*The 3rd International Conference on Structural Health Monitoring of Intelligent Infrastructure*
*Vancouver, British Columbia, Canada*
*November 13-16, 2007*

# DATA ANOMALY IDENTIFICATION IN COMPLEX STRUCTURES USING MODEL FREE DATA STATISTICAL ANALYSIS

Francesca Lanata
University of Genoa, Italy

Daniele Posenato
SMARTEC SA, Switzerland

Daniele Inaudi
SMARTEC SA, Switzerland

## Abstract

Civil engineering structures are difficult to model accurately and their real behaviour is hard to predict. Such situations encourage the enhancement of traditional approximate structural assessments through in-service measurements and interpretation of monitored data. While some proposals have recently been made, in general, no current methodology for detection of anomalous behaviour from measurement data can be reliably applied to complex structures in practical situations. This paper presents a model-free data interpretation methodology to identify and localize anomalous behaviours in civil engineering structures. A statistical method based on Moving Principal Component analysis has demonstrated to be useful for damage detection during continuous static monitoring of civil structures. The algorithm is designed to learn characteristics of time series generated by sensor data during a period called the initialisation phase where the structure is assumed to behave normally. This phase subsequently helps identify anomalous behaviours. No explicit knowledge of structural characteristics is necessary. The methodology will be first tested on a numerically simulated benchmark with sensors at a range of damage severities, and then the algorithm will be validated on a real structure. A comparative study with wavelets and other statistical analyses demonstrates superior performance for identifying damage.

## INTRODUCTION

Structural Health Monitoring systems have been recently developed and applied to an increasing extent in order to prevent serious failures of civil structures. Monitoring produces data, either continuously or periodically, that is analyzed to assess the safety and performance of structures. At the core of any health monitoring system stays the ability to automatically identify, localize and quantify structural damage [1]. In general, the presence of structural damage determines a loss of structural stiffness and it manifests itself as a change in the static and dynamic structural response. Damage detection procedures can be classified into static and dynamic approaches. It is generally easier to obtain dynamic data. This has been one of the reasons for the development of a much larger number of dynamic analyses than static ones. In complex structures, damage can be observed through changes in modal parameters. One of the main problems encountered is that modal parameter estimates are affected by different factors which are sometimes difficult to evaluate and control, such as environmental conditions, excitation techniques, data acquisition parameters, data processing methods and human factors. If the variation of the modal parameters is of the same order of magnitude than the variation due to damage, wrong conclusions could be drawn concerning the state of health of the structure [2]. The applicability of dynamic monitoring is limited since even significant damage may cause only small changes in natural frequencies [1,3,4]. On the contrary, changes in mode shapes are much more sensitive to damage when compared to changes in natural frequencies. However, damage is a local phenomenon and may not significantly influence mode shapes of the lower modes that are usually measured from vibration tests of complex structures and the number and location of sensors may have a crucial effect on the accuracy of the damage detection procedure [5]. The success of these techniques is also affected by the presence of noise in experimental data [5,6,7].

In spite of the large amount of work done in the field of dynamic identification, relatively few papers have considered the possibility of using static data for damage detection [8-10]. The measured quantities are typically displacements or strains due to environmental effects and applied loading, in order to control static parameters during the service conditions. Static monitoring can lead to damage identification by comparing static structural response with predictions from behavior models [5]. This is often referred to as system identification. However, models can be expensive to create and may not accurately reflect real behavior. Difficulties and uncertainties increase in presence of complex civil structures so that a well defined and unique behavior model cannot be clearly identified [6]. Furthermore, multiple-model system identification may not succeed in identifying the right damage [11]. When no models are used, long periods are required to produce reliable information [7]. This type of monitoring needs the realization of complex permanent instrumental monitoring systems, enabling in principle a continuous evaluation of the actual safety conditions of structures. Despite continuous evolution of research, for continuous static monitoring, no reliable strategies for identifying damage have been proposed and verified for broad classes of civil structures [9,12]. Static monitoring produces great amounts of measurements in different formats and the statistical evaluation of data is usually required [13,14]. This approach involves the study of changes in time series over time [15]. The methodology is completely data driven; the evolution of data is estimated without information of physical processes [16].

Although obtaining data from Structural Health Monitoring may be easy and fast, the phase of extraction of knowledge (processing) from measurements is usually complex and expensive in terms of time and resources. A generic monitoring system should have the following architecture (Fig. 1).

The data processing phase for damage detection will be discussed in this work. A general procedure able to discover anomalous behaviour in data generated by sensors, without using behaviour models, will be presented. The discussion will be based on the Principal Component Analysis [17] which has been modified using a moving window. The application of the algorithm to a simulated numerical benchmark will be presented as an example. Furthermore, when a great number of sensors are installed in the structure, the procedure will be integrated with a pre-processing algorithm (clustering technique) with the advantage of gathering together sensors with a similar behaviour. The application of the multi-algorithm to a real database will be finally presented. A comparative study with wavelets and other statistical analysis demonstrates different sensitivities to damage.

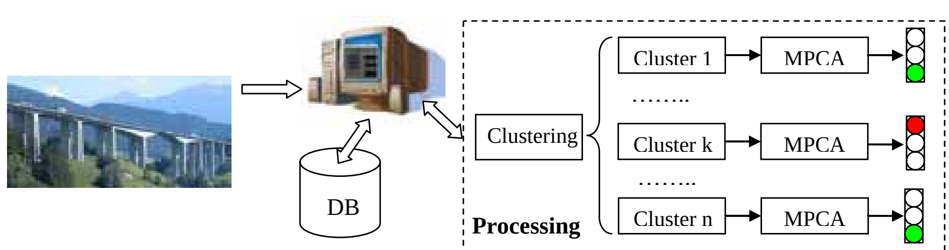


Figure 1. Generic monitoring system architecture.

## MOVING PRINCIPAL COMPONENT ANALYSIS (MPCA)

The PCA or Proper Orthogonal Decomposition [18,19] is a data reduction tool that is capable of transforming a number of related process variables into a smaller set of uncorrelated variables [20]. A key step is finding those principal components that contain most of the information. PCA is based on an orthogonal decomposition of the covariance matrix of the process variables along the direction that explain the maximum variation of the data. In this study PCA has been proposed to extract a spatial and temporal correlation between the installed sensors, related in a certain way to the normal modes contained in the static structural response. The measurements have to be arranged to form the measurement matrix: each column represents a displacement history at a particular sensor location and each row represents the spatial distribution of the response at a given time instant. After subtracting the mean displacements of each sensor location, the covariance matrix has to be computed and the relative proper orthogonal modes can be obtained. It is worthwhile to observe that the modal parameters extracted with the PCA are the static modal properties. While in usual dynamic applications modal parameter modifications are not considered very sensitive to damage, an intrinsic advantage of strain and displacement measurements consists of an assessment of static modes which are much more sensitive to damage.

With the classic PCA, the time necessary to compute the covariance matrix increases with the number of measurements and there is a delay in detecting a new situation in the time series. This is due to the fact that with the increasing of the number of measurements, the effect of new points in the covariance matrix is lower and lower because they are averaged by the total number of points. Previous work has proposed solutions to this problem through following the evolution of series over time, usually through recursive strategies [20]. The algorithm proposed here, called MPCA, computes the covariance matrix inside a moving window of constant size [21], in such a way that the old measurements don't buffer the results. The modified procedure has the advantage of reducing the computational time in presence of long-term time series even if it has shown to be less sensitive to damage initiation in comparison with the classic POD [17]. A reference period is necessary to estimate robust thresholds that will be used for discovering anomalies. The thresholds are used to compare future behavior of the structure with the one observed during the reference period. An appropriate period of reference in which the structure has to remain in its reference state is necessary to reduce the number of false alarms.

The eigenvalues $\lambda_i$ and the orthogonal eigenvectors $\mathbf{u}_i$ are calculated from the covariance matrix (with measurements inside the active window only). As $\lambda_i$ catch the time behaviour of the system, the eigenvectors associated with the main eigenvalues represent the main spatial behaviours of the structure. If the sensors are correlated, the main eigenvalue $\lambda_i$ catches the main behaviour, common to all sensors. The time history of each sensor can be reconstructed using only the first few principal components that contain most information. As a consequence, the following sum can be truncated:

$$\mathbf{s} = \lambda_1\mathbf{u}_1 + \lambda_2\mathbf{u}_2 + \ldots + \lambda_n\mathbf{u}_n \cong \lambda_1\mathbf{u}_1 + \lambda_2\mathbf{u}_2 + \ldots \tag{1}$$

The vector $\mathbf{u}_i$ tells the contribute of $\lambda_1$ to each sensor. As a consequence, if a generic sensor *j* has a problem, its contribute to the main behaviour is low, because $\lambda_1$ is a function of all the *n* involved sensors, and a variation on the $u_{j1}$ should appear. For this reason the first eigenvalue has been excluded from the damage function *DF* that indicates the insurgence of damage and is defined as:

$$DF_i(t) = u_i(t)\lambda_i \qquad for\ i = 2,\ldots,n \tag{2}$$

A damage localization index *DLI* indicating the location of damage and a threshold value that allows the determination of damage intensity, have to be defined in order to complete the damage identification procedure [17]. The *DLI* is defined as the square of the difference of the eigenvectors at two generic observation times:

$$DLI_i(t_1) = \left(u_i(t_1) - u_i(t_2)\right)^2 \qquad for\ i = 1,\ldots,n \tag{3}$$

The threshold value has to be defined case by case from the statistical observation of an initial phase in which the structure is assumed to be undamaged (reference period). This period is normally one or two years. Once thresholds have been fixed, the parameters of the process are monitored to ensure that they are inside predefined ranges. For damage localization the rule has been used that candidate damage zones are close to sensors that have measurements values exceeding a threshold.

## NUMERICAL SIMULATION

Long-term databases from real structures with a range of damage intensities are not easy to retrieve. As a consequence, in order to evaluate the efficiency of the proposed algorithm to detect damage, a finite element model of a beam has been modeled. The aim of the numerical benchmark is to simulate the behavior of a bridge (two span continuous beam) in healthy and various damaged states [12]. As the literature shows that a reliable bridge monitoring program should be conducted under environmental conditions [2], the model is able to simulate thermal loads in order to detect anomalies in the operational behavior under ambient excitation. The structural response is measured by means of a 'virtual' monitoring system installed on the beam. The sensor system has been supposed to be constituted of long-gauge-length fiber optic sensors like SOFO [22] having a resolution of 2 microns and a long-term stability. A set of 12 elongation sensors have been installed at the reference state of the structure; measurements deriving from the virtual monitoring are the elongation and the shortening of the sensors. Each sensor has a measuring base of 1 meter (Figure 2). The simulated responses contain measurement errors and noise.

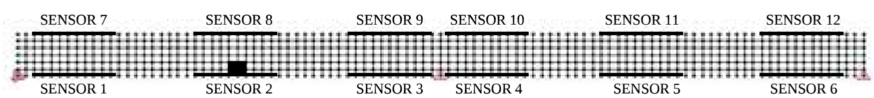


Figure 2. Two span continuous beam discretization and sensors location. (The beam is 10x0.5x0.3 m. The black cells represent an example of a simulated damage scenario.)

Various location and intensities of damage have been simulated; cracks were simulated in order to model local degradation in material properties. Damage has been represented by a finite element with a reduced stiffness [17]. The example with four damaged elements with a stiffness reduction of 80% located within the Sensor 2 is represented in Figure 2. The monitoring system has been designed in order to give the long-term response of the beam during eight years, assuming four measurements in a day. The elongation time series show harmonic variations due to the seasonal variations of temperature in both healthy and damaged states. However, no clear correlation to damage formation has been observed directly from the elongation-sensor time histories.

MPCA algorithm has been applied to the data. The reference period has been defined to be two years and the size of the moving window is one year. In this way the reference period includes one-year environmental cycle and leaves extra time to have a better estimation of thresholds. MPCA algorithm has been applied to the twelve time series in the simulation, one for each sensor, in order to simulate measurement evolution during monitoring phases. MPCA uses all temporal values inside the active window to compute the covariance matrix and at each temporal step the window is shifted. The application of the MPCA to the time histories of the undamaged and damaged responses has shown to be reliable for the detection of small damages, interesting only one element of the mesh with a stiffness reduction of 80%. MPCA diagnostic plots of eigenvectors related to the main eigenvalues in the damage scenario with four damaged elements are shown in Figure 3. The time when damage occurs and its location are visible in the graphs simply following the evolution of the two main eigenvectors. Specifically, eigenvector 11 indicates the transition to a new state associated with damage, while eigenvector 12 indicates when damage occurred. The location of damage can be detected by observing that the changing eigenvectors are associated with sensors close to the damaged sections.

Figures 4 and 5 show the *DF* and the *DLI* as defined in the previous section. The *DF* corresponds to the damage scenario in Fig. 2: the initiation of damage can be clearly identified. The *DLI* shown in Fig. 5 is related to the damage scenario with only one damaged element with a stiffness reduction of 50% that represents the minimum detectable damage of the beam with this algorithm. The plot shows that: i) in Section 2 the parameter related to damage overcomes the threshold previously defined through the statistical analysis of the undamaged response of the beam; ii) damage is localized near the Sensor 2. The threshold is overcame after 1-2 months after damage initiation [17].

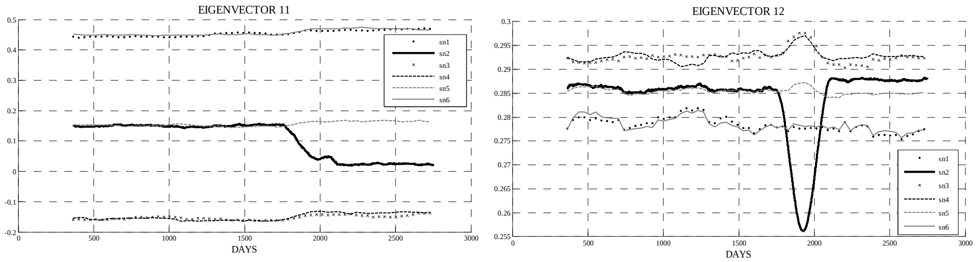


Figure 3. MPCA plots of eigenvectors related to the two main eigenvalues.

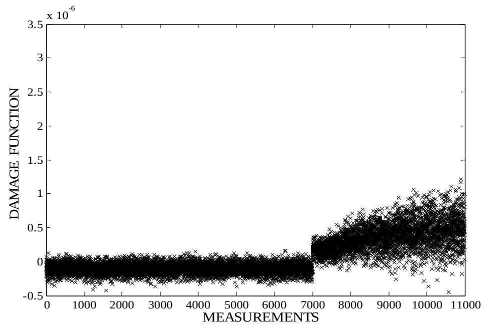


Figure 4. Plot of the *DF* with time for the damage scenario with 4 damaged elements within the Sensor 2.

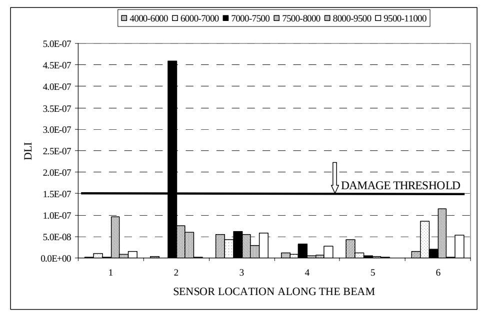


Figure 5. *DLI* for one element with a stiffness reduction of 50% within the Sensor 2.

## MPCA ON A REAL STRUCTURE WITH SEVERAL SENSORS

In order to validate the proposed algorithm, the procedure of damage detection has been also applied to a real database deriving from a continuous static monitoring of an existing structure. The structure under study is the Colle Isarco viaduct on the Italian Brenner-Highway A22, an example of a global monitoring approach in establishing a bridge management system [23]. The photo of the bridge can be seen in Fig. 6. A wide set of sensors have been installed, including both traditional and fiber optic sensors (137 SOFO sensors [22]). A data acquisition system able to collect widely distributed sensing units was also installed due to the large dimensions of the instrumented section.
In comparison with the modeled benchmark, real monitored structures usually have a lot of installed sensors of different types and characterized by different behaviors. As a consequence, a data pre-processing procedure can be necessary to optimize the damage detection algorithm in real structures. A clustering algorithm will be presented in this work, with the aim to cluster and gather together the sensors with comparable behaviors and similar sensitivity to damage. The application of MPCA to small clusters seems to be faster and easier than the application to the complete set of sensors in the structure. Clustering is a useful technique for grouping data points such that points within a single group/cluster have similar characteristics (or are close to each other) [24]. Clustering techniques have been studied extensively in statistics [25], pattern recognition [26], and machine learning [27]. Current clustering techniques can be broadly classified into two categories [28]: partitional and hierarchical. Given a set of objects and a clustering criterion, partitional clustering obtains a partition of the objects into clusters such that the objects in a cluster are more similar to each other than to objects in different clusters. The *K*-means and *K*-medoid methods determine *K* cluster representatives and assign each object to the cluster with its representative closest to the object such that the sum of the Euclidean distances squared between the objects and their representatives is minimized. A hierarchical clustering is a nested sequence of partitions; it starts by placing each object in its own cluster and then merges these atomic clusters into larger and larger clusters until all objects are in a single cluster.

A cluster algorithm has been applied on the Colle Isarco viaduct due to the high number of installed sensors. The chosen algorithm is an extension of the *K*-means method with overlapping clusters, in which each sensor can be included in one or more clusters. This choice permits to have a data redundancy for a better identification of damage position. In this application the correlation matrix of the complete set of sensors has been used as distance criterion instead of the traditional Euclidean distance; as a consequence, the correlation matrix on all the 137 fibre optic sensors for a significant observation period has been computed. Each cluster is formed by the *N* sensors more correlated to each one. In this example each cluster contains 10 sensors. Fig. 6 shows an example of the spatial location of the 10 sensors contained in a cluster in two generic sections of the bridge. After the generation of the 137 clusters, the MPCA has been applied to each group in order to test the effectiveness of the algorithm. Fig. 7 shows a plot of the 10 eigenvectors corresponding to the main eigenvalue within a single cluster. The 10 eigenvectors show a good correlation and they are stable in time. The plot shows that no damage has occurred in the bridge during the monitoring period and the MPCA for the real structure stops at this level.

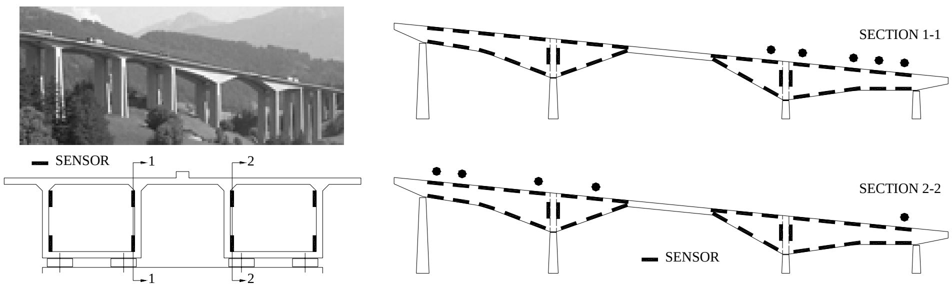


Figure 6. General view, transversal section of the bridge and spatial position along the bridge of the 10 sensors into a generic cluster (marked with the black circles).

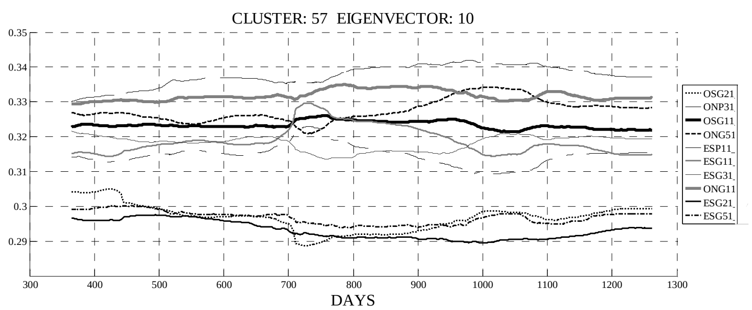


Figure 7. MPCA plot of eigenvector related to the main eigenvalue of a generic cluster of sensors in the bridge.

## COMPARATIVE ANALYSIS BETWEEN MPCA AND OTHER STATISTICAL ALGORITHMS - CONCLUSIONS

The main difficulty encountered in the field of the damage detection from a continuous static monitoring is that the existing studies apply different methods to different structures, rendering side-by-side comparison more difficult. On the contrary, the focus of this work was to apply different methods to the same simulated structural response for comparing damage sensitivity of the various algorithms.

The considered statistical methods are Continuous Wavelet Transform (CWT) and Correlation Analysis (CA). A theoretical treatment of wavelets and wavelet analysis may be found in [29,30]. The CWT has been applied to the difference of the time series of the two sensors that are closest to damage and normalized according the values of the first year. Concerning the CA, the correlations between all the couple of sensors have been computed. During the reference period (first two years), correlations have been computed taking into account the measurements inside a moving window of one year in order to define for each sensor the group of sensors higher correlated and to estimate their variations. After the reference period the variation of the correlations has been monitored in order to see if some of them were out of the defined thresholds.

The performances of the compared statistical algorithms are presented in Table 1. MPCA results to perform better that the other algorithms for all the simulated damage scenarios of the beam.

Table 1. Results of a comparative study between MPCA, CWT and CA using data derived from the benchmark in various damages states. D = detected, ND = not detected.

| **Damage scenario** | **CWT** | **Moving correlation** | **MPCA** |
|---|---|---|---|
| 4 damaged elements at 80% within Sensor 2 | D | D | D |
| 2 damaged elements at 80% within Sensor 2 | D | D | D |
| 1 damaged element at 50% within Sensor 2 | ND | ND | D |
| 4 damaged elements at 80% between Sensors 2 and 3 | ND | D | D |

## REFERENCES


1. Bisby, L.A., ‘An introduction to Structural Health Monitoring’, ISIS Educational Module 5 (2005).
2. Wenzel, H. and Pickler, D., ‘Ambient Vibration Monitoring’, (Wiley, England, 2005).
3. Doebling, S.W., Farrar, C.R., Prime, M.B. and Shevitz, D.W., ‘A review of damage identification methods that examine changes in dynamic properties’, Shock and Vibration Digest 30(2) (1998) 91-105.
4. Yan, Y.J., Cheng, L., Wu, Z.Y. and Yam, L.H., ‘Development in vibration-based structural damage detection technique’, Mech. Syst. & Sign. Proc. (2007) in press.
5. Tim, J.-T., Ryu, Y.-S., Cho, H.-M. and Stubbs, N., ‘Damage Identification in Beam-type Structures: Frequency-based Method vs Mode-shape-based Method’, Engineering Structures 25(1) (2003) 57-67.
6. Hong, J.-C., Kim, Y.Y., Lee, H.C. and Lee, Y.W., ‘Damage detection using Lipschitz exponent estimated by wavelet transform: applications to vibration modes of beam’, Int. J. of Solid and Structures 39 (2002) 1803-1846.
7. Messina, A., ‘Detecting damage in beams through digital differentiator filters and continuous wavelet transform’, J. of Sound and Vibration 272 (2004) 385-412.
8. Biondani, E. R., Vardanega, C., Vaccaro, G., Mirone, E. and Saveri, E., ‘Observed and predicted settlements of two Italian nuclear power plants’, Trans. 9th Int. Conf. on Structural Mechanics in Reactor Technology (Balkema, Rotterdam, 1987) D 293-298.
9. Omenzetter, P., Brownjohn, J.M.W. and Moyo, P., ‘Identification of unusual events in multi channel bridge monitoring data’, Mech. Syst. & Sign. Proc. 18 (2004) 409-430.
10. Del Grosso, A., Inaudi, D. and Lanata, F., ‘Strain and displacement monitoring of a quay wall in the Port of Genoa by means of fibre optic sensors’, in Proc. of the 2nd Europ. Conf. on Structural Control, Paris, July 2000 (2000).
11. Saitta, S., Raphael, B. and Smith, I.F.C., ‘Data mining techniques for improving the reliability of system identification’, Advanced Engng. Informatics 19 (2005) 289-298.
12. Lanata, F., ‘Damage detection algorithms for continuous static monitoring of structures’, PhD Thesis Italy University of Genoa (2005).
13. Sohn, H., Czarneski, J.A. and Farrar, C.R., ‘Structural Health Monitoring using statistical process control’, J. of Structural Engineering 126 (2000) 1356-1363.
14. Lanata, F. and Del Grosso, A., ‘Damage detection algorithms for continuous static monitorino: review and comparison’, in Proc. of the 3rd Europ. Conf. on Structural Control, Wien, July 2004 (R. Flesch, H. Irschij and M. Krommer, Wien, Austria, 2004).
15. Omenzetter, P. and Brownjohn, J.M.W., ‘Applications of time series analysis for bridge monitoring’, Smart Materials and Structures 15 (2006) 129–138.
16. Camerino, M. and Peters, K., ‘Non-model-based approach to damage localization’, Proc. of SPIE, Non-destructive Evaluation for HM and Diagnostics 5393 (2004) 111-122.
17. Lanata, F. and Del Grosso, A., ‘Damage detection and localization for continuous static monitoring of structures using a proper orthogonal decomposition of signals’, Smart Materials and Structures 15(6) (2006) 1811-1829.
18. Kreuzer, E. and Kust, O., ‘Analysis of long torsional strings by proper orthogonal decomposition’, Archive Applied Mechanics 67 (1996) 68-80.

19. Solari, G., Carassale, L. and Tubino, F., 'POD methods and applications in wind engineering', 6th Asia-Pacific Symposium on Wind Engineering, Seoul (2005).
20. Hubert, M., Rousseeuwy, P.J. and Brandenz, K.V., 'ROBPCA: a New Approach to Robust Principal Component Analysis', Technometrics 47(1) (2005) 64-79.
21. Posenato, D., Lanata, F., Inaudi, D. and Smith, I.F.C., 'Model free data interpretation for continuous monitoring of complex structures', Adv. Engng. Inf. (2006) in press.
22. Inaudi, D. and Glisic, B., 'Long-gage sensor topologies for structural monitoring', Pro. 1st fib Congress on Concrete Structures in the 21st Century, Osaka, Japan (2002).
23. Bergmeister, K. and Santa, U., 'Global monitoring concepts for bridges', Proceedings of the fib-TG 5.1, Institute of Structural Engineering, Vienna (1999).
24. Anderberg, M.R. 'Cluster Analysis for Applications' (Academic Press, 1973).
25. Arabic, P. and Hubert, L.J., 'An overview of combinatorial data analysis, Clustering and Classification', (World Scientific Pub., New Jersey, 1996) 5-63.
26. Fukunaga, K., 'Introduction to Statistical Pattern Recognition', (Academic Press, 1990).
27. Cheeseman, P. and Stute, J., 'Bayesian classification (autoclass): Theory and results', Advances in Knowledge Discovery and Data Mining, (AAAI/MIT Press, 1996) 153-180.
28. Jain, A.K. and Dubes, R.C., 'Algorithms for Clustering Data', (Prentice Hall, 1988).
29. Chui, C.K., 'Introduction to Wavelets', (Academic Press, San Diego, 1992).
30. Daubechies, I., 'Ten Lectures on Wavelets', (Philadelphia: Society for Industrial and Applied Mathematics, 1992).