HAL open science

Adaptive Detection of Structural Changes Based on Unsupervised Learning and Moving Time-Windows

João Santos, Luís Calado, André Orcesi, Christian Crémona

▶ To cite this version:

João Santos, Luís Calado, André Orcesi, Christian Crémona. Adaptive Detection of Structural Changes Based on Unsupervised Learning and Moving Time-Windows. EWSHM - 7th European Workshop on Structural Health Monitoring, IFFSTTAR, Inria, Université de Nantes, Jul 2014, Nantes, France. hal-01021195

HAL Id: hal-01021195
https://inria.hal.science/hal-01021195

Submitted on 9 Jul 2014

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Adaptive detection of structural changes based on unsupervised learning and moving time-windows

João Santos[1], Luís Calado[2], André D. Orcesi[3], Christian Crémona[4]

[1] *LNEC, Av. Brasil 101, 1700-066, Lisboa, Portugal.*
[2] *University of Lisbon, IST, Av. Rovisco Pais, 1049-001 Lisboa.*
[3] *IFSTTAR, 14-20 Boulevard Newton, Champs sur Marne, F-77447 Marne la Vallée 2, France.*
[4] *CÉREMA, 110 rue de Paris 77171 Sourdun, France*

[1] josantos@lnec.pt, [2] calado@ist.utl.pt, [3] orcesi@lnec.pt, [4] christian.cremona@cerema.fr

ABSTRACT

The present paper addresses data-driven structural health monitoring to propose a real-time strategy for adaptive structural assessment. The adaptive character is achieved using unsupervised discrimination machine-learning methods, widely known as clustering algorithms. Real-time capability is based on the definition of symbolic data, which allow describing large amounts of information without loss of related information. The efficiency of the proposed methodology is illustrated using an experimental case study in which structural changes were imposed to a suspended bridge during an extensive rehabilitation program.

KEYWORDS: *cluster analysis; symbolic data; suspended bridge; adaptive assessment.*

INTRODUCTION

Structural health monitoring (SHM) is generally defined as the process of implementing a damage identification strategy for any engineering structure, infrastructure or system [1]. Damage occurrences, which can be defined as changes affecting present or future structural performance may have significant human and social impact. In this context, an effective SHM strategy should be able to provide real-time structural feedback and detect any abnormal structural behaviour before damage occurrences take place and affect structural performance.

Structural health monitoring techniques may follow inverse or forward approaches [2] for identifying structural changes. Inverse approaches are frequently named model updating and consist of fitting analytical or numerical responses to experimental data with the objective of inferring structural quantities that cannot be directly measured [2]. Forward SHM approaches do not require the development of models [2]. Instead, they rely on data mining techniques for extracting sensitive information from time-series acquired on site [3]. Their computational simplicity makes them the most suitable candidates for carrying out real-time SHM in large-scale structures [4].

Feature extraction and discrimination are common steps to all data-driven SHM strategies [1], [3]. A feature can be defined as a type of information, extracted from time-series acquired on site, which may highlight structural changes by being correlated with them [2]. Its extraction is generally necessary since raw data is not informative about the occurrence of structural changes. Widely used examples of SHM features are modal quantities [5], autoregressive models [6], principal components [7] or symbolic data [7], [8]. The latter was only recently applied to SHM and proved particularly effective in fusing multi-sensor data without loss of damage-related information.

Feature discrimination consists of classifying the extracted features as related to monitored structural behaviours (known or unknown). This step is generally based on training machine-learning algorithms with baseline references of extracted features, where the structure is assumed to be undamaged and unchanged. New features are then analysed, by the pre-trained machine-learning algorithms, to assess if the structure's behaviour remains unchanged. Statistical

Copyright © Inria (2014)

process control [9], multi-layer perceptron neural networks [6] or support vector machines (Kim, Chong, Chon & Kim, 2013) are among the most common algorithms for discrimination strategies. Cluster analysis can be seen as an alternative approach since it does not require a prior baseline to perform feature discrimination, thus being more flexible and adaptable. Even though this type of unsupervised analysis has been reported as effective [11], its computational complexity has discouraged its use in SHM of civil structures. To circumvent these disadvantages, symbolic data has been coupled to cluster analysis, providing baseline-free classification with sufficient computational efficiency [7], [8].

SHM discrimination methodologies have essentially been reported, in the literature, as off-line (posterior) analyses. Instead of classifying structural behaviour upon detection of the first discrepant feature, these strategies have been mostly applied to data sets composed of large amounts of data, acquired before and after the occurrence of structural changes. The identification of structural changes is then performed by analysing different densities of discrepant features, such as outliers, in similar monitoring periods.

The objective of this paper is to define an adaptive strategy for detection of structural changes upon their occurrence. To test and validate the proposed concepts, data acquired from the suspended Samora Machel Bridge during retrofitting works, in which structural changes were imposed to the structure, is used. This study is divided into three main sections. After this introduction, section 1 describes the suspended bridge, the monitoring program, the acquired data and a brief preliminary analysis. Section 2 describes the data-driven techniques and shows their usefulness when analysing acquired data. In section 3, the real-time strategy is presented and tested in the Samora Machel Bridge data. Afterwards, the main conclusions from this analysis are drawn.

1 CASE STUDY – THE SAMORA MACHEL BRIDGE

The five-span suspended Samora Machel Bridge is located in the province of Tete, Mozambique. The bridge has four concrete towers, a deck 720m long, two main cables and 144 diagonal hangers (see Figure 1a). The cables have a nominal diameter of 170mm and the hangers are composed of 63 steel wires, with 5mm diameter each. Suspended by each pair of hangers are 72 precast transverse beams with variable cross-sections (Figure 1b,c). Simply supported on each pair of transverse beams are 71 post-tensioned deck modules shown in Figure 1b,c. Each module is composed of nine precast concrete longitudinal beams of variable depth.

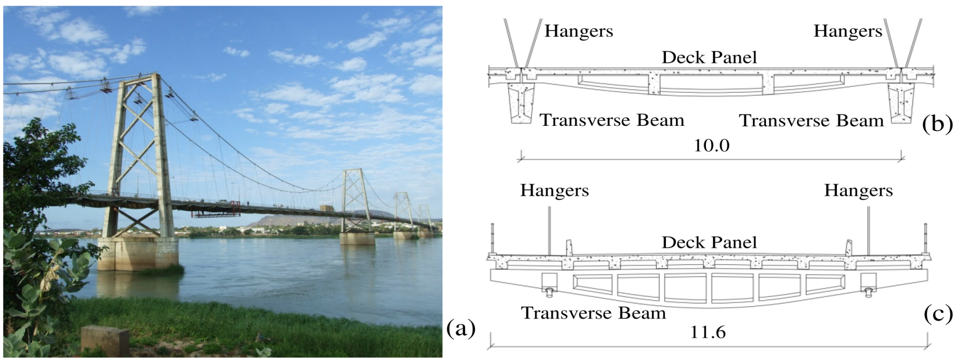

Figure 1 - The Samora Machel Bridge: (a), perspective view (b) longitudinal section, (c) transverse section.

An extensive refurbishment program was conducted on the Samora Machel Bridge from 2009 to 2012. Some of the most important refurbishment tasks included the generalized repairing / replacement of concrete, replacement of suspension cables, hangers and supports as well as the installation of stabilization systems in each of the seventy-two transverse beams. The latter consisted of the greatest change to the original structural system of the bridge and was designed to reduce the range of rotations generated on the transverse beams along their longitudinal axes, when heavy trucks and medium-weight off-road vehicles cross the bridge. These rotations were observed to be the main cause for damage, not only in transverse beams but also in the adjacent deck modules, supports and anchoring devices. To stabilize each transverse beam, two 140mm diameter steel tubes (Figure 2a,b) were installed to connect each transverse beam to the adjacent deck panel. Both connections are pinned and anchored to concrete with metal plates (Figure 2d).

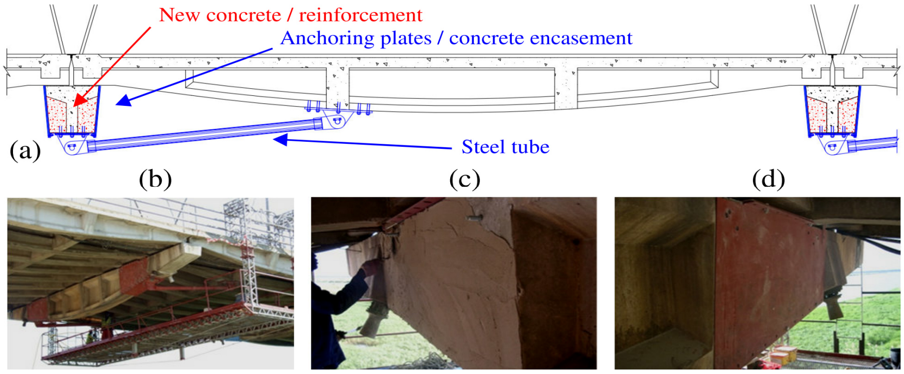

Figure 2 - Stabilization of the transverse beams: (a) longitudinal section, (b) overall view, (c) placement of new concrete, (d) anchoring plates and concrete encasement.

The Samora Machel Bridge SHM was monitored with the objective of assessing the effectiveness of the transverse beams stabilization systems. For this purpose, a transverse beam considered representative of the structural degradation was monitored before and after the stabilization procedure described in the previous section. Data was obtained from two precision tilt-meters fixed to one transverse beam near the tubes' anchorages. The tilt-meters fixed near the upstream and the downstream plates are named herein TM1 and TM2. To achieve these objectives, monitoring of the structural response was conducted at 10 Hz under operational loading, or regular traffic, in three phases:

1. before retrofitting of any beam (phase 1);
2. after stabilization of the instrumented beam (phase 2);
3. after stabilization of the two beams adjacent to the one monitored (phase 3).

The data obtained from the monitoring campaign consisted of the rotations' ranges, registered at 10-second intervals since only rotations generated by heavy trucks loads and medium-weight off-road vehicles were of interest to analyse the efficiency of the stabilizing system.

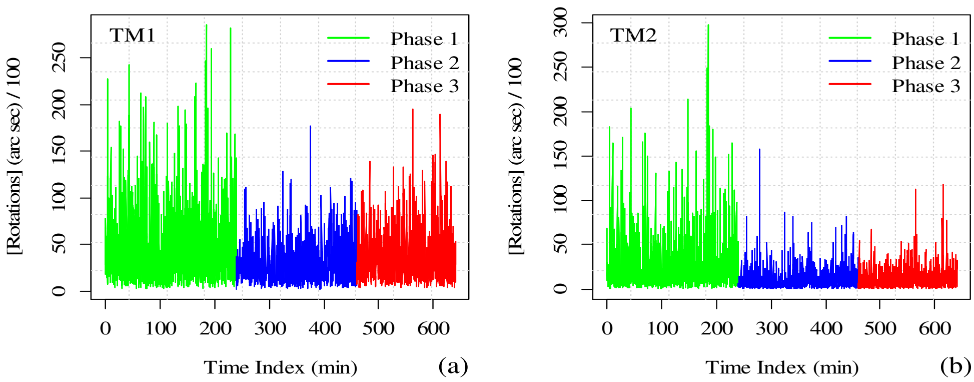

Figure 3 - Time series of rotation ranges acquired from the tilt-meter: (a) TM1, (b) TM2.

The time-series of Figure 3a,b reveal rotations ranges higher in phase 1 than in phases 2 and 3. In the first phase, the measured ranges were reflecting the rotations exhibited by the unrestrained transverse beams. After the installation of the stabilization system, the transverse beams became rigidly connected to the adjacent deck panels and their rotation ranges are much smaller, reflecting only the global response (deflection) of the entire superstructure.

2 FORWARD STRUCTURAL HEALTH MONITORING

2.1 Feature extraction

Feature extraction is addressed in this paper using symbolic data objects, which are less voluminous and more generic types of information when compared to classical data such as SHM time-series measurements [8]. While classical data mining focuses on finding groups or patterns in individual measurements, symbolic data deals with concepts, such as time-intervals (Santos, Crémona, Orcesi & Silveira, 2013), structural elements or structural systems, described by statistical quantities such as interquartile intervals or histograms [8]. In this paper, interquartile intervals are chosen over histograms since they lead to a greater degree of data fusion [12] while exhibiting similar ability to describe SHM data [8]. A symbolic data object, T, described by interquartile intervals of multiple sensors, can be defined as

$$T = \left[\left(\mathrm{T}_{\inf}^{(1)};\mathrm{T}_{\sup}^{(1)}\right) \ \cdots \ \left(\mathrm{T}_{\inf}^{(i)};\mathrm{T}_{\sup}^{(i)}\right) \ \cdots \ \left(\mathrm{T}_{\inf}^{(p)};\mathrm{T}_{\sup}^{(p)}\right)\right] \tag{1}$$

where p is the number of sensors describing the symbolic object. In the present work, p=2 and thus symbolic objects exhibit a rectangular shape in a bi-dimensional space, as shown in Figure 4a-c, where the limits of interquartile intervals define the four sides of the rectangles. For p=3, the symbolic objects can be represented by parallelepiped solid shapes, and for p>3 by hyper-volumes of dimension p.

The usefulness of symbolic data objects in the realm of SHM heavily relies on the definition of techniques that allow the analysis of this type of data using conventional machine-learning algorithms [12]. Among these techniques, symbolic distances have assumed particular importance since they enable cluster analysis to be applied to large data sets, such as those generated in the framework of SHM. Symbolic distances can take a variety of forms and some applications might require specific ones. In the present work, the normalized Euclidean Ichino-Yaguchi distance [12] is used since it exhibits a high sensitivity to structural changes, when compared to other measures found in literature [7]. To present its theoretical background, two symbolic objects $\mathrm{T_i}$ and $\mathrm{T_j}$, obtained from a data set of $\mathrm{s} = 1, \ldots, \mathrm{N}$ objects are defined. The normalized Euclidean Ichino-Yaguchi distance [12] between these objects is defined as

$$\mathrm{d_{ij}} = \left(\frac{1}{\mathrm{p}}\sum_{\mathrm{r}=1}^{\mathrm{p}}\frac{1}{\left|\mathrm{Y_r}\right|}\left[\varphi_{\mathrm{r}}\left(\mathrm{T_i},\mathrm{T_j}\right)\right]^2\right)^{1/2} \tag{2}$$

where φ_r is given by

$$\varphi_{\mathrm{r}}\left(\mathrm{T_i},\mathrm{T_j}\right) = \left|\mathrm{T_i^{(r)}} \oplus \mathrm{T_j^{(r)}}\right| - \left|\mathrm{T_i^{(r)}} \otimes \mathrm{T_j^{(r)}}\right| + \gamma\left(2\left|\mathrm{T_i^{(r)}} \otimes \mathrm{T_j^{(r)}}\right| - \left|\mathrm{T_i^{(r)}}\right| - \left|\mathrm{T_j^{(r)}}\right|\right) \tag{3}$$

and γ a pre-specified constant ranging from 0 to 0.5. The Cartesian join ⊕ and meet ⊗, as well as the norm |[…]| and the normalizing term $|\mathrm{Y_r}|$, are defined by

$$\mathrm{T_i^{(r)}} \oplus \mathrm{T_j^{(r)}} = \left[\min\left(\mathrm{T_{i,inf}^{(r)}},\mathrm{T_{j,inf}^{(r)}}\right), \max\left(\mathrm{T_{i,sup}^{(r)}},\mathrm{T_{j,sup}^{(r)}}\right)\right] \tag{4}$$

$$\mathrm{T_i^{(r)}} \otimes \mathrm{T_j^{(r)}} = \left[\max\left(\mathrm{T_{i,inf}^{(r)}},\mathrm{T_{j,inf}^{(r)}}\right), \min\left(\mathrm{T_{i,sup}^{(r)}},\mathrm{T_{j,sup}^{(r)}}\right)\right] \tag{5}$$

$$\left|\mathrm{T}\right| = \left|\left[\mathrm{T_{inf}},\mathrm{T_{sup}}\right]\right| = \mathrm{T_{sup}} - \mathrm{T_{inf}} \tag{6}$$

$$\left|\mathrm{Y_r}\right| = \left|\max_{\mathrm{s}}\left(\mathrm{T_{sup}^{(r)}}\right) - \min_{\mathrm{s}}\left(\mathrm{T_{inf}^{(r)}}\right)\right| \tag{7}$$

For the data acquired from the Samora Machel Bridge SHM, a symmetric Ichino-Yaguchi distance matrix of size 131x131, obtained according to Eq.(3) was defined using symbolic objects of 5 minutes. This matrix is presented in Figure 4d, where brighter colours stand for closer data objects while darker colours represent larger distances. This figure suggests the existence of two distinct groups in the data set, corresponding to two distinct structural behaviours. The first one being associated with data acquired during phase 1 and the second one comprising data acquired during phases 2 and 3.

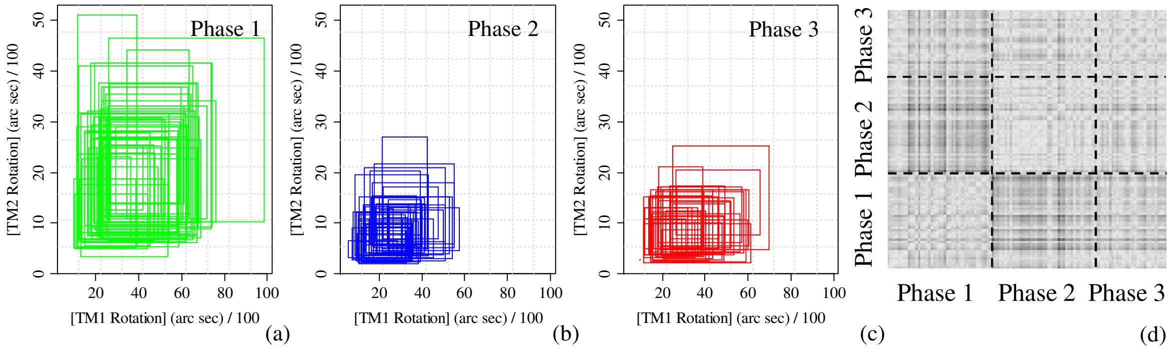

Figure 4 - Symbolic data objects defined for phases (c) 1, (d) 2 and (e) 3; symbolic distance matrix.

2.2 Feature discrimination

Symbolic objects and distance measures can help perform data analysis and suggest the existence of different structural behaviours within a data set, as shown in the previous section. However, the development of real-time SHM strategies requires the use of machine-learning algorithms which can analyse these features and help decide whether one or more distinct structural behaviours is being observed. Hence, feature discrimination is addressed herein using cluster analysis, which appears to be the family of most promising machine-learning methods for conducting adaptive SHM. The goal of a clustering process is to divide a data set into groups, which must be as compact and separate as possible [7]. This aim can be mathematically posed as the attempt to minimize the overall within-cluster distance $W(P_k)$ which, consequently, maximizes the overall between-cluster distance [8]. Considering a given partition containing K clusters, the overall within-cluster distance $W(P_k)$ can generally be defined as [13]

$$W(P_k)=\frac{1}{2}\sum_{k=1}^{K}\sum_{C(i)=k}\sum_{C(j)=k}d_{ij} \tag{8}$$

where C(i) is an allocation rule which assigns element i to cluster k, based on the distance measure. The total distance of data can be defined as in Eq.(10), where N is the number of symbolic objects defined for the analysed data set. The overall between-cluster distance $B(P_k)$ is simply obtained by subtracting the other two distances (Eq.(11)).

$$T=\frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N}d_{ij} \tag{9}$$

After partitioning the data set into K clusters, the data allocated to each of them can be simply described by a prototype, which generally consists of an object of the same type as those being clustered. The location and features of the prototype are sufficient to represent the structure of clustered data. The location of each prototype is obtained herein by computing the centroid of each cluster's members.

Several families of clustering methods can be found in the literature, from which the most well-known and used are the combinatorial methods. This type of cluster analysis is iterative in nature and requires the input of an initial set of K<N cluster prototypes and their centroids, generally randomly defined [13]. Afterwards, each iteration starts by allocating data objects to the closest cluster prototype. After the allocation, the new set of prototypes is defined by computing the centroids of the new clusters (defined through the allocation). This procedure is then repeated until a general criterion, which is usually a function of the overall within-cluster distance, is met. The most used clustering method is based on greedy descent and is named k-means. This algorithm uses as loss function the squared overall within-cluster distance. A generalization of this method, named dynamic cloud, was proposed by Diday for application to symbolic data sets [8] and is used herein.

The structural engineering problem posed for the presented case study consists in determining which of the two monitored structural changes, generated by the installation of the retrofitting systems (in the monitored beam and in the adjacent ones), were significant in changing the monitored structural behavior. Since the number of significant structural changes is not known in advance, this type of problems must be tackled by allocating data objects to all possible cluster partitions. For the case study used herein, the analysis of cluster partitions comprising 2 and 3 structural behaviors, respectively, suffices since only two structural changes were imposed to the structural system.

Cluster analysis performed on the data set acquired from the Samora Machel SHM was based on the dynamic cloud algorithm applied to the Ichino-Yaguchi distance matrix, shown in Figure 4d and obtained from 5-minute symbolic objects. Two and three clusters were, respectively, defined in the initialization phases of these analyses and their outputs, presented as object allocations, are shown in Figure 5.

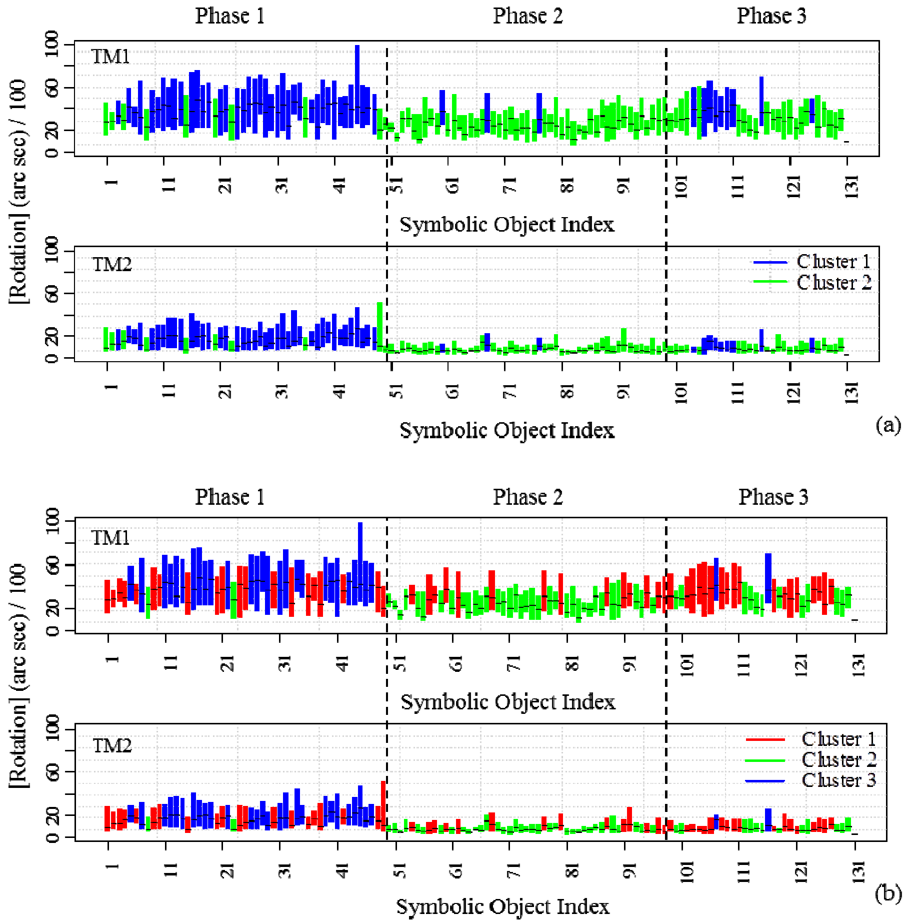

Figure 5 - Data objects allocation to partitions comprising (a) 2 clusters, (b) 3 clusters.

When studying the possibility of one significant structural change, by assuming two clusters (Figure 5a), it was observed that most data objects allocated to cluster 1 were acquired during phase 1 whereas objects assigned to cluster 2 were mostly obtained during phases 2 and 3. These results strongly suggest that the most important structural change observed in the Samora Machel Bridge took place between phases 1 and 2, thus generating two distinct structural behaviors: the first one observed during phase 1 and the second one during phases 2 and 3. If two significant structural changes have occurred in the Samora Machel bridge, a cluster analysis comprising three clusters should allocate the objects belonging to each experimental phase into each distinct cluster. However, the allocations shown in Figure 5b indicate otherwise. Even though cluster 3 is almost completely composed of objects acquired during phase 1, cluster 2 is assigned with objects from phases 2 and 3 and cluster 3 is scattered among the three phases. Finally, the possibility of having observed unchanged structural behavior throughout the SHM program of the Samora Machel Bridge must also be considered. Provided that cluster analysis cannot analyse single-cluster partitions, this possibility must be considered by analyzing the remaining cluster partitions. A data set obtained during an unchanged structural behavior would generate random allocation of data objects among any number of considered clusters. The results discussed in the two previous paragraphs clearly show that this is not the case for the present case study, thus leading to the conclusion that structural changes were observed on the Samora Machel bridge.

Clustering algorithms are able to allocate data objects to any amount of predefined structural behaviours (each represented by one cluster), irrespective of whether this number of distinct behaviours has really been observed in the structural system. Even though an analysis of object allocations, such as the one performed in the previous paragraph, is able to suggest the number of structural behaviours (and structural changes), its conclusions might be subjective.

To obtain a quantitative and objective estimation of the most truthful number of observed behaviours (clusters) existent in a data set, an evaluation known as cluster validity [7] must be conducted on the allocations provided by cluster analyses. Cluster validity consists of computing indices for all possible cluster partitions [13]. Each validity index provides a quantitative measure of cluster compactness and separation, regardless of the number of clusters present in the evaluated partition. The most truthful number of clusters (or structural behaviours) is then assessed by comparing the indices obtained from each of the evaluated partitions. In the present work the global Silhoutte index was used since it achieved best performance in previous works [7], and allowed concluding that two clusters (and distinct structural behaviours) are present in the analysed data set.

3 ADAPTIVE REAL-TIME ASSESSMENT

An adaptive strategy capable of conducting automatic structural assessment in real-time is proposed in this section. Its real-time capabilities are based on the definition of fixed-length moving time-windows, each of which is defined upon the acquisition of new data. Its ability to perform baseline-free assessment is achieved through the use of cluster analysis (explained in the previous section), followed by an outlier analysis.

Conventional methods proposed in SHM-related literature cannot generally perform real-time assessment due to their difficulty in providing reliable conclusions upon detection of the first discrepant value (outlier). This is due to the high probability of obtaining false detections when using features that normally describe single-sensor data or exhibit small generalization capacity. Herein, the false detections are avoided by using a single-valued quantity based on cluster analysis and which describes the entire data set analysed within each time-window: the distance between clusters, B(Pk). At each time-window defined over time, this index is computed along with its confidence interval defined in the previous window. This procedure results in an on-going process with an adaptive confidence boundary capable of outlining structural changes in real-time.

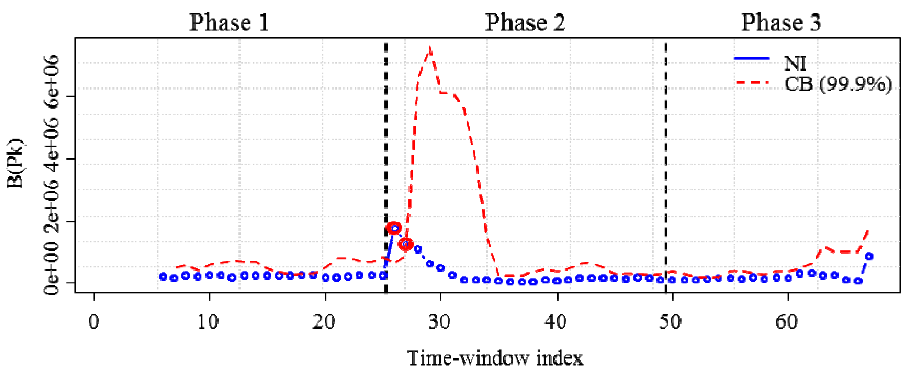

Figure 6. Real-time structural assessment output using a 10-minute time-resolution.

This strategy was applied to the Samora Machel Bridge data set using time-windows of one hour and 10-minute symbolic objects. Each time-window was defined 10 minutes after the previous one, thus simulating a real-time assessment procedure with a time resolution of 10 minutes. The output obtained from this procedure consists of the $B(P_k)$ values along with their adaptive confidence boundary, as show in Figure 6. In this plot, the $B(P_k)$ exceeds the adaptive boundary in the first two time-windows containing data from phases 1 and 2 ($B(P_k)$) (values outlined with a red circle), thus detecting the structural change 10 minutes after its occurrence. This time resolution evidences the effectiveness of the proposed strategy.

CONCLUSIONS

This paper presents a data-driven SHM methodology for conducting adaptive real-time structural assessment based on cluster analysis and symbolic data. Its ability to detect structural changes in real-time without resorting to baselines were tested using data acquired from the Samora Machel Bridge. It was observed that symbolic objects based on interquartile intervals are able to accurately generalize the information present in data, while performing significant data fusion. Furthermore, it was observed that the Ichino-Yaguchi distance is sensitive to structural changes concealed in symbolic data sets. Concerning cluster analysis, it was observed that the dynamic cloud method is effective in allocating data objects.

The moving time-windows strategy proposed herein proved capable of describing multi-sensor data and point out structural changes imposed in the experimental study only 10 minutes after their occurrence. In addition, it was concluded that the variations exhibited by the proposed single-valued quantity, $B(P_k)$, over time reflect structural changes measured across multiple sensors and that the proposed discrimination method based on cluster analysis and adaptive confidence boundaries allows detecting structural changes upon their occurrence, while avoiding false detections.

REFERENCES

[1] C. R. Farrar and K. Worden, "An introduction to structural health monitoring.," *Philos. Trans. A. Math. Phys. Eng. Sci.*, vol. 365, no. 1851, pp. 303–315, Feb. 2007.

[2] S. D. Glaser and A. Tolman, "Sense of Sensing: From Data to Informed Decisions for the Built Environment," *J. Infrastruct. Syst.*, vol. 14, no. 1, pp. 4–14, 2008.

[3] H. Sohn, C. Farrar, F. M. Hemez, D. D. Shunk, D. W. Stinemates, and B. R. Nadler, "A Review of Structural Health Monitoring Literature: 1996 – 2001," Los Alamos, USA, 2004.

[4] D. Posenato, "Model-Free Data Interpretation for Continuous Monitoring of Complex Structures," École Polytechnique Fédérale de Lausanne, 2009.

[5] A. Alvandi and C. Crémona, "Assessment of vibration-based damage identification techniques," *J. Sound Vib.*, vol. 292, no. 1–2, pp. 179–202, Apr. 2006.

[6] O. R. Lautour and P. Omenzetter, "Damage classification and estimation in experimental structures using time series analysis and pattern recognition," *Mech. Syst. Signal Process.*, vol. 24, no. 5, pp. 1556–1569, Jul. 2010.

[7] J. Santos, C. Crémona, A. D. Orcesi, and P. Silveira, "Multivariate Statistical Analysis for Early Damage Detection," *Eng. Struct.*, vol. 56, pp. 273–285, 2013.

[8] A. Cury, C. Crémona, and E. Diday, "Application of symbolic data analysis for structural modification assessment," *Eng. Struct.*, vol. 32, no. 3, pp. 762–775, 2010.

[9] D. Posenato, P. Kripakaran, D. Inaudi, and I. F. C. Smith, "Methodologies for model-free data interpretation of civil engineering structures," *Comput. Struct.*, vol. 88, no. 7–8, pp. 467–482, Apr. 2010.

[10] Y. Kim, J. W. Chong, K. H. Chon, and J. Kim, "Wavelet-based AR–SVM for health monitoring of smart structures," *Smart Mater. Struct.*, vol. 22, no. 1, pp. 3–15, Jan. 2013.

[11] H. Sohn and S. D. Kim, "Reference-Free Damage Classification Based on Cluster Analysis," *Comput. Civ. Infrastruct. Eng.*, vol. 23, no. 5, pp. 324–338, 2008.

[12] L. Billard and E. Diday, *Symbolic Data Analysis*, vol. 52, no. 2. Chichester, UK: John Wiley and Sons, 2006, p. 321.

[13] S. Theodoridis and K. Koutroumbas, *Pattern Recognition*, 4th ed. London, UK: Elsevier, 2009, p. 961.