REVISTA INTERDISCIPLINAR DE PESQUISA EM ENGENHARIA

CILAMCE 2016

XXXVII IBERIAN LATIN AMERICAN CONGRESS ON COMPUTATIONAL METHODS IN ENGINEERING

BRASÍLIA - DF - BRAZIL

DAMAGE IDENTIFICATION THROUGH THE USE OF HIGH-ORDER STATISTICS

Alan Torres

Alexandre Cury

alan.torres@engenharia.ufjf.br

alexandre.cury@engenharia.ufjf.br

Federal University of Juiz de Fora

Rua José Lourenço Kelmer, 36036-330, Juiz de Fora, Minas Gerais, Brazil

Abstract. *Structural Health Monitoring is based on the development of reliable and robust indicators capable to detect, locate, quantify and predict damage. Studies related to damage detection in civil engineering structures have a noticeable interest for researchers in this area. Indeed, the detection of structural changes likely to become critical can avoid the occurrence of major dysfunctions associated with social, economic and environmental consequences. Recently, many researchers have focused on dynamic assessment as part of structural diagnosis. Most of the studied techniques are based on time or frequency domain analyses to extract compressed information from modal characteristics or based on indicators built from these parameters. This work has as its main interest the use of high-order statistics (HOS) coupled with clustering techniques i.e. the k-means algorithm to detect structural modification (damage). The approach is applied directly to dynamic measurements (accelerations) obtained on site. In order to attest the efficiency of the proposed methodology, two investigations are carried out: a numerical model of a simply supported beam and a real case railway bridge, in France. It is shown that HOS coupled with clustering methods is able to distinguish structural conditions with adequate rates.*

Keywords: *Damage detection, High-Order Statistics, Clustering methods, Raw Data.*

1 INTRODUCTION

Structural Health Monitoring (SHM) is of great importance to Civil Engineering, once it allows the detection of modifications in the physical properties of structures. SHM permits, if possible and/or necessary, the use of recovery procedures in suitable time. These procedures are often based on evidence collected from tests performed on the structure that would ideally allow 'detecting, locating, quantifying and even predicting damages'. Recently, researches have focused on the dynamic evaluation as part of structural diagnoses (Cury et al. 2012 & Alves, 2015), by the extraction of modal parameters or data built from these parameters that, so far, have provided promising results. However, some problems remain unsolved, such as the sensitivity of the damage identification methods, their need of a reference state and their reliability when it comes to the detection of false alarms (Alves, 2015 & Cury et al. 2010).

Traditional methods of damage detection and health monitoring are often based on the variation of structural vibration characteristics, i.e. natural frequencies, damping ratios and mode shapes. These modal parameters are directly affected by changes in the physical properties of the structure including its mass and stiffness. Nevertheless, modal parameters identification is a sort of filtering process, leading to a loss of information compared to the raw data. This compression process can erase any small changes due to a structural modification. In turn, using raw dynamic measurements (especially if high sampling frequencies are used) leads to the storage of large set of data. However, several damage detection methods exist in the literature based on signature principles, but they usually fail when making them practical (Santos et al., 2013). In this sense, despite the current computers' processing power, the necessary computational effort to manipulate large data sets remains a problem. Furthermore, and this is certainly the major drawback when using modal parameters, is that modal components are essentially describing an equivalent linear behavior, a feature which may be not exact for the analysis of specific degraded systems.

In general, data acquisition campaigns in civil engineering structures gather thousands of accelerations values measured by several sensors. Consequently, analyzing all of these data directly may usually be time-consuming or even prohibitive. In this sense, transforming this massive quantity of data into a compact but also rich descriptive type of data becomes an attractive approach. In statistics, the term Higher-Order Statistics (HOS) refer to functions which use the third and higher powers of a sample, as opposed to more conventional techniques of lower-order statistics, which use constant, linear, and quadratic terms (zeroth, first, and second powers). The third and higher moments, as used in the skewness and kurtosis, are examples of HOS, whereas the first and second moments i.e. arithmetic mean and variance are examples of low-order statistics.

In this paper, raw data obtained from dynamic tests (acceleration measurements) are transformed into a more compact arrangement by computing HOS for each sensor e.g. accelerometer used. Then, these quantities are applied to a clustering algorithm in order to discriminate different structural states or, in other words, to detect damage. This paper is based on the conjecture that the HOSs are sensitive to damage, meaning that they can provide information regarding variations on the physical properties of structures that could indicate the existence of damage. Thus, the clustering algorithm would be able to eventually identify different structural scenarios.

The main objective of this study is to develop a methodology capable of detecting structural damage by clearly differentiating physical states that correspond to an "undamaged" configuration from a "damaged" configuration. It is important to emphasize that such a methodology is original, since it uses HOS coupled with clustering techniques. However, it must also be reminded that this is a rather complex problem, since the proposed methodology is applied directly to the raw data i.e. the raw dynamic measurements.

2 METHODOLOGY

This section presents the main concepts within the framework of this paper. First, a brief description of HOS is presented. Then, a short explanation about the k-means clustering algorithm is given (more details can be found in reference Madhulatha, 2012).

2.1 High-Order Statistics

As previously explained, HOS refer to functions which use the third and higher powers of a given sample. The HOS used in this paper are summarized in Table 1 where the variable "y_i" represents each sensor (accelerometer) measuring "n" acceleration values for each dynamic test.

Table 1. High order statistics used as damage sensitive characteristics. Adapted from (Farrar & Worden, 2013).

Statistic	Formula
Peak value	$y_{peak} = \max\lvert y_i \rvert$
Mean	$\bar{y} = \frac{1}{n}\sum_{i=0}^{n} y_i$
Mean Square	$\overline{y_{sq}} = \frac{1}{n}\sum_{i=0}^{n} y_i^2$
Root Mean Square (rms)	$rms = \sqrt{\frac{1}{n}\sum_{i=0}^{n} y_i^2}$
Variance	$\sigma^2 = \frac{1}{n}\sum_{i=0}^{n}\left(y_i - \bar{y}\right)^2$
Standard Deviation	$\sigma = \sqrt{\frac{1}{n}\sum_{i=0}^{n}\left(y_i - \bar{y}\right)^2}$
Skewness	$\gamma = \frac{\frac{1}{n}\sum_{i=0}^{n}\left(y_i - \bar{y}\right)^3}{\sigma^3}$

Kurtosis	$k = \dfrac{\frac{1}{n}\sum_{i=0}^{n}\left(y_i - \bar{y}\right)^4}{\sigma^4}$
Crest factor	$X_{CF} = \dfrac{y_{peak}}{rms}$
K-Factor	$X_K = \left(y_{peak}\right)\left(rms\right)$

2.2 k-means clustering algorithm

This clustering method is based on a generalization of the classical dynamic clusters method. Clustering of a dataset is the partition of that data into groups, named clusters, so that the data inside a cluster has the highest degree of similarity among all possible combinations. What defines this ‘similarity’ is the clustering algorithm. In this paper, the chosen algorithm - k-means - uses the spatial distance between a data point (a dynamic test) to the centroid of the cluster. The idea is to identify the best combination of data that produces the clusters with the highest degree of similarity, that is, the smallest data-centroid sum of distances. The k-means algorithm has several metrics to calculate those spatial distances. In this paper, the metrics used are the square Euclidean and the *cityblock*, shown in Equations 1 and 2, respectively:

$$d\left(E_i, E_j\right) = \sqrt{\sum_{k=1}^{n}\left(E_{ik} - E_{jk}\right)^2} \tag{1}$$

$$d\left(E_i, E_j\right) = \sum_{k=1}^{n}\left|E_{ik} - E_{jk}\right| \tag{2}$$

Once the statistics were extracted, the resulting datasets were analyzed through clustering algorithm k-means.

3 RESULTS

Before presenting the numerical and experimental applications explored in this paper, it must be kept in mind that the aforementioned clustering method was already applied to modal parameters (natural frequencies and mode shapes) obtaining very good results (Alves, 2015). Now, the authors want to further explore the potentialities of the proposed approach using uniquely raw data i.e. accelerations measured directly *in situ*.

The procedure conducted henceforth in this paper follows these steps:

1. Evaluate the HOS of each accelerometer as explained in section 2.1;
2. Use HOS (step 1) as inputs for the clustering technique (k-means).

Finally, it is important to emphasize that this entire procedure strongly depends on the quality of the input data. In this case, if accelerations measurements present any type of problem (bad sampling, missing data, incorrect measurement, etc.), the results obtained from the clustering methods will be compromised. Thus, it is imperative to assure, in first hand, that the data used in the analysis is adequate.

Two sets of data were used for the validation of the proposed methodology. In both cases, it was previously known which data corresponded to the 'undamaged' and 'damaged' states. Therefore, the goal of the k-means algorithm was to allocate the 'undamaged' tests into one cluster and the 'damaged' tests into another. Thus, it would show that this technique is able to distinguish between two groups related to two different physical structural conditions.

3.1 Numerical application – Simply supported beam

The first dataset corresponds to a numerical simulation of the vibration response of a 6-meter simply supported steel beam. The beam is discretized into 200 finite elements using Matlab. A random force is applied at 0,69m from the right support as shown in Figure 1. The beam has the following physical and geometric properties:

- Elastic modulus – 210 GPa;
- Volumetric mass – 7850 kg/m³;
- Cross-sectional area – $2{,}81\times10^{-3}$ m²;
- Moment of inertia – $2{,}81\times10^{-8}\ \mathrm{m}^4$.

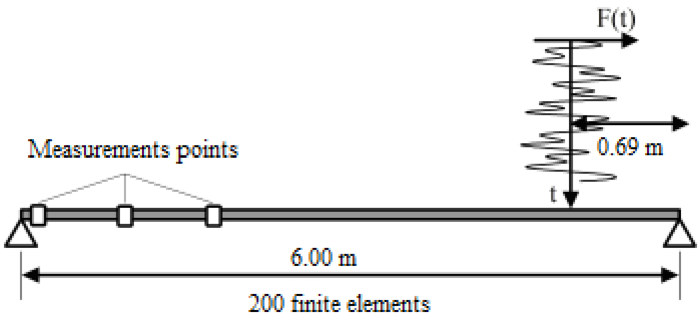

Figure 1 – Discretized beam.

Dynamic measurements are taken at 10 equidistant points of the beam during 100s. The sampling frequency is 0.01 Hz, which corresponds to 10.000 acceleration measurements per sensor.

Damage is simulated as the reduction of the elastic modulus of the beam in three progressive states, yielding three damage levels:

- Undamaged (D0): unaffected beam;
- Level 1 (D1): reduction of 20% in the elastic modulus of elements 96-105;
- Level 2 (D2): same as level 1 + reduction of 10% in the elastic modulus of elements 146-155.

Additionally, three levels of white noise (0, 5% and 10%) are added to the data.

The main objective here is to separate two different damage levels at a time (D0 from D1 and D0 from D2) for each noise level. Each dataset of the numerical simulations contains measurements from 10 accelerometers and, for each accelerometer, 10 HOSs are calculated (as shown in Table 1). Thus, each dataset yields a 20x100 matrix (20 tests by 100 HOS).

These sets are used as inputs to the k-means method and the results are presented in Tables 2 and 3.

Table 2. Percentages of correctly classified data within cluster (levels D0 and D1).

	No noise		5% noise		10% noise	
	sqeuclidian	*cityblock*	*sqeuclidian*	*cityblock*	*sqeuclidian*	*cityblock*
D0	80%	80%	80%	80%	80%	80%
D1	70%	70%	70%	70%	70%	70%

Table 3. Percentages of correctly classified data within cluster, (levels D0 and D2).

	No noise		5% noise		10% noise	
	sqeuclidian	*cityblock*	*sqeuclidian*	*cityblock*	*sqeuclidian*	*cityblock*
D0	80%	80%	80%	80%	80%	80%
D2	60%	60%	60%	60%	60%	60%

From Tables 2 and 3 it is possible to notice that the proposed methodology is rather capable of distinguishing the two damage levels, although the correct rates were lower for the second case. However, it is important to remark that this procedure is completely insensitive to noise, since all rates remained unchanged throughout the simulations.

3.2 Experimental application – TGV viaduct

The second application comprehends the measurements taken in the PK 075+317 viaduct in Southeast France, between Paris and Lyon, over which passes high-speed rails for TGV (*Train à Grande Vitesse*) trains. The viaduct is 17.5 meters wide (Figure 2).

Figure 2 – Side view of the viaduct.

This bridge was built in the early eighties; the increase of the operating speed of TGVs has moved the excitation frequency of the trains close to the first natural frequency of the bridge. This risk of resonance was furthermore increased by the uncertainties in the mass of the ballast disposed on the bridge. The first natural frequency was 5.86 Hz and the excitation frequency was around 4.0 Hz. The French railways SNCF considered that this difference was not enough and ballast recharging in connection with new operating speeds could reduce it even more. This is why SNCF set up a system of rods near the bearings tightened by torque wrench (Figure 3); this strengthening brought stiffness and increased the natural frequencies. In 2003, a strengthening intervention was scheduled and led to a change in natural frequencies.

Figure 3 – Tightening procedure.

In order to assess the efficiency of the tightening procedure, eight vertical accelerometers were installed under the bridge's deck, having the sampling frequency fixed at 4096 Hz. Twenty-eight dynamic tests were carried out i.e. 15 tests before and 13 after the procedure. Thereby, this dataset can be represented by a 28x80 matrix containing 10 HOS for each accelerometer. This dataset was inputted to the k-means algorithm and results are shown in Table 4.

Table 4. Percentages of correctly classified data within cluster.

	sqeuclidian	*cityblock*
Before	67%	67%
After	23%	23%

As shown in Table 4, the proposed methodology was not able to differentiate the two structural scenarios. Therefore, it was necessary to enhance the analysis. This was achieved by performing a series of combinations of the HOS, two at a time, for each accelerometer. The goal was to find the best combination i.e. the one that yields the smallest within cluster sum of distances. This criterion was adopted, since it is believed to provide the most accurate classification. Results are shown in Table 5.

Table 5. Percentages of correctly classified data within cluster.

	sqeuclidian	*cityblock*
Before	79%	79%
After	85%	100%

From Table 5, it is clear to observe a significant improvement in the classification rates. Moreover, for the *cityblock* metric, 100% of the tests after the tightening procedure were correctly classified into on cluster.

4 CONCLUSIONS

This paper introduced a novel approach based on the coupling of High-Order Statistics with the k-means clustering algorithm. The main goal was to discriminate different structural behaviors using only raw information for feature extraction.

In order to attest the robustness of the proposed approach, two applications were studied: the cases of a FEM beam model and of a real-case railway viaduct. For the first case, different damage and noise levels were simulated. It was noticed that although the procedure was insensitive to noise, it was not properly capable of distinguishing and separating the structural states. For the experimental application, the same methodology was applied, yielding poor results. Thus, the combination of the HOS, 2 at a time, proved itself rather suitable for the TGV, since it made possible for the k-means algorithm to identify two different states in the dataset, as shown in Table 5.

It is important to highlight the complexity of such an analysis, since it deals directly with raw data measurements. However, the authors are aware that results must be improved before using this methodology to untested structures.

ACKNOWLEDGEMENTS

The authors would like to thank UFJF (Universidade Federal de Juiz de Fora - Federal University of Juiz de Fora), CAPES (Coordenação de Aperfeiçoamento de Pessoal de Nível Superior), CNPq (Conselho Nacional de Desenvolvimento Científico e Tecnológico - "National Council of Technological and Scientific Development") and FAPEMIG (Fundação de Amparo à Pesquisa do Estado de Minas Gerais) for the financial support.

REFERENCES

Alves, V., Cury, A., Roitman, N., Magluta, C., Cremona, C., “Structural modification assessment using supervised learning methods applied to vibration data”, Engineering Structures 99, 439-448.

Cury, A., Cremona, C., Diday E., “Application of Symbolic Data Analysis for structural modification assessment”, Engineering Structures 2010, 32(3), 762-775.

Cury, A., Cremona, C., “Assignment of structural behaviors in long-term monitoring: Application to a strengthened railway bridge”. Structural Health Monitoring 2012, 1, 1-20.

Farrar, C., Worden, K, “Structural Health Monitoring: a machine learning perspective.” Chichester. Wiley. 2013.

Madhulatha, T.S., “An overview on clustering methods”, IOSR Journal of Engineering 2012, 2(4), 719-725.

Santos, J.P., Cremona, C., Orcesi, A.D., Silveira, P., “Multivariate statistical analysis for early damage detection”, Engineering Structures 2013, 56, 273-285.