XLIII CILAMCE
21–25, November, 2022
FOZ DO IGUAÇU – PR – BRAZIL

Unsupervised feature selection-based technique for locating structural deterioration: a multi-domain approach

Victor H. M. Alves[1], Alexandre A. Cury[1]

[1]*Universidade Federal de Juiz de Fora*
Rua José de Lourenço Kelmer, 36036-900, São Pedro, Juiz de Fora, Minas Gerais, Brazil
victor.meneguitte@engenharia.ufjf.br, alexandre.cury@engenharia.ufjf.br

Abstract. Structural monitoring methods have been extensively researched in recent years due to developments in Artificial Intelligence (AI) technology. In this regard, the purpose of this work is to offer an automated data-driven approach for deterioration localization based on the extraction of features from raw vibration data utilizing domain knowledge and a filtering procedure. To diversify information retrieval, feature extraction is conducted concurrently in temporal, frequency, and quefrency domains. This filtering process is known as feature selection (FS) and is used to reduce redundancies and raise the relevance of the feature set by removing a subset based on a predefined criterion. The key idea is that the proposed approach may be tuned to the structure while offering generality for whatever shape, material, or excitation it comes across. The deterioration index is calculated via outlier analysis referenced by the structure's healthy condition. The technique was successfully tested in a full-scale bridge, demonstrating a performance that is encouraging for real-world monitoring scenarios.

Keywords: Structural Health Monitoring, Damage Localization, Automated, Feature Selection, Multi-domain.

1 Introduction

The lifespan of a given structure relies on a variety of factors, including the weather, usage, design, and others which are difficult to predict and sometimes impracticable to evaluate. As a result, real-time Structural Health Monitoring (SHM) systems have drawn considerable attention from scientists in a variety of research areas, including civil, aerospace, and structural engineering. Recent developments in sensor technology tend to improve robustness and support ever-cheaper pricing for implementation in large-scale applications. This is a noteworthy accomplishment since the localization becomes more precise the more sensors are arranged on a certain building.

Model-driven and data-driven are the two primary classifications of SHM. This first class entails developing a FE model of the structure, which is frequently related to a model updating procedure. The data-driven approach, on the other hand, employs vibration measurements (such as acceleration or displacement), frequently using feature extraction combined to a kind of pattern recognition approach. Data-driven methods generally provide a more useful solution for deterioration identification due to the computational model's complexity and the inevitable differences between the vibration attributes of the real building and the FE model. Thus, this research proposes a completely automated data-driven and multi-domain approach for localizing structural deterioration enhanced with an unsupervised feature selection technique.

2 Multi-domain approach

The effectiveness of the extraction step has a major impact on all feature-based techniques. Therefore, it becomes an attractive idea to divide the number of features into multiple domains, i.e., time, frequency, and quefrequency domains. This "multi-domain association" is expected to improve the feature set's representation of the dynamic responses (i.e., acceleration signals), as a structural change (e.g., in stiffness, due to local deterioration) might have an influence on all of these domains concurrently.

2.1 Time domain features

To gather as much information and nuances about the vibration signals as possible, this paper uses a variety of statistics in time domain evaluated directly from the acceleration data. Typically, its computation has low complexity. In this study, 17 time domain features are extracted, these are listed in table 1.

Table 1. Time domain features definition

Feature	Definition	Feature	Definition
Peak (PE)	$Y_{PE} = \max(\boldsymbol{y_i})$	Kurtosis (KU)	$Y_{KU} = \frac{\sum_{i=1}^{N}(\boldsymbol{y_i} - Y_\mu)^4}{(N-1){Y_\sigma}^4}$
Root-Mean-Square (RMS)	$Y_{RMS} = \left(\frac{1}{N}\sum_{i=1}^{N}{\boldsymbol{y_i}}^2\right)^{\frac{1}{2}}$	5th Moment (5th M)	$Y_{5^{th}m} = \frac{\sum_{i=1}^{N}(\boldsymbol{y_i} - Y_\mu)^5}{(N-1){Y_\sigma}^5}$
Square-Mean-Root (SMR)	$Y_{SMR} = \left(\frac{1}{N}\sum_{i=1}^{N}\lvert\boldsymbol{y_i}\rvert^{\frac{1}{2}}\right)^2$	6th Moment (6th M)	$Y_{6^{th}m} = \frac{\sum_{i=1}^{N}(\boldsymbol{y_i} - Y_\mu)^6}{(N-1){Y_\sigma}^6}$
Range (RG)	$Y_{RG} = \max(\boldsymbol{y_i}) - \min(\boldsymbol{y_i})$	7th Moment (7th M)	$Y_{7^{th}m} = \frac{\sum_{i=1}^{N}(\boldsymbol{y_i} - Y_\mu)^7}{(N-1){Y_\sigma}^7}$
Mean (M)	$Y_M = \frac{1}{N}\sum_{i=1}^{N}\boldsymbol{y_i}$	Shape Factor (SF)	$Y_{SF} = \frac{Y_{RMS}}{\frac{1}{N}\sum_{i=1}^{N}\lvert\boldsymbol{y_i}\rvert}$
Mean Square (MS)	$Y_{MS} = \frac{1}{N}\sum_{i=1}^{N}{\boldsymbol{y_i}}^2$	Crest Factor (CF)	$Y_{CF} = \frac{\max(\lvert\boldsymbol{y_i}\rvert)}{y_{RMS}}$
Variance (VAR)	$Y_{\sigma^2} = \frac{1}{N-1}\sum_{i=1}^{N}\lvert\boldsymbol{y_i} - Y_\mu\rvert^2$	Impulse Factor (IF)	$Y_{IF} = \frac{\max(\lvert\boldsymbol{y_i}\rvert)}{\frac{1}{N}\sum_{i=1}^{N}\lvert\boldsymbol{y_i}\rvert}$
Standard deviation (SD)	$Y_\sigma = \sqrt{Y_{\sigma^2}}$	Latitude Factor (LF)	$Y_{LF} = \frac{\max(\lvert\boldsymbol{y_i}\rvert)}{Y_{SMR}}$
Skewness (SK)	$Y_{SK} = \frac{\sum_{i=1}^{N}(\boldsymbol{y_i} - Y_\mu)^3}{(N-1){Y_\sigma}^3}$		

2.2 Frequency domain features

The proposed methodology aims to obtain valuable attributes from the frequency domain responses in addition to the time domain characteristics. Therefore, acceleration responses are used to calculate the well-known Power Spectral Densities (PSD). Being a commonly utilized tool in the field of signal processing, PSD are regularly used to represent the distribution of a signal over frequency. The work of Beskhyroun et al. [1] is suggested for a deeper knowledge of PSD applied to SHM tasks.

Frequency-related features are retrieved from the already mentioned PSD via an analysis of the k highest peaks. The key hypothesis is that changes in the highest k peaks cause changes in the PSD, frequently associated with anomalous structural vibration behavior, hence indicating potential damage spots. A distance vector $\boldsymbol{D}$ is calculated between succeeding peaks as follows in equation 1, where k is an integer, and h_i is the location of the i^{th} peak in the PSD (taking into account the ascending frequency order).

$$\boldsymbol{D} = [(h_2 - h_1), (h_3 - h_2), \dots, (h_{i-1} - h_i), \dots, (h_k - h_{k-1})]^T = [d_1, d_2, \dots, d_i, \dots, d_{k-1}]^T \quad \text{for } i = 1,2,\dots,k-1. \tag{1}$$

where d_i is the Euclidean distance between the $(i+1)^{\text{th}}$ and i^{th} peaks. Thus, a vector $\boldsymbol{D}$ of dimension $(k-1)$ is produced from a given set of k peaks.

As a result, five frequency-related properties are retrieved from $\boldsymbol{D}$: mean, standard deviation, variation,

skewness, and kurtosis. This method is repeated twice to increase the amount of detail in the analysis. As shown in Alves and Cury [2], adjusting the parameter *k* to 6 and 8 yields optimal results. Finally, 10 frequency-related features are developed. Note that the same peak might be calculated numerous times due to its width, and closely spaced peaks can bias the result. To solve this problem, a minimum distance between the peaks of 50 points was considered.

2.3 Quefrency domain features

Mel-frequency cepstral coefficient (MFCC), which belongs to the quefrency domain, is added to complete the collection of features for the multi-domain approach. The initial use of cepstrum-based features with a focus on damage identification may be attributed to Zhang et al. [3], who developed noise cancellation and damage detection algorithms in concrete bridge decks. In that survey, MFCCs outperformed other examined features in terms of repeatability and separability.

The steps below (Sahidullah and Saha [4]) illustrate how to compute the MFCC in brief:

I. Windowing: A Hamming window is used to multiply the signal.
II. Power spectrum computation: Construct the discrete Fourier transform (DFT) coefficients of a (windowed) signal, which are subsequently utilized to compute the power spectrum.
III. Log energy filter bank: The power spectrum is processed through a triangle filter based on the Mel scale.
IV. Discrete cosine transform (DCT): computing the DCT of the log Mel spectrum list
V. Obtaining MFCC features.

It is recommended to examine the literature (Sahidullah and Saha [4]) for a more in-depth look at the computations necessary. In this research, 9 features from the quefrency domain are employed.

2.4 Feature selection (FS)

It is clear that the proposed approach is dependent on the extracted features' ability to represent local changes in structural behavior caused by damage. To address the feature relevance problem, the Feature Selection (FS) technique is introduced. FS is becoming a must-have solution for managing projects with extensive and expanding databases (Big Data). This technique is excellent for removing unnecessary features and preventing overfitting. FS methods are frequently divided into three basic categories: filter, wrapper, and embedding. Filter-type approaches choose features devoid of any learning mechanism and independent of the classification model. Information-theoretic measures are used to examine the variables in this situation, and the best features are then picked while the least important ones are discarded. This category is used on the present study, because are known to be quick and unaffected by a specific classifier, which makes them ideal for implementing FS in real-world applications involving structural damage location.

2.5 Unsupervised Infinite Feature Selection (Inf-FS_U)

The Inf-FS_U is an unsupervised filter-type feature selection methodology. It entails treating feature subsets as a path in a fully connected graph, where nodes are features and edges assess the relationship between two linked nodes (features) in terms of relevance and redundancy metrics. The detailed computation process is demonstrated in study of Roffo et al. [5], which performed tests in numerous benchmark datasets and indicated that Inf-FS_U outperformed several other well-known FS approaches. For the present research the set of selected features are determined by the top 80% (truncated form) regarding it scores.

3 Methodology

The 36 multi-domain features in this work are filtered applying Inf-FS_U, leaving the top 28 ranked features to evaluate the deterioration index. Notice that each sensor has its own ranking. The outlier identification system employed in this research is based on Alves and Cury [2].

3.1 Percentile intervals

Percentiles can be used to divide a sample frequency distribution into two sections. In essence, the percentile P_k is a number that k percent of all samples fall below it, where $0 < k \leq 100$. The most common are the first quartile, median, and third quartile, which correspond to P_{25}, P_{50}, and P_{75}, respectively. It is worth noting that the interquartile range in given by $(P_{75} - P_{25})$. The percentile measure was chosen because it is unaffected by outliers. It is a powerful metric as it tends to eliminate noise influence in the signal induced by environmental fluctuations (e.g., temperature and wind velocity) or vehicle traffic.

The percentile interval, $[P_L, P_U]$, is defined as a criterion to represent the structure's healthy behavior trend in the present work. In fact, for a set interval of extreme pairings P_L and P_U, the likelihood of finding a particular sample within it is $(U - L)\%$. For the sake of an automated technique, an interval of 85% is employed as the baseline, which means that for all tests analyzed, the interval $[P_{7.5}, P_{92.5}]$ was selected based on the healthy data. This parameter is evaluated in line with prior research (Alves and Cury [2]), where experiments on various intervals revealed that a little modification (~5%) does not significantly influence the outcomes.

3.2 Deterioration index

As already mentioned, the percentile intervals are obtained using the healthy state features F^h. Therefore, the deterioration index is calculated for each sensor (i.e., accelerometer) based on the number of features in the damaged state F^d that are beyond its range $[P_L, P_U]$. For a particular sensor, it is essential to partition its whole signal into N samples; hence, after extracting the features, there is a distribution of samples of the same feature to offer the creation of its interval and subsequent outlier analysis. In matrix notation, features, i.e., F^h and F^d, are described as:

$$F^h = \left[f_1^h, f_2^h, \dots, f_i^h, \dots, f_s^h\right] \text{ and } F^d = \left[f_1^d, f_2^d, \dots, f_i^d, \dots, f_s^d\right], \text{for } i = 1, 2, \dots, s. \tag{2}$$

where s is the total number of sensors mounted along the construction, and where the definitions of the matrices $\boldsymbol{f_i^h}$ and $\boldsymbol{f_i^d}$ are:

$$\boldsymbol{f_i^h} = \begin{bmatrix} g_{1,1}^h & \cdots & g_{1,j}^h & \cdots & g_{1,nf}^h \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ g_{n,1}^h & \cdots & g_{n,j}^h & \cdots & g_{n,nf}^h \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ g_{N,1}^h & \cdots & g_{N,j}^h & \cdots & g_{N,nf}^h \end{bmatrix} \text{ and } \boldsymbol{f_i^d} = \begin{bmatrix} g_{1,1}^d & \cdots & g_{1,j}^d & \cdots & g_{1,nf}^d \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ g_{n,1}^d & \cdots & g_{n,j}^d & \cdots & g_{n,nf}^d \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ g_{N,1}^d & \cdots & g_{N,j}^d & \cdots & g_{N,nf}^d \end{bmatrix} \tag{3}$$

$$\text{for j} = 1, 2, 3, \dots, nf \text{ and } n = 1, 2, 3, \dots, N.$$

where nf is the total number of features extracted from each signal sample (36 features for this work), and $g_{n,j}$ is a generic variable representing a feature j of a signal sample n. As a result, F^h is used to calculate the standard healthy interval for each feature j, that is also detailed in the previous section. An auxiliary counter matrix $\boldsymbol{AC}$ is initialized with zero-valued components. Its goal is to determine the number of outliers for each feature j and sensor i considering their corresponding healthy intervals.

$$\boldsymbol{AC} = \begin{bmatrix} ac_{1,1} & \cdots & ac_{1,i} & \cdots & ac_{1,s} \\ \vdots & \ddots & \vdots & & \vdots \\ ac_{j,1} & \cdots & ac_{j,i} & \cdots & ac_{j,s} \\ \vdots & & \vdots & \ddots & \vdots \\ ac_{nf,1} & \cdots & ac_{nf,i} & \cdots & ac_{nf,s} \end{bmatrix} \tag{4}$$

The $\boldsymbol{AC}$ counter matrix is updated interactively. A normalization procedure is used to ensure that all features have the same potential to influence the output. Finally, the proposed deterioration index DI_i is determined by Eq. 5.

$$DI_i = \sum_{j=1}^{nf} ac_{j,i}, \text{for } i = 1, \dots, s. \tag{5}$$

4 Application: Z24 bridge

The Z24 bridge was constructed in 1963 in the canton of Bern, near Solothurn, Switzerland. It connected the communities of Koppigen and Utzenstorf by crossing the A1 motorway, which links Bern to Zurich (Figure 1(a)). The building geometric dimensions are illustrated in Figure 1(b). Progressive damage studies were carried out before the building was set to be demolished in late 1998. The tests, which lasted a month, were designed to simulate real bridge deterioration for scientific objectives (Wahab and De Roeck [6]). Nonetheless, some sensors, notably those on the Bern side span, were found to be defective given the existence of an abnormal shift in the offset of the signal, which are being disregarded for this work.

To recreate genuine damage on Z24, the pier on side Koppigen was cut and rebuilt with steel fill plates and three hydraulic jacks (Wahab and De Roeck [6]). This installed mechanism can lower the pier to replicate real-world damage sources such as subsurface settlement and erosion. The period following the installation of the lowering system is used as a baseline for the healthy condition. Four different damage scenarios are being investigated: (I) 20 mm settlement; (II) 80 mm settlement; (III) 95 mm settlement; (IV) pier lifting and foundation tilt. Note that both temperature and wind conditions on this full-scale bridge fluctuate over time, allowing the assessment of the proposed method's robustness to noise.

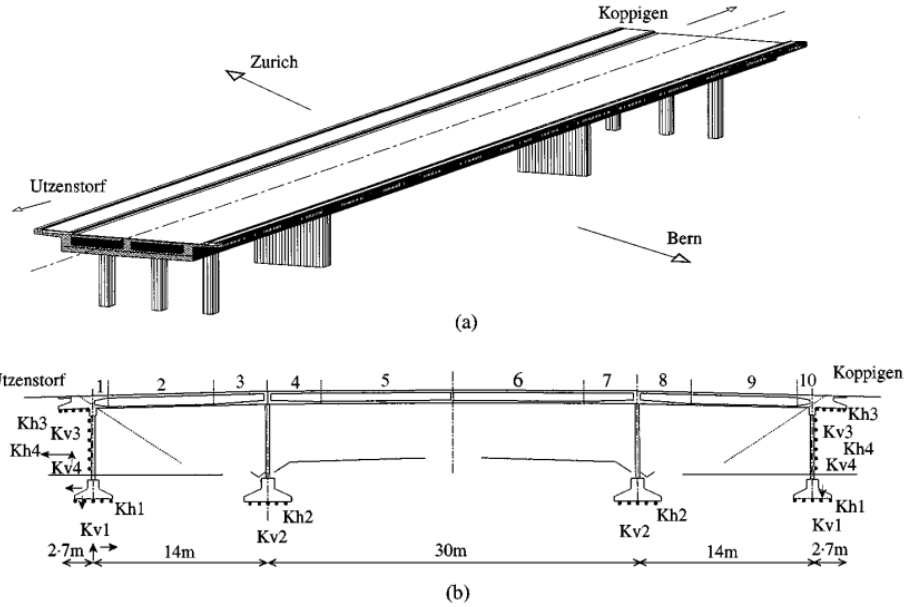

Figure 1. Z24 Bridge: (a), Global view; (b) Elevation. (adapted from Wahab and De Roeck [6])

4.1 Results

Reynders and De Roeck [7] asserts that the loss of local stiffness in this area was caused by fractures appearing at the midspan in an area 6.6 m from the Koppigen pier because of the forced settlements. Therefore, the Koppigen pier neighborhood should be identified as the true damage site. The spatial localization is shown in Fig. 2. For negative indices, the color is set to green; for positive indices, it ranges from green ($DI_i = 0$) to red (maximum DI_i). Other colors are assigned in proportion to the proposed deterioration index magnitude (e.g., yellow, orange, etc.). The greatest deterioration indices in red (Fig. 2) demonstrate that the approach can locate the damage in the actual spot in all four damage situations.

Sensors located to the right of the Utzenstorf pier (about 5 meters towards Koppingen) show low positive indices (as indicated by the weak positive indexes in scenarios I, II, and III in Fig. 2(a-c)). This settlement might have bent the bridge span along that pier and caused minor cracks in that area. The assessment of the condition of elevating the pier to its original height followed by foundation tilt supports this theory (i.e., scenario IV). Such lifting tends to close fractures on the bridge span, increasing stiffness to the section in which the technique no more identifies positive index in that area. This event demonstrates the great potential of the methodology in recognizing structural reinforcement.

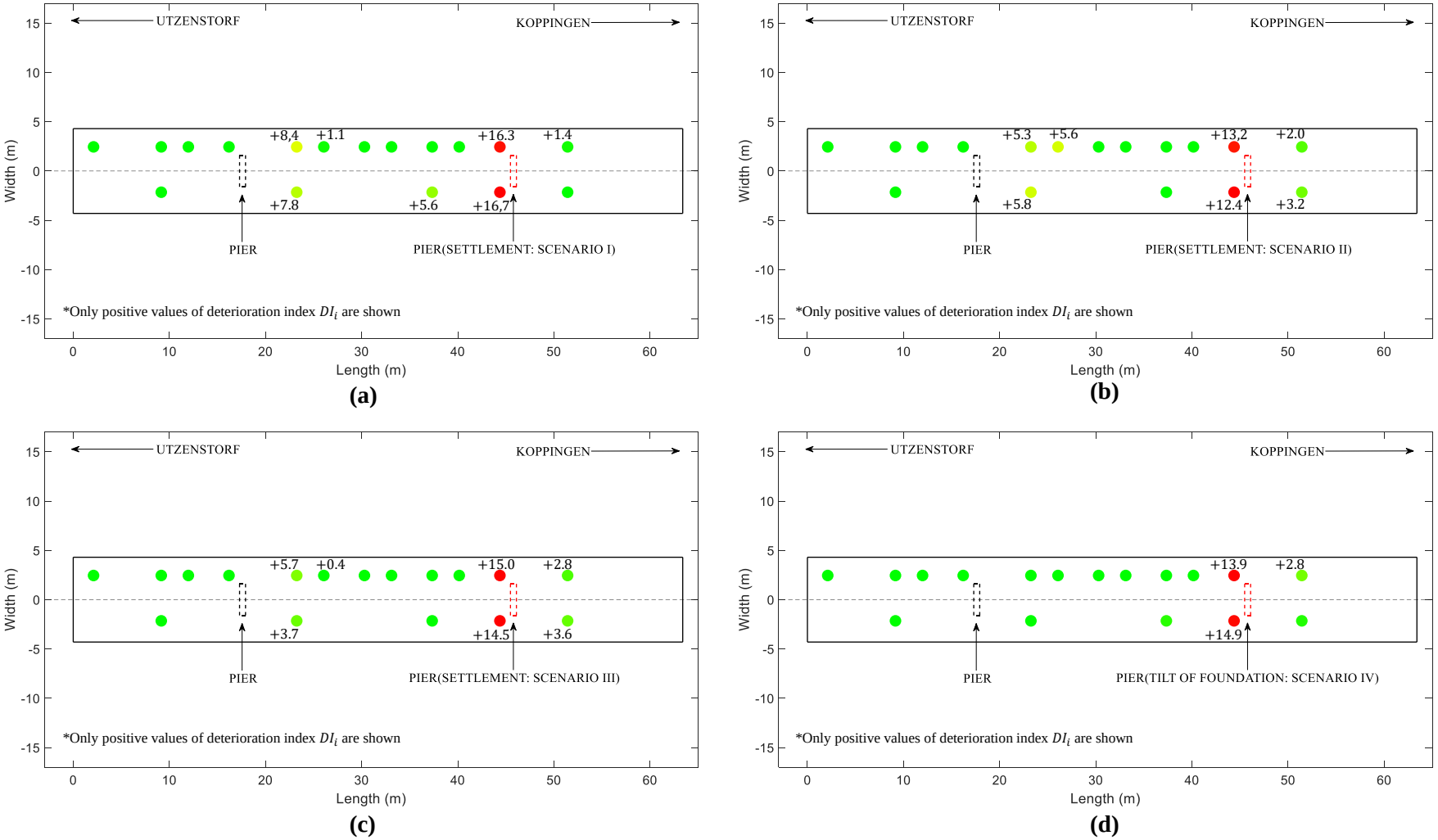

Figure 2. Z24 bridge - Spatial visualization: (a) Scenario I; (b) Scenario II; (c) Scenario III; (d) Scenario IV.

5 Conclusions

The automated technique produced promising outcomes on locating deterioration in all scenarios tested for the Z24 bridge. The employed feature selection (FS) strategy demonstrated to be an effective noise filtering procedure. Changes in ambient factors (e.g., wind and temperature) were noticed in the full-scale bridge experiment and did not interfere with the final damage localization assessment. This robustness is associated to the unsupervised FS to some degree. During a specific scenario, the approach identified structural reinforcement based on stiffness gain due to crack closure. This was directly reflected in the reduction in damage indexes in the reinforced zone. Hence, this paper proposes a significant accomplishment in validating the features' sensitivity to structural deterioration.

Acknowledgements. This study was financed by CNPq (Conselho Nacional de Desenvolvimento Científico e Tecnológico - grant 304329/2019-3) and FAPEMIG (Fundação de Amparo à Pesquisa do Estado de Minas Gerais – Grant PPM-00001-18).

Authorship statement. The authors hereby confirm that they are the sole liable persons responsible for the authorship of this work, and that all material that has been herein included as part of the present paper is either the property (and authorship) of the authors, or has the permission of the owners to be included here.

References

[1] S. Beskhyroun, T. Oshima, S. Mikami, and Y. Tsubota, “Structural damage identification algorithm based on changes in power spectral density”. *Journal of Applied Mechanics*, vol. 8, n. 1, pp. 73–84, 2005.
[2] V. H. M. Alves and A. A. Cury, “A fast and efficient feature extraction methodology for structural damage localization based on raw acceleration measurements”. *Structural Control and Health Monitoring*, vol. 28, n. 7, 2021.
[3] G. Zhang, R. S. Harichandran, P. Ramuhalli, “Application of noise cancelling and damage detection algorithms in NDE of concrete bridge decks using impact signals”. *Journal of Nondestructive Evaluation*, vol. 30, n. 4, pp. 259-272, 2011.
[4] M. Sahidullah and G. Saha. “Design, analysis and experimental evaluation of block based transformation in MFCC computation for speaker recognition”. *Speech Communication*, vol. 54, n. 4, pp. 543–565, 2012.
[5] G. Roffo, S. Melzi, U. Castellani, A. Vinciarelli and M. Cristani, "Infinite Feature Selection: A Graph-based Feature Filtering Approach," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 43, no. 12, pp. 4396-4410, 2021.
[6] M. M. A. Wahab, and G. De Roeck, “Damage detection in bridges using modal curvatures: application to a real damage scenario”. *Journal of sound and vibration*, vol. 226, n. 2, pp. 217–235, 1999.
[7] E. Reynders, and G. De Roeck, “A local flexibility method for vibration-based damage localization and quantification”. *Journal of Sound and Vibration*, vol. 329, n. 12, pp. 2367–2383, 2010.