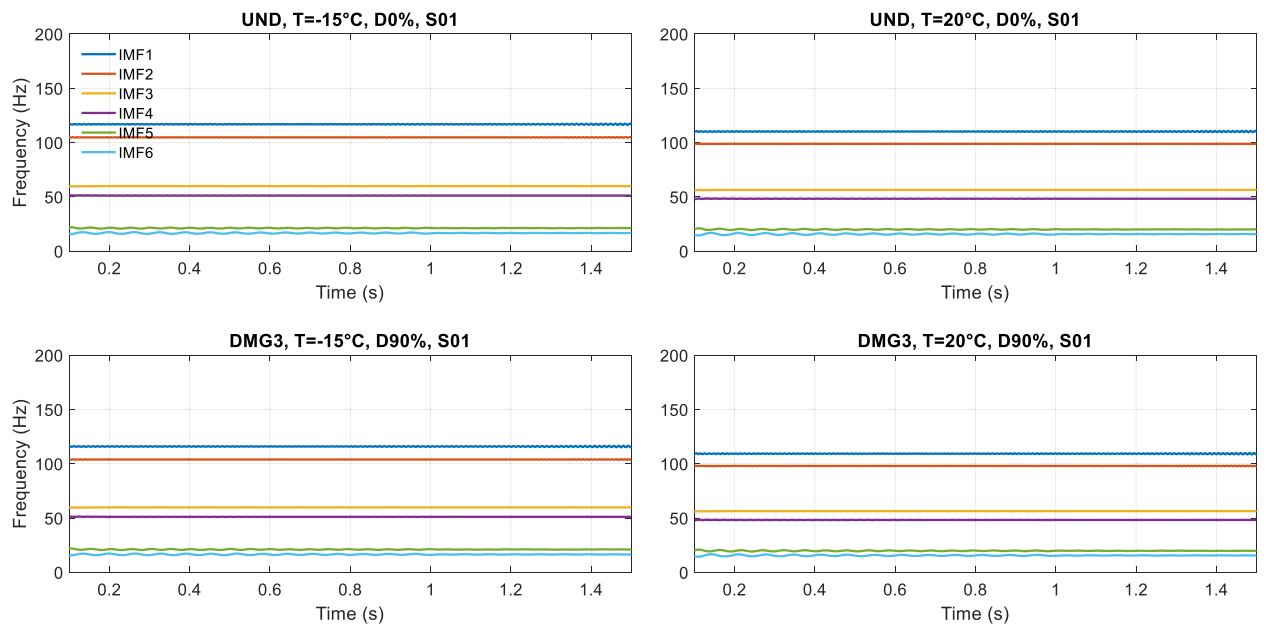


Figure 9. Concatenated instantaneous frequencies for each scenario in Sensor S-01.

The basic data set for typical PCA is a 2-D matrix,  $X_{n,m}$ , consisting of  $n$  observations (e.g., temperature variability) and  $m$  measured variables (e.g., modes). The correlation between these variables is examined using the standard PCA method. However, because the observed variables (e.g., instantaneous frequencies) are continuous in time in this case study, a third dimension,  $k$ , is added, resulting in  $X_{n,m,k}$ .

In this case study, correlations between a particular measured variable (instantaneous frequency) obtained for each IMF  $m$  and several temperature observations,  $n$ , are analyzed. Therefore, each time sample and each sensor are treated independently. As an illustration, Figure 11 shows the unfolding in time samples of the data set corresponding to a sensor resulting in a 3-D data matrix,  $X_{n,m,k}$ .



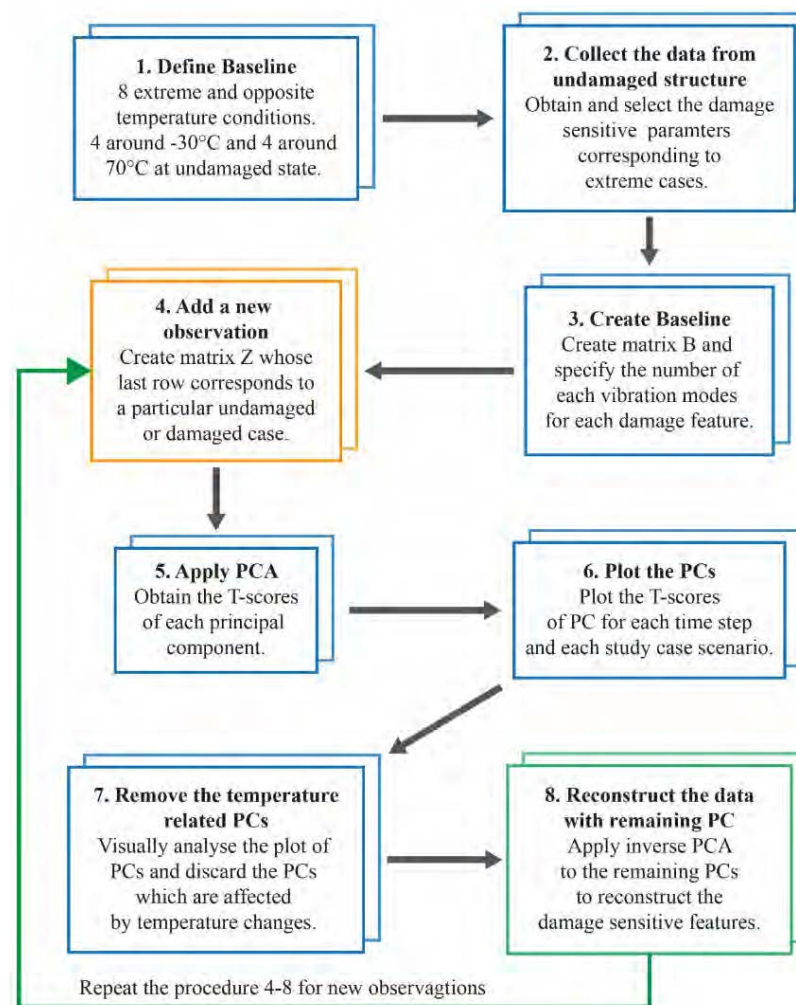


Figure 10. Flow chart of the PCA-based filtering method.

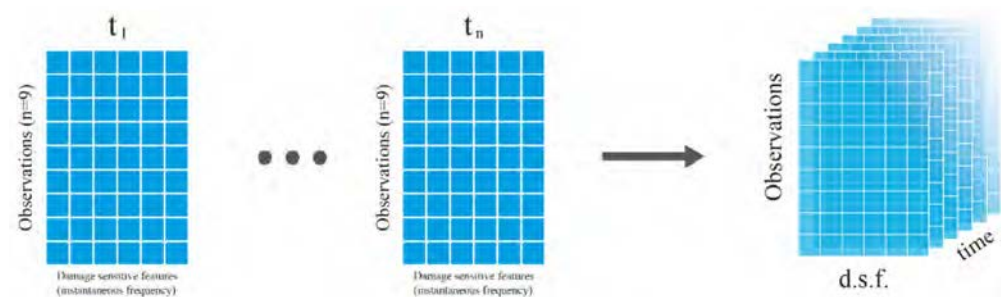
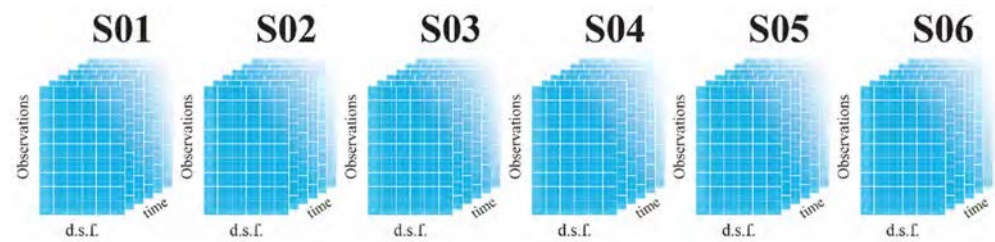


Figure 11. Unfolding of PCA matrix for a sensor.

Moreover, a fourth dimension,  $l$ , must be added to the data matrix when various sensors are examined. As a result, a four-dimensional data matrix  $X_{n,m,k,l}$  with a high number of connected variables is created, as illustrated in Figure 12.

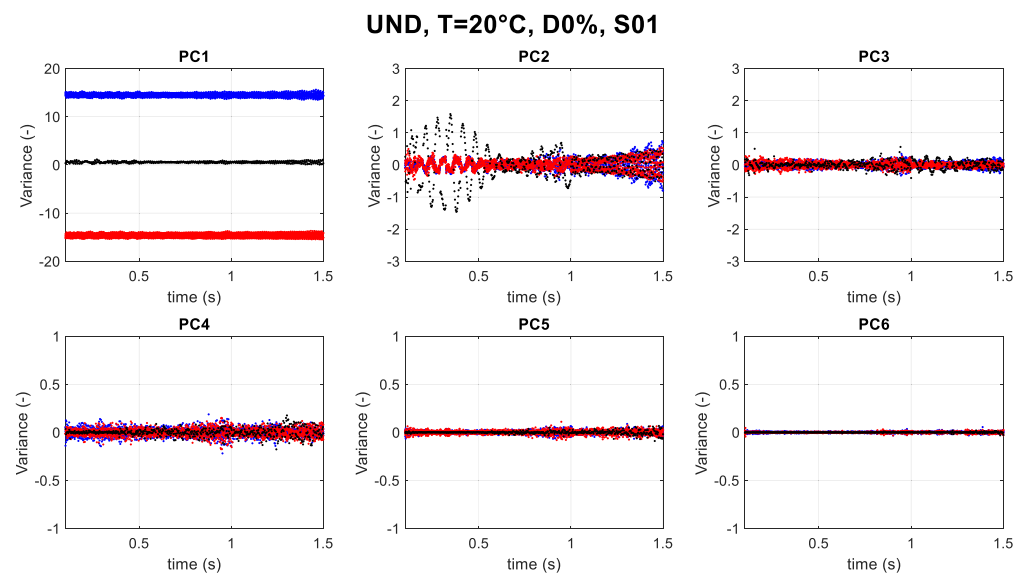
On the one hand, the number of observations is  $n = 9$  corresponding to the 8 extreme temperatures (case observations) to create the baseline plus one new observation. The number of modes,  $m$ , represents the number of IMFs in which the original signal was decomposed. As seen in the previous section, the number of IMFs  $m$  varies from one sensor to another:  $m = 6$  for Sensors S01, S03, S04, and S06, and  $m = 5$  for Sensors S02 and S05. As a reminder, the instantaneous modal parameters were obtained for the interval of 0.1–1.5 s and a sampling frequency of 400 Hz was selected to capture the first three bending modes of vibration, hence dimension  $k = 560$ . The vertical acceleration response obtained for any

temperature condition was decomposed by means of the VMD-based method with the following parameters:  $\alpha = 1000$ ,  $\varepsilon_a = 0.1$ ,  $\varepsilon_r = 10^{-5}$ ,  $\tau = 0.1$ ,  $\varepsilon_r = 10^{-5}$  and  $O = 100,000$ .



**Figure 12.** Complete data matrix for PCA.

In the following, the graphical results for the undamaged cases and damaged cases corresponding to scenarios at 20 °C temperature are presented (black dots) in Figures 13 and 14 (for PC2 only). These represent a representative sample of the results over different scenarios. Moreover, to create the baseline, four extreme cases at low temperatures of around −30 °C (blue dots) and four extreme cases at high temperatures of around 70 °C (red dots) are considered.

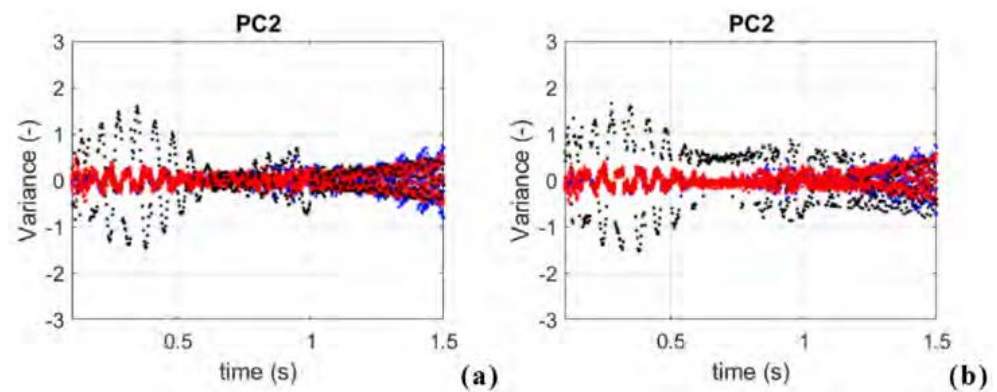


**Figure 13.** PCA of UND T = 20 °C, D0%, S01 (black dots); blue dots: extreme negative temperatures (−30 °C, −29 °C, −28 °C, −27 °C); red dots: extreme positive temperatures (67 °C, 68 °C, 69 °C, 70 °C).

The instantaneous frequency  $x_k(t)$  of each IMF was determined. Figure 13 illustrates the six components for the undamaged scenario cases of 20 °C for Sensor S01. On the vertical axis, the variance of the principal components (dimensionless) is reported. Furthermore, it can be noted that the magnitude of the observations in PC1 is much larger than that in PC2 for all considered scenarios.

As mentioned, PC1 related to the data shows the highest variance related to the temperature. As scenarios under consideration are at −15 °C and 20 °C, they both fall between extreme cases, both in damaged and undamaged conditions, thus making it clear that PC1 is definitively temperature related.

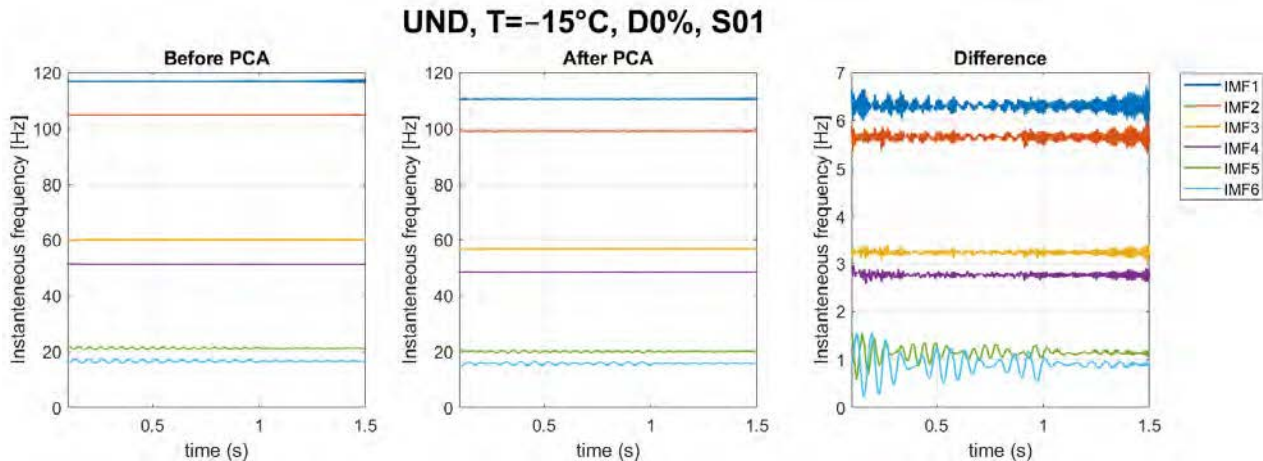
Moreover, it is important to note that PC2 can reveal some changes in the structure response as well. If two cases of damaged and undamaged conditions at T = 20 °C are compared (Figure 14), the difference in PC2 can be observed.



**Figure 14.** Comparison of PC2: (a) UND  $T = 20\text{ }^{\circ}\text{C}$ , D0%, S01 (black dots); blue dots: extreme negative temperatures ( $-30\text{ }^{\circ}\text{C}$ ,  $-29\text{ }^{\circ}\text{C}$ ,  $-28\text{ }^{\circ}\text{C}$ ,  $-27\text{ }^{\circ}\text{C}$ ); red dots: extreme positive temperatures ( $67\text{ }^{\circ}\text{C}$ ,  $68\text{ }^{\circ}\text{C}$ ,  $69\text{ }^{\circ}\text{C}$ ,  $70\text{ }^{\circ}\text{C}$ ) (b) DMG3,  $T = 20\text{ }^{\circ}\text{C}$ , D90%, S01 (black dots); blue dots: extreme negative temperatures ( $-30\text{ }^{\circ}\text{C}$ ,  $-29\text{ }^{\circ}\text{C}$ ,  $-28\text{ }^{\circ}\text{C}$ ,  $-27\text{ }^{\circ}\text{C}$ ); red dots: extreme positive temperatures ( $67\text{ }^{\circ}\text{C}$ ,  $68\text{ }^{\circ}\text{C}$ ,  $69\text{ }^{\circ}\text{C}$ ,  $70\text{ }^{\circ}\text{C}$ ).

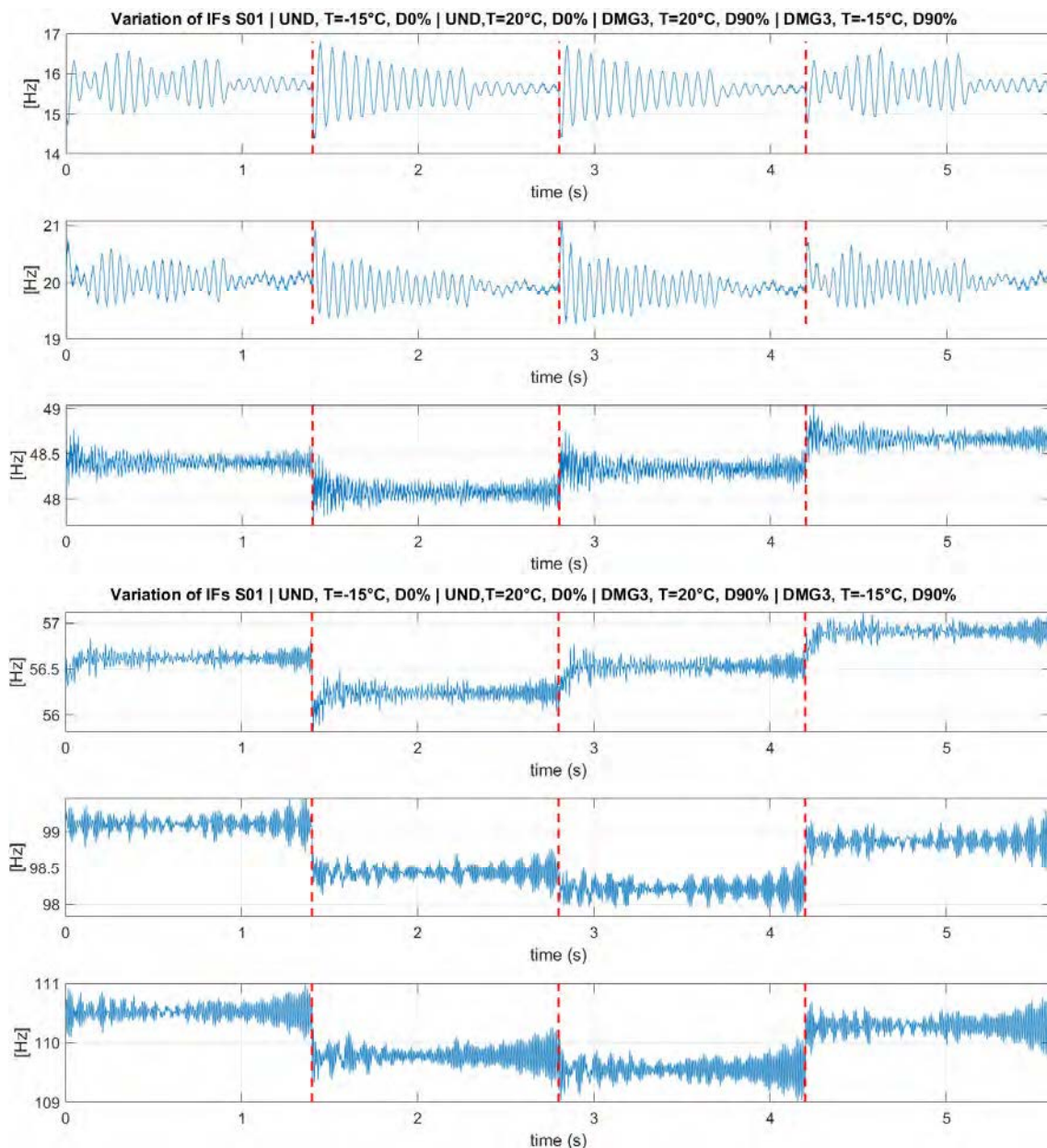
As a next step, the damage-sensitive features (instantaneous frequencies) are reconstructed using the five last principal components (from PC2 to PC6). It is worth noting that the reconstructed data should be rescaled to have a physical meaning.

In Figure 15 below, the IF from a representative scenario (undamaged:  $-15\text{ }^{\circ}\text{C}$ ) at S01 is shown before and after the application of PCA dimensionality reduction. It is worth noting that IF at  $20\text{ }^{\circ}\text{C}$  had a lower change in magnitude between, after, and before the application of PCA with respect to  $-15\text{ }^{\circ}\text{C}$ .



**Figure 15.** Instantaneous Frequencies before and after the application of PCA dimensionality reduction.

Figure 16 shows the variation of the instantaneous frequencies at each damage scenario from S01. As can be observed, although PCA dimensionality reduction was applied at high frequencies ( $\sim 48\text{ Hz}$ ,  $\sim 56.5\text{ Hz}$ ,  $\sim 98.7\text{ Hz}$ ,  $\sim 110\text{ Hz}$ ), the difference between damage and temperature scenarios in their value is visibly clear. The application of cluster analysis to such data will not be appropriate, as the goal of the methodology is to distinguish between damaged and undamaged states when the influence of the temperature is not easily observed. Moreover, signals from S02 and S05 were disregarded as well, as the mode mixing problem due to the VMD created data mixing after the PCA analysis, making instantaneous frequencies unreliable.

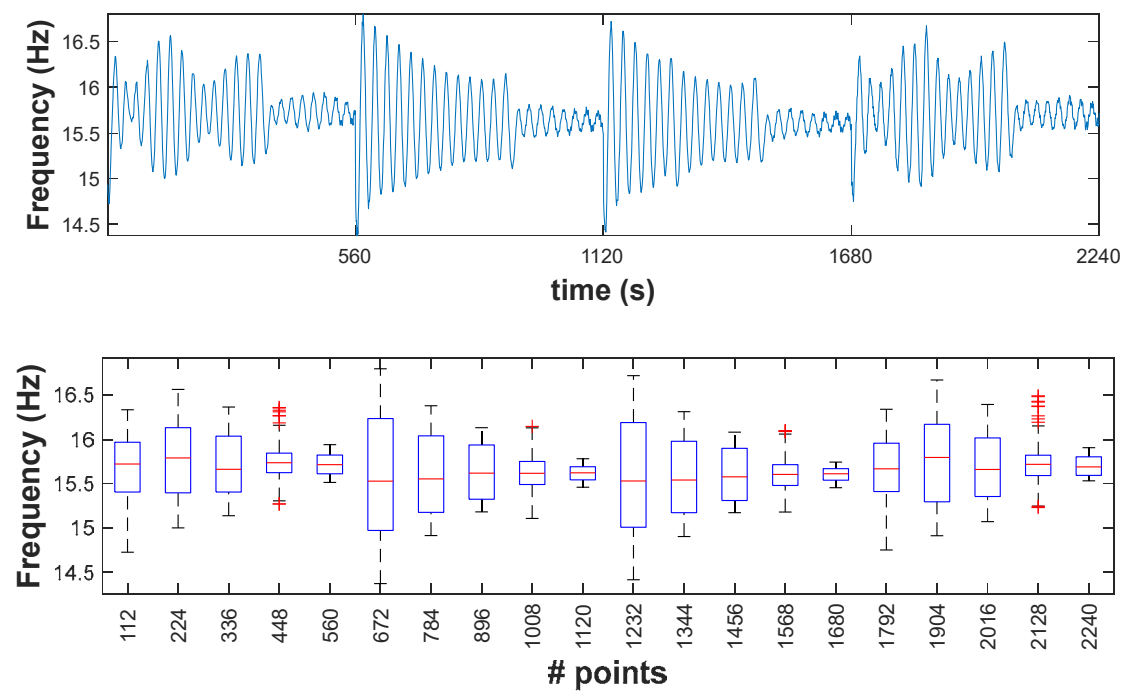


**Figure 16.** Variation of the instantaneous frequencies of different scenarios related to S01.

In conclusion, only the first instantaneous frequencies ( $\sim 15.6$  Hz) from S01, S03, S04, and S06 were considered for the clustering algorithm.

#### 3.4.4. K-Means

Before the application of the K-means algorithm, the data must be reduced to more generic types and less voluminous information in contrast with classical data used in SHM. The extracted damage-sensitive features, in this case the instantaneous frequency of the four scenarios with a total time of 5.6 s, are converted into symbolic data by using the interquartile interval. The resulting total number of points was 2240. A symbolic data length of  $L = 112$  points was considered, and a respective boxplot for each sensor was constructed (Figure 17).



**Figure 17.** First instantaneous frequency of all scenarios for S01 and corresponding box plot.

Moreover, in the implementation of the K-means algorithm in MATLAB, the suitable method “*Start*” to determine the initial clusters centroid positions (or seeds) should be set among the options (Table 6).

**Table 6.** Method for choosing initial cluster centroid positions.

“cluster”	Perform a preliminary clustering phase on a random 10% subsample of $X$ when the number of observations in the subsample is greater than $k$ . This preliminary phase is itself initialized using “sample”. If the number of observations in the random 10% subsample is less than $k$ , then the software selects $k$ observations from $X$ at random.
“plus” (default)	Select $k$ seeds by implementing the $k$ -means++ algorithm for cluster center initialization.
“sample”	Select $k$ observations from $X$ at random.
“uniform”	Select $k$ points uniformly at random from the range of $X$ . Not valid with the Hamming distance.
“numeric matrix”	$k$ -by- $p$ matrix of centroid starting locations. The rows of <i>Start</i> correspond to seeds. The software infers $k$ from the first dimension of <i>Start</i> , so you can pass in [] for $k$ .
“numeric array”	$k$ -by- $p$ -by- $r$ array of centroid starting locations. The rows of each page correspond to seeds. The third dimension invokes replication of the clustering routine. Page $j$ contains the set of seeds for replicate $j$ . The software infers the number of replicates (specified by the “Replicates” name-value pair argument) from the size of the third dimension.

For this study, the “sample” case was found as a consistent and appropriate method for the initial cluster centroid positions.

As explained, the detection method is based on the value of the  $DC$ , which is the difference between clusters. To calculate  $DC$ , the clusters of moving windows must be

defined. In this work, the size of windows was selected as  $S = 5$  symbolic data, each comprising  $L = 112$  points of the damage-sensitive feature. Figure 18 shows a sequence of mobile windows, while Figure 19 shows the DC values obtained for each time window.

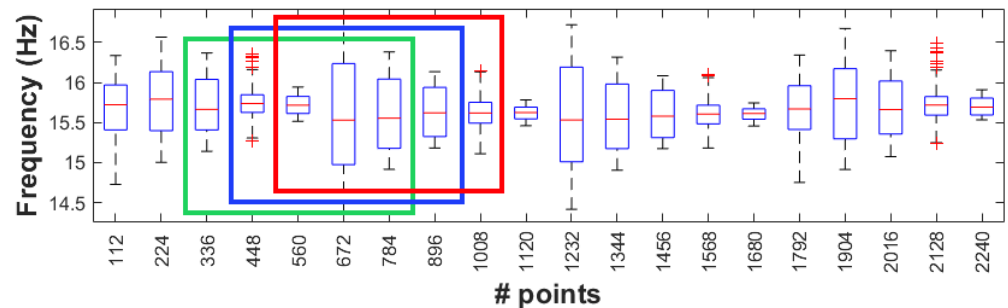


Figure 18. A sequence of mobile windows.

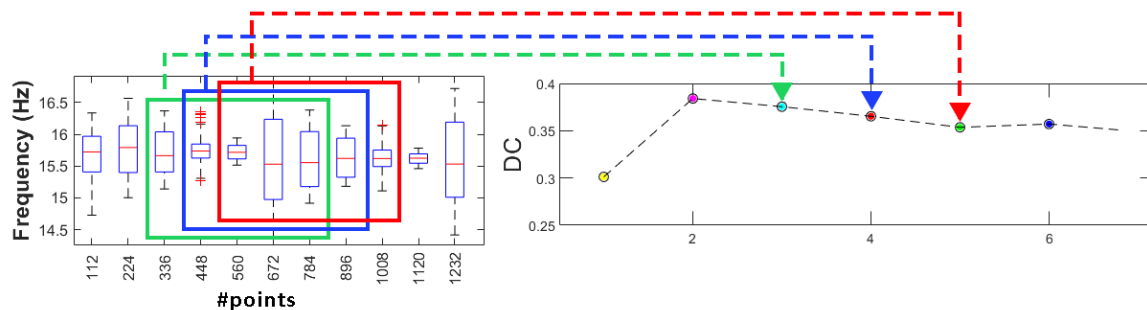


Figure 19. DC values obtained for each time window.

The total number of symbolic data equals 20, and the number of mobile windows and DC values is 16.

It should be noted that if the time window increases, then higher sensitivity to damage detection is obtained, but the time between damage occurrence and damage detection increases. Moreover, if the samples within a time window decrease, it is possible to detect the damage earlier, but the probability of false detections increases (there is less sensitivity to damage detection). Therefore, it is necessary to find a balance between the two objectives: detect damage as soon as possible but obtain high confidence in the detection.

After the  $DC$  values are obtained, they must be statistically tested by the confidence boundary test.  $CB$  is defined for each time window containing five values of  $DC$ .

Finally, using  $DC$  and  $CB$  values the original detection index,  $DI$  is calculated. When the  $DI$  value is negative or zero, there is no damage in the structure, and when it is positive, it indicates the presence of damage in the structure.

Figure 20 shows the whole procedure for the IF ( $\sim 15.6$  Hz) from S01 for the damage scenarios. It can be seen that the clustering algorithm exactly detects damage (positive value of  $DI$ ) when it crosses the scenarios between UND T = 20 °C D0% and DMG3 T = 20 °C D90% (marked with the black dotted vertical line). However, temperature change either in the undamaged or damaged scenarios does not create a positive  $DI$  value, therefore, not warning on the existence of damage (false detection).

One of the disadvantages of the K-means algorithm is its dependence on initial values. As the initial cluster centroids are assigned randomly, they may create different outputs each time the algorithm runs. However, in this case, this disadvantage was found useful to localize the damage. Specifically, an innovative aspect highlighted by this research is the following: K-means gives consistent detection of damage when using the data coming from sensor S01, the closest to DM3. In fact, as presented in Figure 21, four trials (experiments) of the algorithm gives always detection of damage when it appears, without false detections.

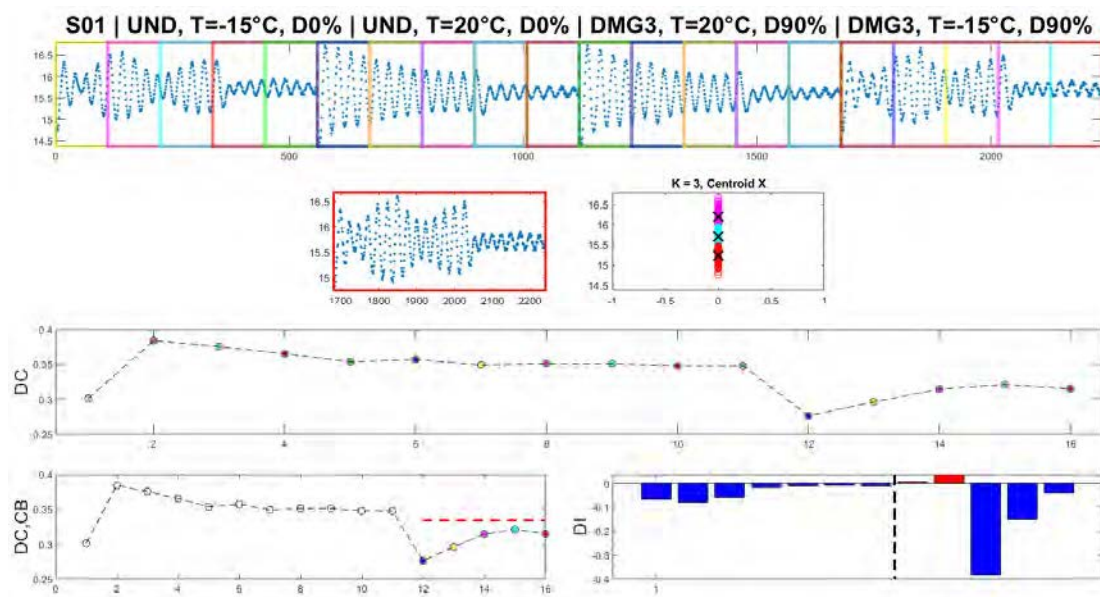


Figure 20. K-means clustering on the S01 first IF from different scenarios (Dashed line on the DI diagram: damage introduction).

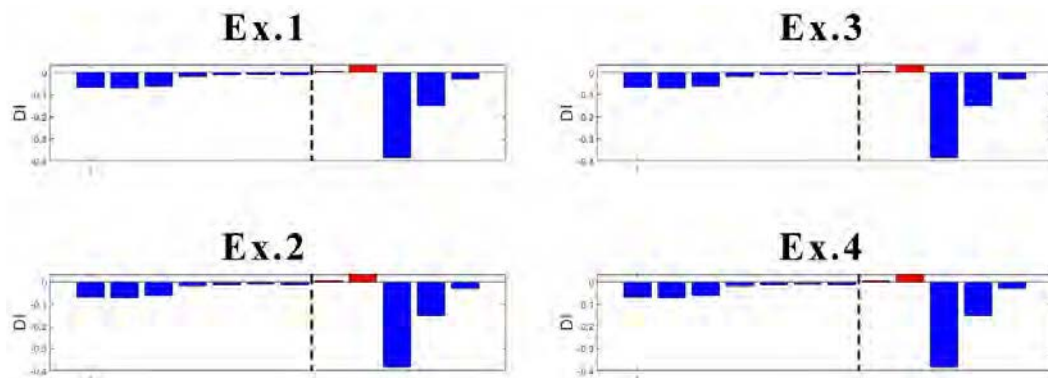


Figure 21. Experiments with DI for S01 (dashed line: damage introduction with 90% stiffness reduction).

In the case of sensor S03 (Figure 22), which is the second sensor closest to the damage, K-means provides early false detection in all cases. However, true damage detection is always obtained.

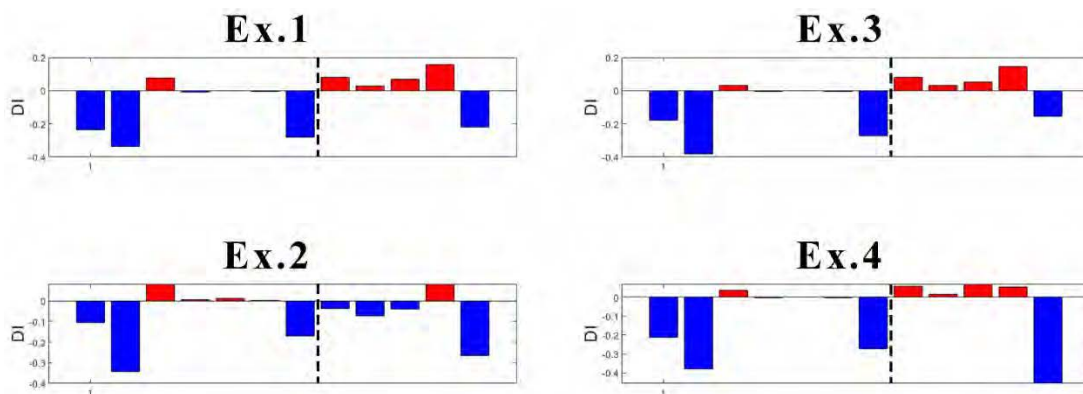
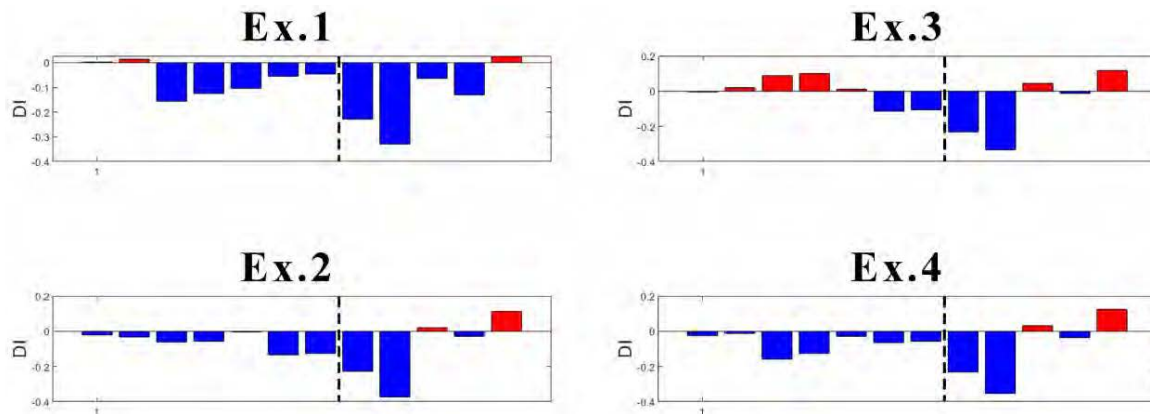


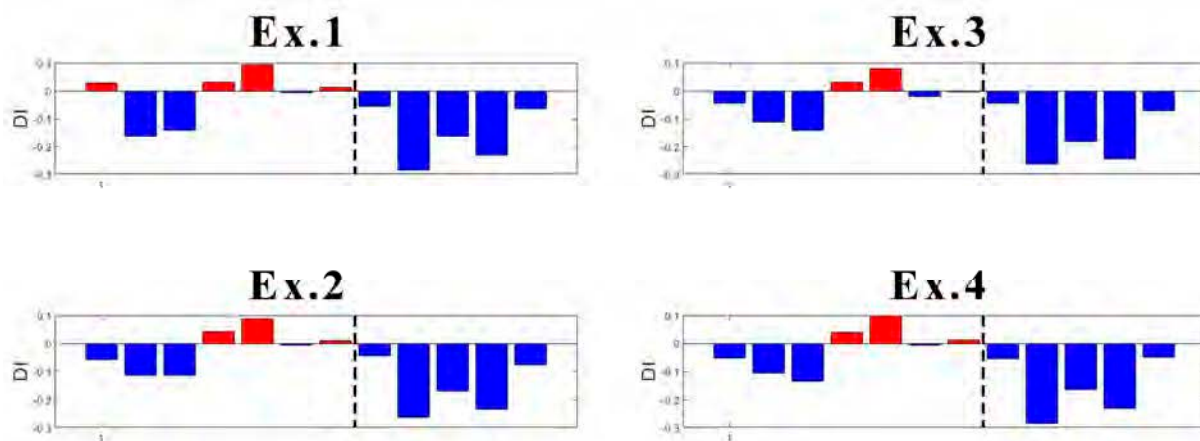
Figure 22. Experiments with DI for S03 (dashed line: damage introduction with 90% stiffness reduction).

What concerns the results of Sensors S04 and S06 located more far away from the damage location is that they both provided inconsistent results. With high randomness in their output, both early false detection and no detection of true damage can be seen in their results (Figures 23 and 24).



**Figure 23.** Experiments with DI for S04 (dashed line: damage introduction with 90% stiffness reduction).

Focusing on the results of DI (Figures 21 and 22), it can be said that the damage is located somewhere between Sensors S01 and S03, which is correct as the damage was introduced at the mid-span of the left span. The results are not conclusive when looking at sensors far away from where the damage is present. Thus, as a further contribution to the present study, the proposed methodology is suitable not only for damage detection, but also for damage localization.



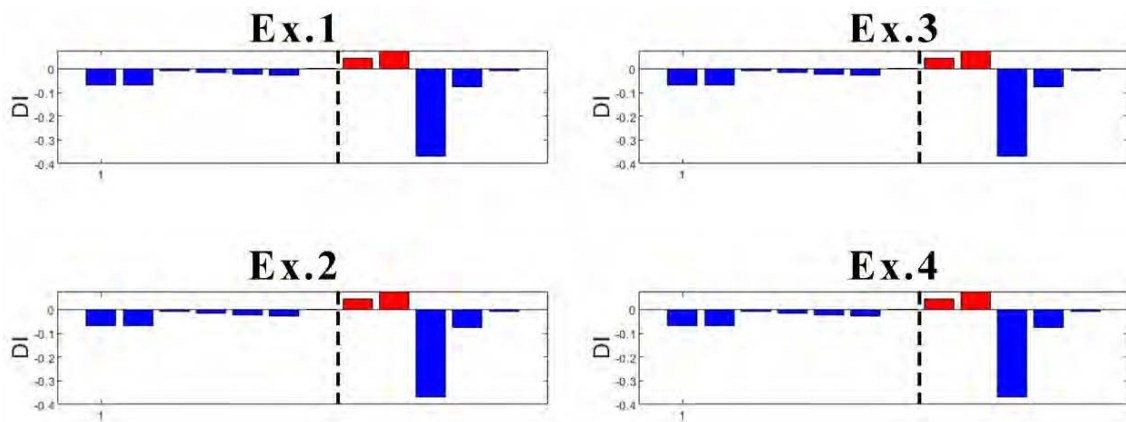
**Figure 24.** Experiments with DI for S06 (dashed line- damage introduction with 90% stiffness reduction).

Finally, the methodology was preliminarily tested out for the case with a smaller severity of damage (30% of reduction of  $E$ ), considering Sensor S01 only and the same damage position already simulated for the previous analysis. In this case, the results are promising regarding the detection of damage, as reported in Figure 25.

However, it should be pointed out that the percent reduction in modulus of elasticity to simulate damage is applied to a single, rather narrow, and localized slice of vertical elements. Thus, the damage is rather sharp on the model, while in reality, the damage, e.g., by degradation of the mechanical properties of an element, is often more widespread. In addition, the ability of the method to localize the damage on the bridge when close to



the measurement point suggests the application of a diffuse monitoring system, such as distributed fiber-optic sensors.



**Figure 25.** Experiments with DI for S01 (dashed line: damage introduction with 30% stiffness reduction).

#### 4. Conclusions

This paper is focused on a methodology for damage detection and localization in bridges under both traffic loading and environmental variability by modeling different damage and temperature scenarios. The presented methodology is based on time-series data collected from a finite-element model.

Due to the non-stationary nature of recorded data, the fast Fourier transform is not applicable. Instead, the vertical acceleration measurements obtained from six sensors spaced along the bridge-like structure were analyzed using the Hilbert–Huang transform with variational model decomposition.

To remove the temperature effects from the damage-sensitive features, principal component analysis with dimensionality reduction was applied. It was observed that the first principal component with the highest variance in the data matrix was responsible for the temperature-related effects. As a result, the damage-sensitive features were reconstructed by removing the first principal component.

Finally, to detect and localize the damage, a machine learning algorithm (K-means) was applied. Using the symbolic data to reduce the amount of data, a technique of moving the time window was used for damage-sensitive features. A confidence boundary was deployed to evaluate damage indices values for each window, and a detection index was then defined to consolidate a result with high confidence.

Results showed that K-means in combination with principal component analysis with dimensionality reduction can identify and localize damages even with changing temperatures and analyze non-stationary vibrations due to traffic in the bridge. The methodology showed its effectiveness and reliability in the field of bridge structural health monitoring and damage detection and localization, under both traffic-induced loads and temperature changes.

This paper is the first step in this direction, but consequently, several limitations are still present, such as the use of a numerical example and not a real bridge due to the lack of vibration data under health and damage conditions and with varying temperatures, the simple modeling of the vehicle as a moving load, and the crossing of only one vehicle present in the bridge. Therefore, further steps of the research will consider different damage levels and locations, multiple damage states, and a real-world case study of a bridge structure to test the proposed methodology in different conditions under operational and environmental effects.

**Author Contributions:** Conceptualization, M.D. and J.R.C.; Methodology, S.N., J.R.C. and R.M.D.; Software, S.N. and R.M.D.; Validation, S.N. and R.M.D.; Formal analysis, S.N., M.D., J.R.C. and R.M.D.; Investigation, S.N., M.D., J.R.C. and R.M.D.; Resources, M.D. and R.M.D.; Data curation, S.N.; Writing—original draft, S.N. and M.D.; Writing—review & editing, S.N., M.D., J.R.C. and R.M.D.; Supervision, M.D. and J.R.C.; Project administration, M.D. and J.R.C. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research received no external funding.

**Acknowledgments:** The authors thank Eleni Chatzi, ETH Zurich, for providing us with the data and documentation from the numerical bridge utilized in the case study.

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

## References

1. Gkoumas, K.; Santos, F.L.M.D.; van Balen, M.; Tsakalidis, A.; Hortelano, A.O.; Grosso, M.; Haq, G.; Pekár, F. *Research and Innovation in Bridge Maintenance, Inspection and Monitoring—A European Perspective Based on the Transport Research and Innovation Monitoring Information System (TRIMIS)*; Publications Office of the European Union: Luxembourg, 2019.
2. Rania, N.; Coppola, I.; Martorana, F.; Migliorini, L. The collapse of the Morandi Bridge in Genoa on 14 August 2018: A collective traumatic event and its emotional impact linked to the place and loss of a symbol. *Sustainability* **2019**, *11*, 6822. [[CrossRef](#)]
3. Rytter, A. *Vibrational Based Inspection of Civil Engineering Structures*; Department of Building Technology and Structural Engineering, Aalborg University: Aalborg, Denmark, 1993.
4. Malekloo, A.; Ozer, E.; Al Hamaydeh, M.; Girolami, M. Machine learning and structural health monitoring overview with emerging technology and high-dimensional data source highlights. *Struct. Health Monit.* **2021**, *21*, 1906–1955. [[CrossRef](#)]
5. Zhang, F.-L.; Yang, Y.; Xiong, J.-H.; Yu, Z. Structural health monitoring of a 250-m super-tall building and operational modal analysis using the fast Bayesian FFT method. *Struct. Control Health Monit.* **2019**, *26*, 23–83. [[CrossRef](#)]
6. Yesilyurt, I.; Gursoy, H. Estimation of elastic and modal parameters in composites using vibration analysis. *J. Vib. Control* **2015**, *21*, 509–524. [[CrossRef](#)]
7. Chen, B.; Zhao, S.-L.; Li, P.-Y. Application of Hilbert-Huang transform in structural health monitoring: A state-of-the-art review. *Math. Probl. Eng.* **2014**, *2014*, 317954. [[CrossRef](#)]
8. Yi, T.H.; Li, H.N.; Sun, H.M. Multi-stage structural damage diagnosis method based on energy-damage theory. *Smart Struct. Syst.* **2013**, *12*, 345–361. [[CrossRef](#)]
9. Gurley, K.; Kareem, A. Applications of wavelet transforms in earthquake, wind and ocean engineering. *Eng. Struct.* **1999**, *21*, 149–167.
10. Xu, Y.L.; Chen, B. Integrated vibration control and health monitoring of building structures using semi-active friction dampers. Part I: Methodology. *Eng. Struct.* **2008**, *30*, 1789–1801. [[CrossRef](#)]
11. Chen, B.; Xu, Y.L. Integrated vibration control and health monitoring of building structures using semi-active friction dampers. Part II: Numerical investigation. *Eng. Struct.* **2008**, *30*, 573–587. [[CrossRef](#)]
12. Mostafa, N.; Di Maio, D.; Loendersloot, R.; Tinga, T. Extracting the time-dependent resonances of a vehicle–bridge interacting system by wavelet synchrosqueezed transform. *Struct. Control Health Monit.* **2021**, *28*, e2833. [[CrossRef](#)]
13. Santos, J.; Crémone, C.; Orcesi, A.D.; Silveira, P. Multivariate statistical analysis for early damage detection. *Eng. Struct.* **2013**, *56*, 273–285. [[CrossRef](#)]
14. Wah, W.S.L.; Chen, Y.-T.; Elamin, A.; Roberts, G.W. Damage detection under temperature conditions using PCA—An application to the Z24 Bridge. *Proc. Inst. Civ. Eng. Struct. Build.* **2022**, *175*, 890–902.
15. Tenelema, F.J.; Delgadillo, R.M.; Casas, J.R. Bridge damage detection and quantification under environmental effects by Principal Component Analysis. In *Proceedings of the 1st Conference of the European Association on Quality Control of Bridges and Structures—EUROSTRUCT 2021, Padua, Italy, 28 August–1 September 2021*; LNCE Series; Springer: Cham, Switzerland, 2022; Volume 200, pp. 183–190.
16. Delgadillo, R.M.; Casas, J.R. Bridge damage detection via Improved Completed Ensemble EMD with Adaptive Noise and machine learning algorithms. *Struct. Control Health Monit.* **2022**, *29*, e2966. [[CrossRef](#)]
17. Wang, X.; Gao, Q.; Liu, Y. Damage Detection of Bridges under Environmental Temperature Changes Using a Hybrid Method. *Sensors* **2020**, *20*, 3999. [[CrossRef](#)] [[PubMed](#)]
18. Huang, H.-B.; Yi, T.-H.; Li, H.-N.; Liu, H. Strain-Based Performance Warning Method for Bridge Main Girders under Variable Operating Conditions. *J. Bridge Eng.* **2020**, *25*, 04020013. [[CrossRef](#)]
19. Meixedo, A.; Santos, J.; Ribeiro, D.; Calçada, R.; Todd, M. Damage detection in railway bridges using traffic-induced dynamic responses. *Eng. Struct.* **2021**, *238*, 112189. [[CrossRef](#)]
20. Meixedo, A.; Santos, J.; Ribeiro, D.; Calçada, R.; Todd, M.D. Online unsupervised detection of structural changes using train-induced dynamic responses. *Mech. Syst. Signal Process.* **2022**, *165*, 108268. [[CrossRef](#)]
21. Zhang, S.; Wang, Y.; Yu, K. Steady-State Data Baseline Model for Nonstationary Monitoring Data of Urban Girder Bridges. *Sustainability* **2022**, *14*, 12134. [[CrossRef](#)]

22. Wang, Y.; Yang, D.-H.; Yi, T.-H. Displacement model error-based method for symmetrical cable-stayed bridge performance warning after eliminating variable load effects. *J. Civ. Struct. Health Monit.* **2021**, *12*, 81–99. [CrossRef]
23. Dragomiretskiy, K.; Zosso, D. Variational Mode Decomposition. *IEEE Trans. Signal Process.* **2014**, *62*, 531–544. [CrossRef]
24. Bagheri, A.; Ozbulut, E.O.; Harris, K.D. Structural system identification based on variational mode. *J. Sound Vib.* **2018**, *417*, 182–197. [CrossRef]
25. Muñoz, F.J.T. Bridge Damage Identification under Operational and Environmental Variability. Master's Thesis, Department of Civil and Environmental Engineering, UPC, Barcelona, Spain, 2020.
26. Zhang, M.; Jiang, Z.; Feng, K. Research on variational mode decomposition in rolling bearings fault diagnosis of the multistage centrifugal pump. *Mech. Syst. Signal Process.* **2017**, *93*, 460–493. [CrossRef]
27. Liu, S.; Tang, G.; Wang, X. Time frequency analysis method for rotary mechanical fault based on improved variational mode decomposition. *J. Vib. Eng.* **2016**, *29*, 1119–1126.
28. Wu, S.; Feng, F.; Zhu, J.; Wu, C.; Zhang, G. A Method for Determining Intrinsic Mode Function Number in Variational Mode Decomposition and Its Application to Bearing Vibration Signal Processing. *Shock Vib.* **2020**, *2020*, 8304903. [CrossRef]
29. Pines, D.; Salvino, L. Health Monitoring of One Dimensional Structures Using Empirical Mode Decomposition and the Hilbert-Huang Transform. *Proc. SPIE Int. Soc. Opt. Eng.* **2002**, *4701*, 127–143.
30. Cohen, L. *Time-Frequency Analysis*; Prentice Hall: Hoboken, NJ, USA, 1995.
31. Santos, J.P.; Crémona, C.; Calado, L.; Silveira, P.; Orcesi, A.D. On-line unsupervised detection of early damage. *Struct. Control. Health Monit.* **2016**, *23*, 1047–1069. [CrossRef]
32. Xu, D.; Tian, Y. A Comprehensive Survey of Clustering Algorithms. *Ann. Data Sci.* **2015**, *2*, 165–193. [CrossRef]
33. Santos, J.P. Smart Structural Health Monitoring Techniques for Novelty Identification in Civil Engineering Structures. Ph.D. Thesis, University of Lisbon, Lisbon, Portugal, 2014.
34. Tatsis, K.; Chatzi, E. A Numerical Benchmark for System Identification under Operational and Environmental Variability. In Proceedings of the 8th International Operational Modal Analysis Conference (IOMAC 2019), Copenhagen, Denmark, 13–15 May 2019.
35. Jiao, Y.; Liu, H.; Wang, X.; Zhang, Y.; Luo, G.; Gong, Y. Temperature Effect on Mechanical Properties and Damage Identification of Concrete Structure. *Adv. Mater. Sci. Eng.* **2014**, *2014*, 191360. [CrossRef]
36. MATLAB. k-Means Clustering. Available online: <https://it.mathworks.com/help/stats/kmeans.html#d123e555240> (accessed on 1 January 2023).
37. Wah, W.S.L.; Chen, Y.-T.; Roberts, G.W.; Elamin, A. Separating damage from environmental effects affecting civil structures for near real-time damage detection. *Struct. Health Monit.* **2018**, *17*, 850–868.
38. Mujica, L.E.; Rodellar, J.; Fernandez, A.; Guemes, A. Q-statistic and T2-statistic PCA-based measures for damage assessment in structures. *Struct. Health Monit.* **2011**, *10*, 539–553. [CrossRef]

**Disclaimer/Publisher's Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.