Extended PCA visualisation of system damage features under environmental and operational variations

Luis E. Mujica[a], José Rodellar[b], Josep Vehí[c], Keith Worden[d] and Wieslaw Staszewski[d]

[a]Department of Applied Mathematics III. Escola Universitària d'Enginyeria Tècnica Industrial de Barcelona (EUETIB). Universitat Politécnica de Catalunya (UPC). Comte d'Urgell 187, 08036, Barcelona, Spain.

[b]Department of Applied Mathematics III. Escola Tècnica Superior d'Enginyers de Camins, Canals i Ports de Barcelona (ETSECCPB). Universitat Politécnica de Catalunya (UPC). Jordi Girona, 1-3, 08034, Barcelona, Spain.

[c]Institut d'Informàtica i Aplicacions.
University of Girona. Campus Montilivi, Edifici P-IV, 17071. Girona, Spain.

[d]Department of Mechanical Engineering.
University of Sheffield. Mappin Street S1 3JD. Sheffield-UK.

ABSTRACT

This paper explores the use of Principal Component Analysis (PCA), an extended form of PCA and, the T^2-*statistic* and *Q-statistic*; distances that detect and distinguish damages in structures under varying operational and environmental conditions. The work involves an experiment in which two piezoelectric transducers are bonded on an aluminium plate. The plate is subjected to several damages and exposed to different levels of temperature. A series of tests have been performed for each condition. The approach is able to determine whether the structure has damage or not, and besides, gives qualitative information about its size, isolating effects of the temperature.

Keywords: Damage Identification under uncertainties, Principal Component Analysis, T^2-*statistic*, *Q-statistic*

1. INTRODUCTION

In general, the objective of Structural Health Monitoring (SHM) is to find out if damage is present in the structure or not, based on measured dynamic responses. For example, changes in the properties of a structure may be indicated by changes in propagated elastic waves or changes in observed natural frequencies or other vibration-based features. Unfortunarley, in the real world, structures are subject to fluctuating environmental and operational conditions that affect measured signals and these variations can mask the changes in dynamical response caused by damage. The result of this may be a missed fault, or alternatively, a false alarm. As a result, the problem of distinguishing a damaged structure from an undamaged one becomes more complicated, and trying to identify different defects is even more difficult.

In this paper, previous work on the use of Principal Component Analysis (PCA) as an unsupervised method for damage identification in SHM is extended. In past work, a number of approaches based on PCA models: T^2-*statistics* and *Q-statistic* etc. have been developed. These are statistical distance measures which not only detect defects, but potentially distinguish between different severities.[1] However, in much of the previous work, the experiments were performed in state-stable operational conditions. In reality, it has been shown that effects of the temperature on the measured signal induce certain patterns; in other words, the dynamic responses of a structure subjected to different levels of temperature will have certain correlations that should be studied as they can confuse damage detection algorithms.[2]

Further author information: (Send correspondence to L.E.M.)
L.E.M.: E-mail: luis.eduardo.mujica@upc.edu, Telephone: +34 934 137 437
J.R.: E-mail: jose.rodellar@upc.edu, Telephone: +34 934 016 865
J.R.: E-mail: josep.vehi@udg.edu, Telephone: +34 972 418 495
K.W.: E-mail: K.Worden@sheffield.ac.uk, Telephone: +44 (0)114 222 7758
W.S.: E-mail: W.J.Staszewski@sheffield.ac.uk, Telephone: +44 (0)114 222 7759

In this paper, the use of PCA, an extended PCA and the previously mentioned statistical measures are exploited in order to detect and distinguish damages in structures under varying operational and environmental conditions. The approach is able to determine whether the structure has damage or not, and can also give a qualitative information about its size, thus isolating the effects of the temperature.

Typically, PCA in SHM research is used to eliminate data redundancy between the data-channels or measured signals. To isolate the effects of the temperature, elimination of the redundancy (or correlations) between signals is critical. PCA can be extended in order to study the correlations between signals and reduce that redundancy. In this approach, the usual training data is reorganized in such a way that a conventional PCA can be applied.

A PCA model is built using data from the undamaged structure. Data from the damaged structure tests are the projected on the model. The principal components, the T^2*-statistics* and the *Q-statistics* are then analyzed. The *Q-statistic* indicates how well each sample conforms to the PCA model; it is a measure of the difference, or residual between a sample and its projection onto the principal components retained in the model. The T^2*-statistic* distance is a measure of the variation in each sample within the PCA model.

This work involves a simple experiment in which two piezoelectric transducers are bonded on an aluminium plate. The plate is subjected to several damages (holes of 1mm, 3mm and 5mm diameter) and exposed to different levels of temperature in four experimental steps: (1) heating (from 35^{o}C to 70^{o}C), (2) cooling from (70^{o}C to 35^{o}C), (3) heating and (4) cooling. A series of tests have been performed for each condition of the plate (intact and damaged) at each level of temperature. Comparison between the effects of damage and temperature change upon the responses is performed.

The structure of the paper is as follow. Section 2 describes in detail the experimental setup. The methodology for visualisation of damage features under environmental and operational variations is introduced in section 3; also, the effects of the temperature changes on Lamb wave responses is explained. In addition, the concepts of Principal Component Analysis (PCA), damage identification indices, and the extensions of PCA to isolate temperature effects are presented. Finally, results of applying a conventional PCA and/or the extensions proposed in this paper are analyzed and the paper concludes in section 4.

2. EXPERIMENTAL SETUP

The work involves a simple experiment in which two piezoelectric transducers (P155, radius 10mm and thickness 1mm) are bonded, in a symmetrical configuration, on an aluminium plate (200mm x 150mm x 2mm) as can be seen from figure 1. This experiment was previously performed to study the effects of variations of temperature on Lamb wave responses.[2]

One piezoelectric disk, used as an actuator, was excited with a five-cycle sine signal with the frequency equal to 75 KHz, it was modulated by a half-cosine function with a maximum peak-to peak amplitude of 10V. This excitation signal was generated using a *TTi TGA 1230* arbitrary waveform generator. The second piezoelectric disk was used as a sensor in order to capture Lamb waves propagated in the specimen. The responses were acquired using a digital 4-channel *LeCroy LT264* oscilloscope. The specimen was placed in a 100 liters *LTE Scientific Oven*. A thermal probe was used to measure the temperature on the surface of the plate. The experimental setup is shown in figure 2.

Several damages were introduced into the middle of the plate (holes of 1mm, 3mm, and 5mm diameter). The plate (undamaged and damaged) was exposed to different levels of temperature in four consecutive steps: (1) heating from 35^{o}C to 70^{o}C, (2) cooling from 70^{o}C to 35^{o}C, (3) heating from 35^{o}C to 70^{o}C, (4) cooling from 70^{o}C to 35^{o}C. Lamb wave responses were recorded at each stage of the plate (undamaged and damaged). Finally, the recorded data were stored in a 3D matrix bfor each state of the plate as shown in figure 3.

In short, 50 experiments were performed (y-axis) at 29 levels of temperature (z-axis) and 4500 samples were recorded by signal (x-axis), in this way, each frontal slice represents all measurements at one level of temperature. D0 is data from the undamaged plate; D1, D3 and D5 are data from the damaged plate with holes of 1mm, 3mm and 5mm, respectively.

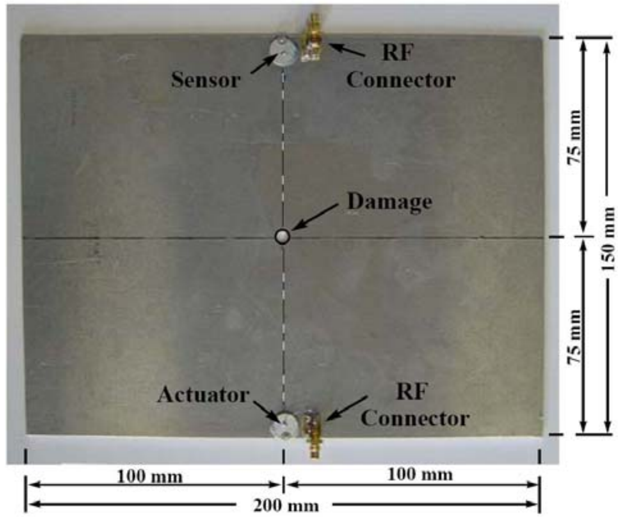

Figure 1. Aluminium plate with sensor locations[2]

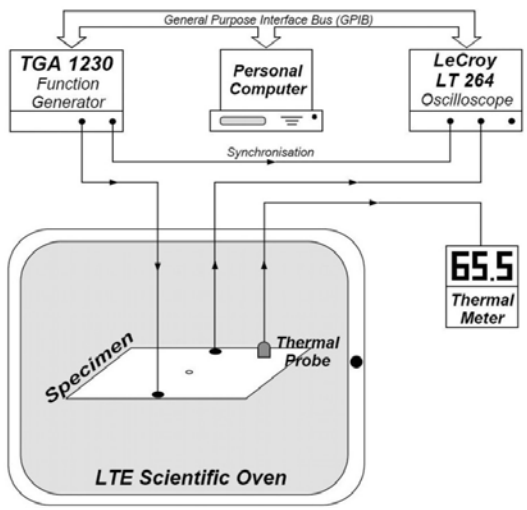

Figure 2. Experimental setup[2]

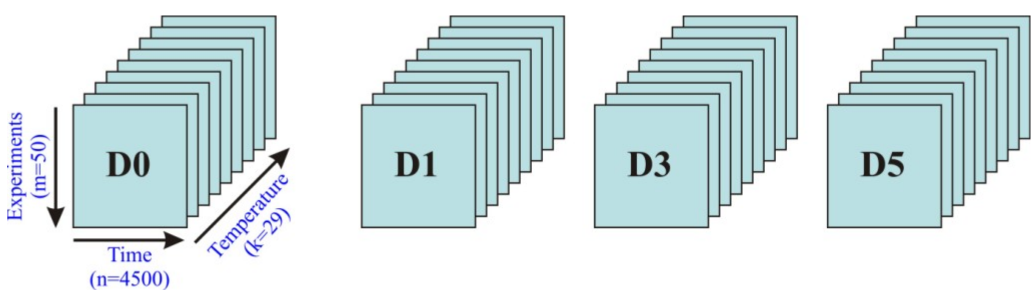

Figure 3. Data set organization

3. METHODOLOGY

In the context of Structural Health Monitoring, especially for damage detection, features (any set of values drawn from or calculated from the measured data) which are capable of distinguishing between the undamaged and damaged states regardless of variations in the environmental and operational conditions are desirable. Using the experimental setup described in the previous section, Lee and co-workers[2] showed that amplitudes and shapes of wave packets change with the variation in temperature and the introduction of damage and besides, the relationship between them is not easily distinguishable. Two different features were considered there: (i) The peak-to-peak value of the signal, and (ii) the two principal components of an area of the signal which is sensitive to damage. The former shows that the amplitudes of the responses decrease with increasing temperature, but the values of amplitude of the same experiment are highly scattered. The latter shows that the difference between two undamaged cases with different temperature is more pronounced than the undamaged and damaged features at the same temperature. The paper concluded that the effect of the temperature changes is so great that it dwarfs the signal changes observed when damage is introduced.

In this work, PCA is proposed as a means to study correlations between data-points in each response signal, signals from the same experiments at different levels of temperature, signals from different experiments at the same level of temperature, and all of them. In this section, the concept of PCA and damages indices are briefly explained. Also, how the data must be organized in order to analyze the correlations of interest is discussed.

3.1 Principal Component Analysis (PCA)

PCA is a linear projection technique used for data compression, and information extraction and interpretation. PCA finds combinations of variables or factors that describe major trends in a data set.[3] PCA is concerned with explaining the variance-covariance structure through a few linear combinations of the original variables. The formulation of PCA is relatively simple. Consider a data set, $\mathbf{X}_{m\times n}$, consisting of m observations and n measured variables. The first step in applying PCA is to standardize the data matrix $\mathbf{X}$, since PCA is scale variant. The mean trajectories are removed and all variables are made to have equal variance. As a consequence, the trends on the observations and their standard deviations, often non-linear in nature, are removed from the data. Once the variables have been standardized, the covariance matrix $\mathbf{S}$ is calculated as:

$$\mathbf{S} = \frac{1}{m-1}\mathbf{X}^T\mathbf{X}. \tag{1}$$

This matrix measures the degree of the linear relationship between the data set of elements to one another. The subspaces in PCA are defined by the eigenvectors and eigenvalues of the covariance matrix as follow:

$$\mathbf{S}\hat{\mathbf{P}} = \hat{\mathbf{P}}\mathbf{\Lambda}. \tag{2}$$

Eigenvectors of $\mathbf{S}$ are the columns of $\hat{\mathbf{P}}$, and eigenvalues of $\mathbf{S}$ are the diagonal terms of $\mathbf{\Lambda}$ (the off-diagonal terms of $\mathbf{\Lambda}$ are zero). The eigenvector with the highest eigenvalue represents the most important pattern in the data, i.e. contains the largest quantity of information. Therefore, this vector is called the *principal component* of the data set. Ordering the eigenvectors by eigenvalue, highest to lowest, gives the components in order of significance. The columns of this new matrix $\mathbf{P}$ (loading matrix) are named loading vectors (p_i) and describe the linear combinations of the original variables that constitute each principal component. In order to reduce the dimensionality, the less important components can be eliminated (information is lost, but if the eigenvalues are small, this information can be disregarded). If only the k-first eigenvectors are chosen, then the final data set will be k-dimensional. The projected matrix $\mathbf{T}$ (or score matrix) in the new space is defined by equation (3), the projection of $\mathbf{T}$ back onto the k-dimensional observation by equation (4). Letting the residual matrix $\mathbf{E}$ be the difference between $\mathbf{X}$ and $\hat{\mathbf{X}}$ (see equation 5), the formulation of PCA is defined by equation (6).

$$\mathbf{T} = \mathbf{X}\mathbf{P}, \tag{3}$$

$$\hat{\mathbf{X}} = \mathbf{T}\mathbf{P}^T, \tag{4}$$

$$\mathbf{X} = \hat{\mathbf{X}} + \mathbf{E}, \tag{5}$$

$$\mathbf{X} = \mathbf{T}\mathbf{P}^T + \mathbf{E}. \tag{6}$$

3.2 Damage Identification Indices

The score matrix $\mathbf{T}$ (its columns consist of score vectors, t_i, associated with the principal components PC_i and the residual matrix $\mathbf{E}$ can be used in order to detect abnormal behaviour in a process or set of responses. With this aim, the *Q-statistic* (or *SPE-statistic*) and T^2*-statistics* (or T^2*-Hotelling's*, *D-statistic*) are used to represent the variability of the projection in the residual subspace and the new space respectively. These methods are based on the assumption (generally stemming from the central limit theorem) that the underlying process follows approximately a multivariate normal distribution where the first moment vector is zero (see figure 4).

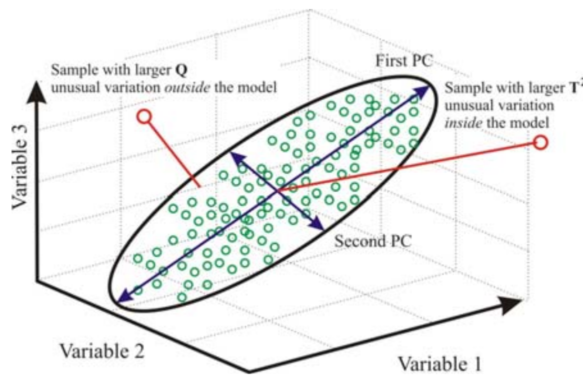

Figure 4. PCA model of three dimensional data set showing T^2*-statistics* and *Q-statistics* outliers[4]

3.2.1 T^2*-statistics*

In statistics, *Hotelling's* T^2*-statistic* is a generalization of *Student's t-statistic* that is used in multivariate hypothesis testing. It denotes the inner change of the principal component model. The T^2*-statistics* of the *i*-th sample is defined by:

$$T_i^2 = t_i \mathbf{\Lambda}^{-1} t_i^T, \tag{7}$$

and only detects variation in the plane of the first k principal components which are greater than what can be explained by the common-cause variations. In other words, T^2*-statistics* is a measure of the variation in each sample within the PCA model.

3.2.2 *Q-statistics*

The *Q-statistic* denotes the components of the changed data which are not explained by the initial model of the principal components. In other words, it is a measure of the difference, or residual between a sample and its projection into the model. The *Q-statistics* of the *i*-th sample is defined by:

$$Q_i = e_i e_i^T = x_i(\mathbf{I} - \mathbf{P}\mathbf{P}^T)x_i^T, \tag{8}$$

where e_i is the principal component model error of the *i*-th sample (*i*-th row of E) and x_i is the *i*-th sample of the whole set of events (*i*-th column of X).

Normally, The *Q-statistic* is much more sensitive than the T^2*-statistic*. This is because Q is typically very small and therefore any minor change in the system characteristics will be observable. In contrast, T^2 has great variance and therefore requires a major change in the system characteristic for them to be detectable.

Information about the events can also be obtained from the plot of scores for the relevant principal components. When there is a change in the system, the scores of the new events will be very different from the previous scores, and the change will be detected. However, this information is also included in the *Hotelling's* T^2*-statistics* since it is calculated using the scores. Moreover, the *Q-statistic* gives us additional information which is not included in the scores plot, because it is related to the variation which is not considered by the model. In this way, the plots of Q and T^2 are a hypothesis test which clearly distinguish any signals with abnormal behavior whereas the inspection of the scores plot is a qualitative tool.[5]

3.3 Principal Component Analysis Extension

In the original formulation of PCA (section 3.1), the original data set, $\mathbf{X}_{m\times n}$, consisting of m observations and n measured variables is considered. The standard PCA analyzes correlations between variables. For chemometricians (pioneers in PCA applications), some variables are continuous in time, therefore they included a third dimension to the data matrix ($\mathbf{X}_{m\times n\times k}$) and they studied different ways to analyze both correlations: between variables and within each time history.[6] The most successful technique was to perform ordinary PCA on a large 2D matrix constructed by unfolding the 3D matrix. Six possible ways of unfolding were suggested.[7]

In SHM, applications of PCA and some extensions were successfully applied to detect and locate an impact on a part of an aircraft wing.[8] In this case, the structure was provided with several sensors which captured the vibration caused by the impact. The data set therefore consisted of m experiments (observations); one by impact, n data-points or samples (variables) and k sensors. To study both correlations: between samples and between sensors (but not between experiments, because every experiment is a different impact), the six possible ways of unfolding were analyzed. Due to the nature of the problem, two of them were applied (two were not applicable and the other two are equivalent to the selected ones).

In this study the data set consists of m experiments (observations), n data-points or samples (variables) and k levels of temperature as can be seen from figure 3. In this way, it is considered that each sample is linearly independent of the others. As the aim of this work is to analyze the correlation between samples, also correlation between experiments (because in each matrix, all experiments are performed using the same defect in the plate) and also, to consider the temperature, the PCA technique has to be extended and adapted. The two ways of adaption (termed unfolding) used previously are applied and a third way is also proposed.

3.3.1 Unfolding in the experiment direction

Each of the front slices of data are put next to each other as in figure 5. The unfolded matrix (to which PCA is applied) therfore has dimensions 50×130500 and every row represents the data from the experiments at all levels of temperature thus, an experiment (at all levels of temperature) is considered as an object. Correlations between samples and between temperatures are analyzed, therefore, differences between experiments, no matter the temperature, are highlighted. In this way, every experiment (losing information about at what level of temperature was performed) can be compared against a group of undamaged experiments.

3.3.2 Unfolding in the time direction

Each front slice is below another as is shown in figure 6. The unfolded matrix has dimensions 1450×4500; in this way every experiment at each level of temperature is considered as an object independent of each other. Correlation between samples at all levels of temperatures are analyzed (effects of the temperature are not considered), thus, differences between experiments and temperatures can be visualized. Therefore, every experiment, including information about temperature can be compared with a a group of undamaged experiments.

3.3.3 Unfolding in the experiment-time direction

In short, unfolding in the experiment direction allows one to study correlations between temperatures, but they can not be visualized. Whereas, unfolding in the time direction does not consider these correlations, although they can be visualized. In order to combine the advantages of these unfolding ways, in this work, a third unfolding way is proposed. Here, each front slice is put next to the other as in the unfolding of the experiment direction. Later, the whole unfolded matrix is duplicated and put below the first, but circularly shifted by n samples. This matrix is duplicated and circularly shifted again and put below the previous one. This process is repeated k times, as seen in figure 7. The final transformed matrix has dimensions 1450×130500, thus an experiment at all levels of temperature is considered as an object, so both correlations between samples and temperature are analyzed, and also differences between experiments and temperatures can be visualized.

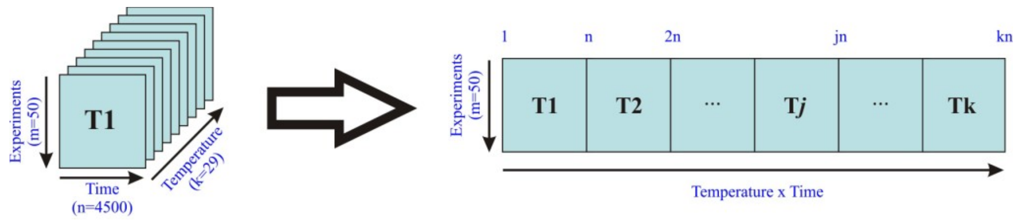

Figure 5. Unfolding of data set in experiment direction.

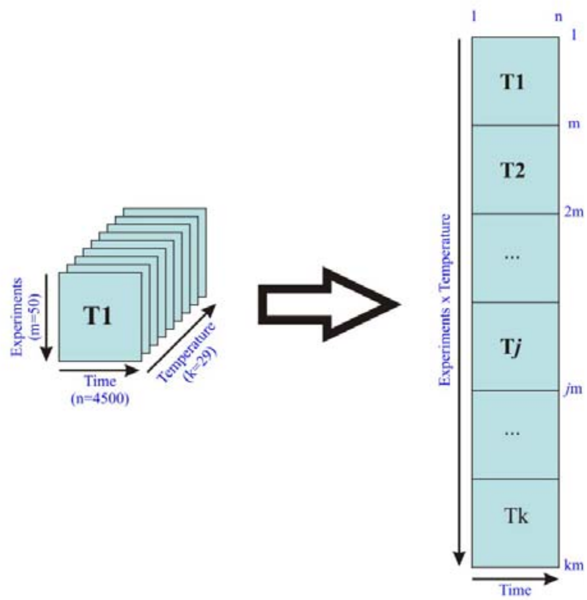

Figure 6. Unfolding of data set in time direction.

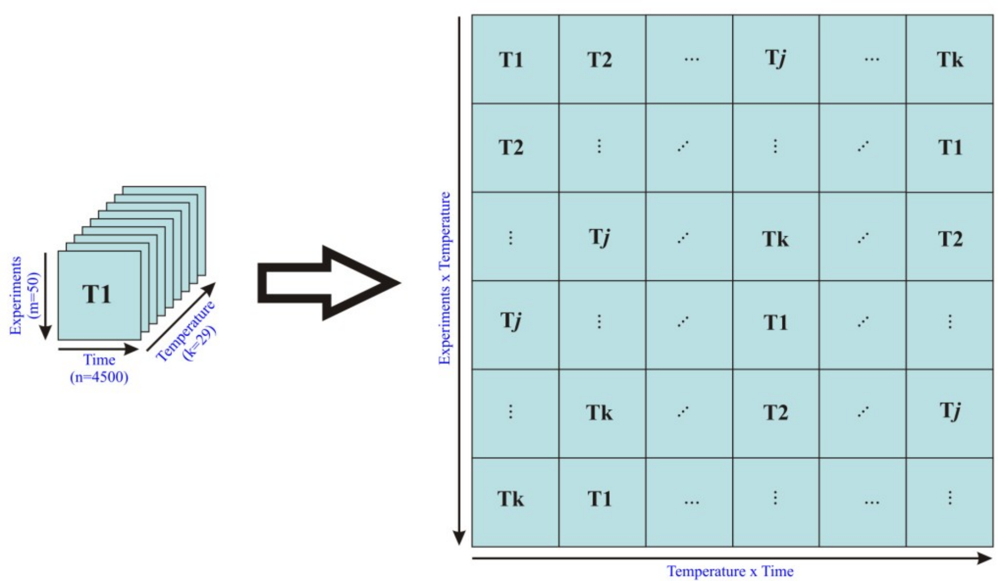

Figure 7. Unfolding of data set in experiment-time direction.

4. RESULTS, ANALYSIS AND CONCLUSIONS

A PCA model was built using the data from the undamaged plate (D0). These data and signals from the damaged plate (D1, D3 and D5) were projected onto the PCA model. The first and second scores of the projected data are shown in figures 8, 10 and 12 for the unfolding in the experiment, time and experiment-time directions respectively. The damage detection indices (T^2*-statistic* and *Q-statistic*) of each experiment are also shown in figures 9, 11 and 13 for the three ways of unfolding. The color represents the temperature at which the experiment was performed and, the shape represents the size of the defect in the plate.

From figure 8, it can be seen that scores using unfolding in the experiment direction can not provide any help in visualizing the different kinds of damage. However, the graph of the damage indices (figure 9) shows how the experiments with the same damage are grouped. In other words, a specific damage appears to be associated with one value of the *Q-statistic*. The T^2*-statistics* does not appear to provide any information. It can also be seen from these two figures that the information on temperature is unavailable, because of the nature of the unfolding.

Analyzing the results of unfolding in the time direction, it is clear that a classification or visualization of defects and/or temperature is not possible. From the score graph (figure 10), it can be seen that some patterns (correlations) exist in these data, although correlations between temperatures were not considered. In fact, this lack of consideration is reflected in the graph of damage indices (figure 11). In the figure, groups of experiments with the same defect and the same temperature can be identified, although many of them are overlapping. In both figures, a visualization of the experiments and temperatures can be performed, but a clear discernment of damages is not available.

Finally, using an unfolding in the experiment-time direction, the distinct damages and temperatures are clearly discernable. From the score graph (figure 12), it is clear that there exists a circular symmetry. The damage can be discerned using the radial distance from the centre of the figure and the temperature by using the angle. More interesting though, is the hysteric behavior in the second score, which was completely unanticiapted. If the plate is heating (increasing the temperature from 35^oC to 70^oC), the second score is negative but, if it is cooling (from 70^oC to 35^oC), this score is positive. In other words, the second score of experiments performed at the same level of temperature (one of them in the heating step and the other one in the cooling step) have the same magnitude but opposite sign. This result is the subject of further study and a physical explanation is sought.

In this case, a clear and accurate visualization and classification can be observed from the graph of damage statistics (figure 13). Here, experiments with the same damage are distinguishable from experiments with other structural condition. The index which provides the information about the damage is the *Q-statistic*.

In conclusion, this paper demonstrates the application of PCA and some of it's extensions in order to isolate the effects that temperature changes (or another environmental and operational variations) produces in structures when damages identification procedures have been implemented. The extensions allows one to study not only correlations between samples of the same signal, but also, correlations between signals, experiments, sensors, temperatures, etc. Another advantage to highlight in the application of the unfolding procedure proposed here, is the fact that information about the environmental and/or operational condition of the experiment is preserved.

ACKNOWLEDGMENTS

This work has been supported by "Ministerio de Ciencia e Innovación" in Spain through the research project DPI2008-06564-C02-02, the visit grant programme "José Castillejo", and the post-doctoral programme "Juan de la Cierva". The authors would like to thank the support of the Technical University of Catalonia and the University of Sheffield. We are grateful also to Dr Boon C. Lee who collected the data in the course of his Ph.D. studies.

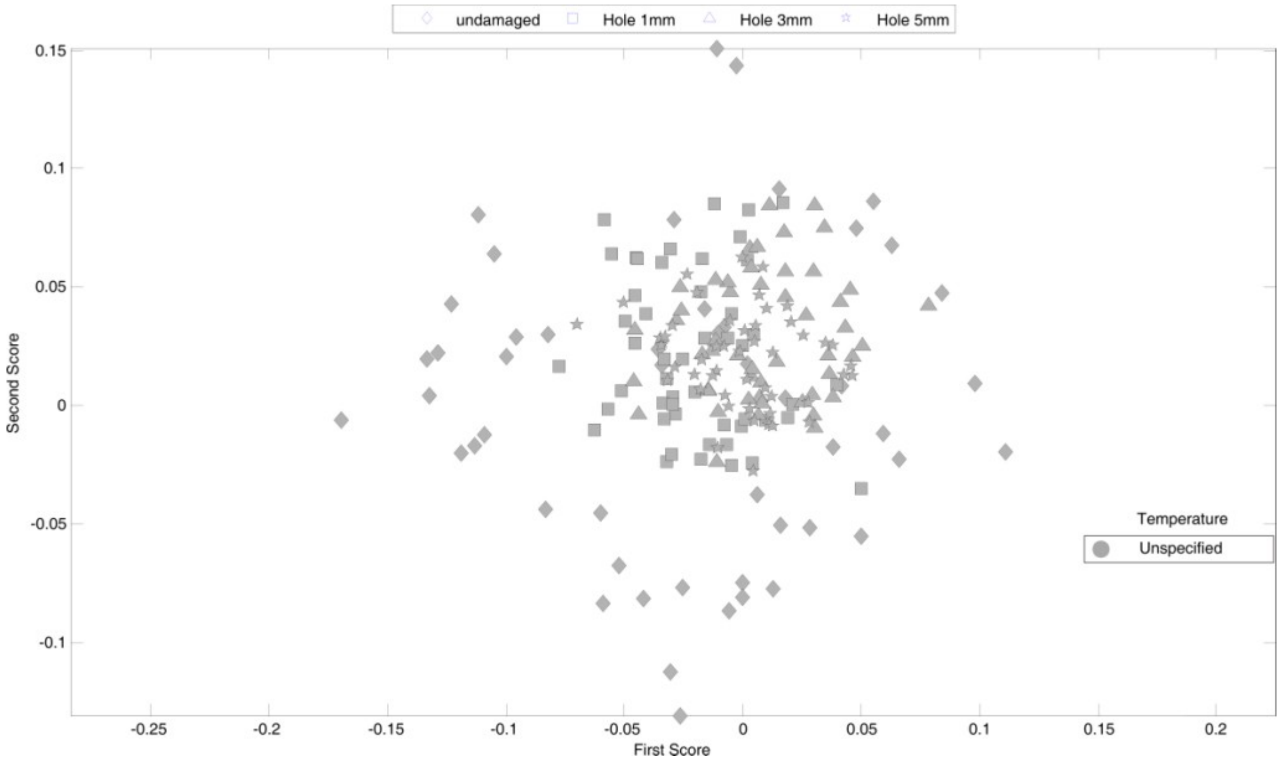

Figure 8. Experiments projected on the PCA model using unfolding in experiment direction.

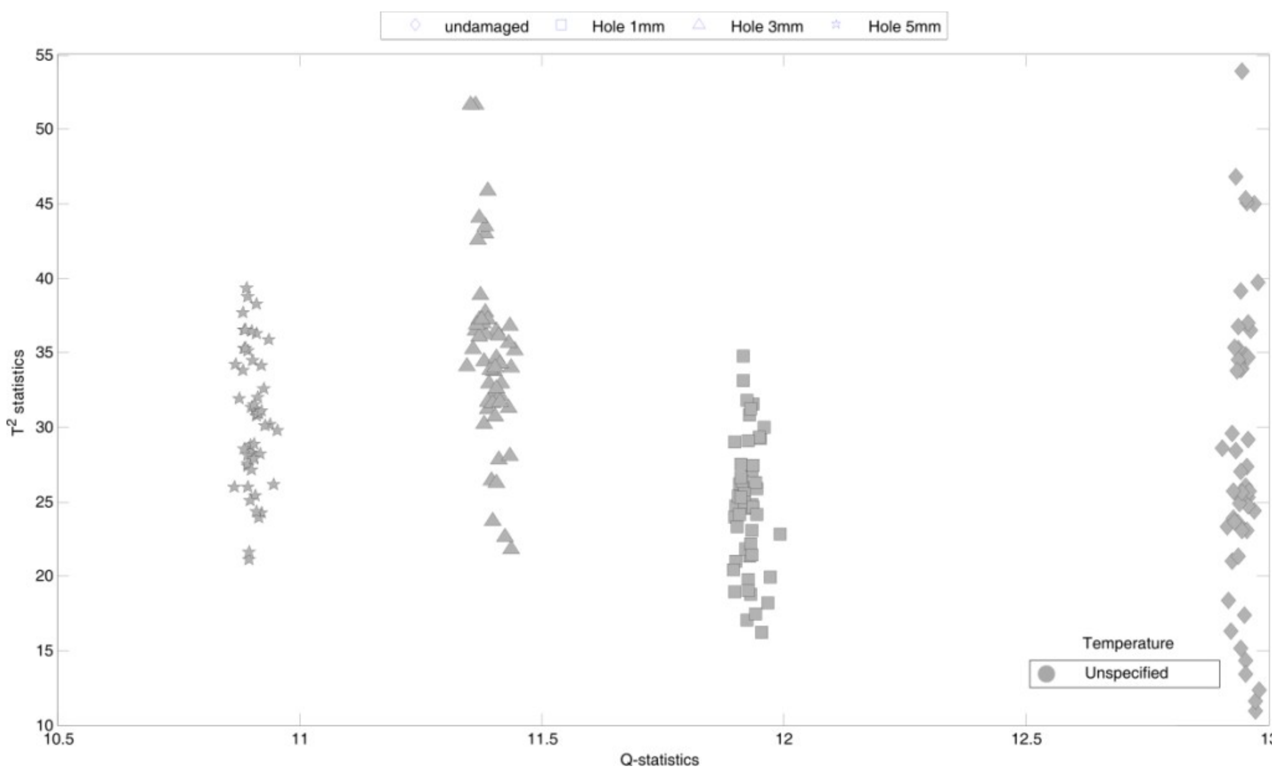

Figure 9. Damage indices of experiments using unfolding in experiment direction.

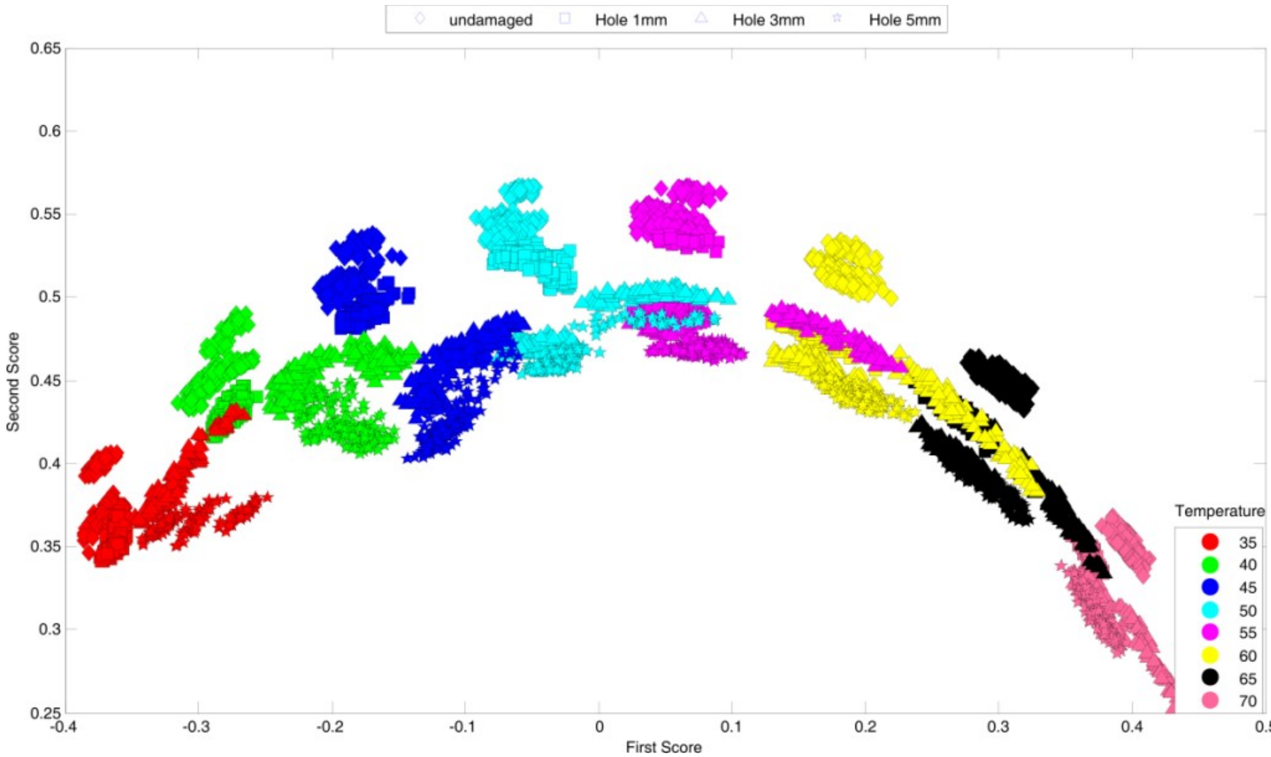

Figure 10. Experiments projected on the PCA model using unfolding in time direction.

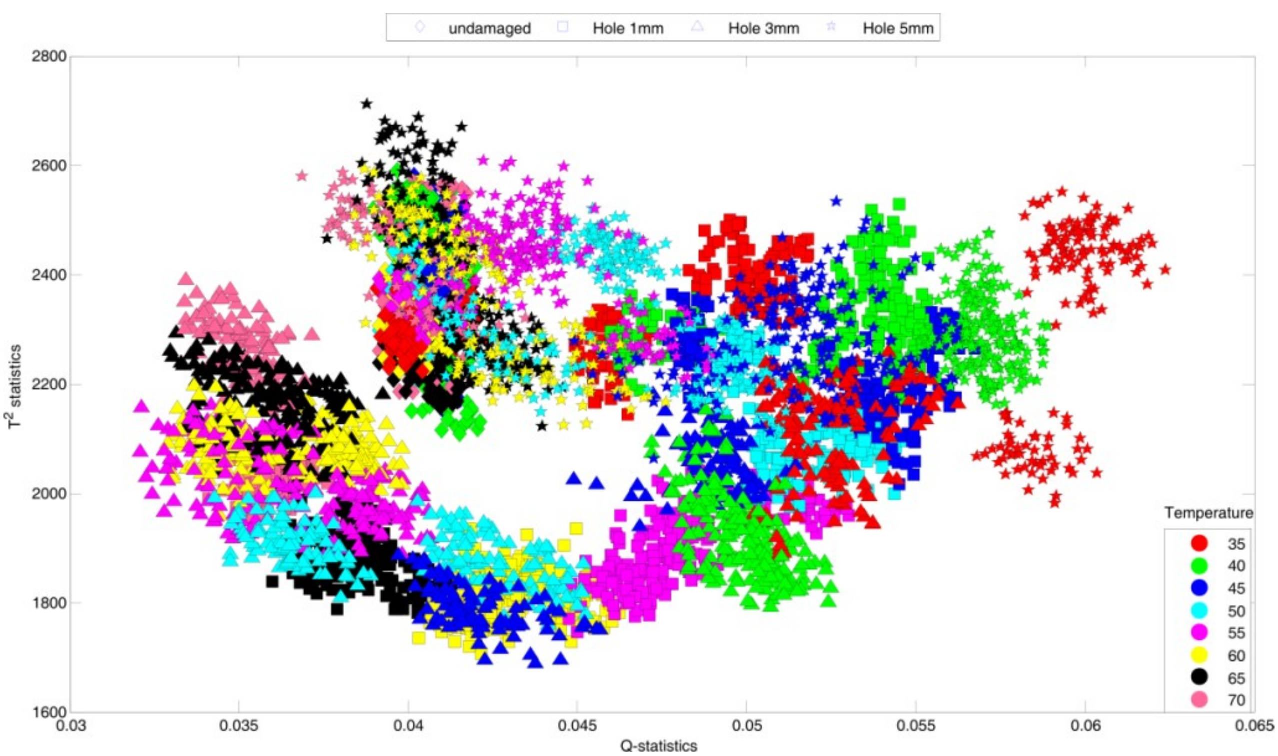

Figure 11. Damage indices of experiments using unfolding in time direction.

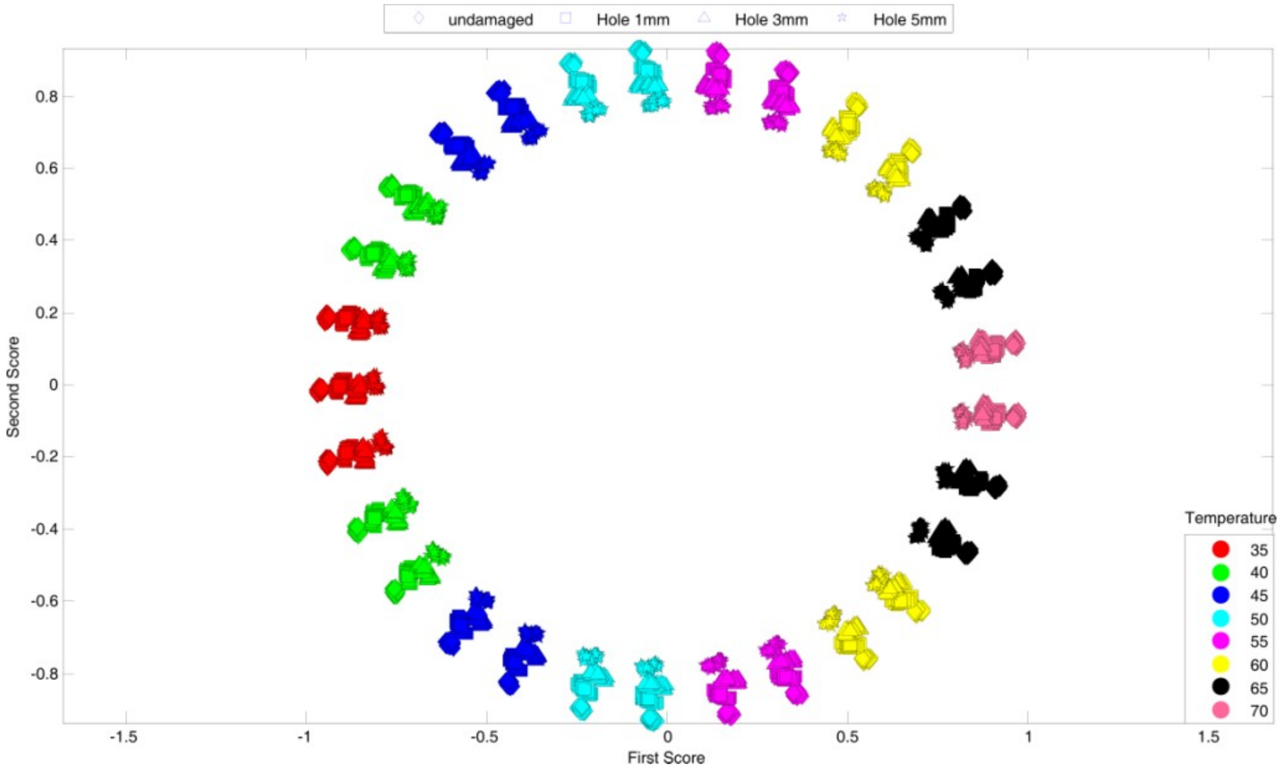

Figure 12. Experiments projected on the PCA model using unfolding in experiment-time direction.

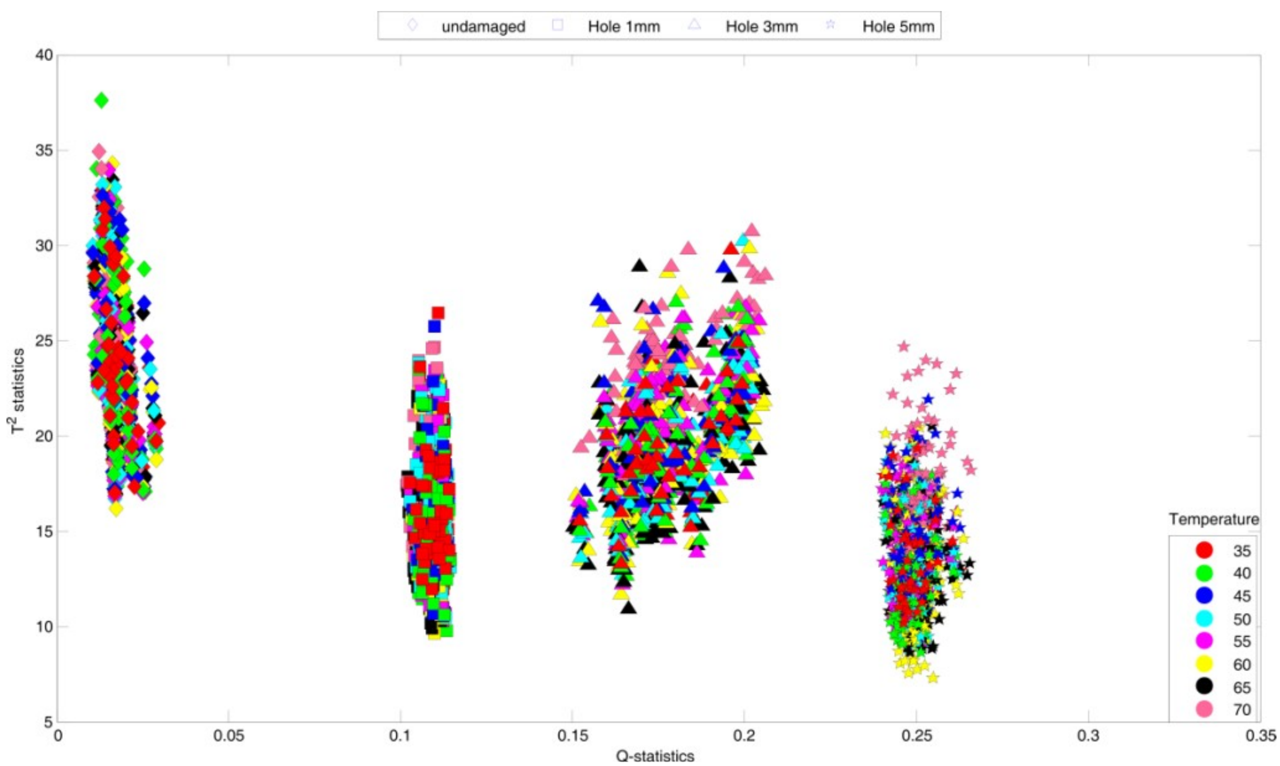

Figure 13. Damage indices of experiments using unfolding in experiment-time direction.

REFERENCES

[1] L. Mujica, J. Rodellar, A. Güemes, and L. López-Diez, "Pca based measures: Q-statistics and t^2-statistics for assessing damages in structure," in *Proceeding of Fourth European Workshop on Structural Health Monitoring*, T. Uhl, W. Ostachowicz, and J. Holnicki-Szulc, eds., pp. 1088–1095, DEStech Publications, Inc., 2008.

[2] B. Lee, G. Manson, and W. Staszewski., "Environmental effects on lamb wave responses from piezoceramics sensors," in *Proceeding of the Modern Practice in Stress and Vibration analysis Conference, 2003.*, pp. 195–202, 2003.

[3] I. Jolliffe, *Principal Component Analysis*, Springer Series is Statistics, Springer-Verlag, second edition ed., 2002.

[4] B. Wise, N. Gallagher, S. Watts, D. White, and G. Barna, "A comparison of pca, multiway pca, trilinear decomposition and parallel factor analysis for fault detection in a semiconductor etch process.," *Journal of Chemometrics* **13**, pp. 379–396, 1999.

[5] S. Qin, "Statistical process monitoring: Basics and beyond," *Journal of Chemometrics* **17**(8-9), pp. 480–502, 2003.

[6] P. Nomikos and J. MacGregor, "Monitoring batch processes using multiway principal component analysis," *AIChE Journal* **40**, pp. 1361–1375, aug 1994.

[7] J. Westerhuis, T. Kourti, and J. MacGregor, "Comparing alternative approaches for multivariate statistical analysis of batch process data," *Journal of Chemometrics* **13**, pp. 397–413, 1999.

[8] L. Mujica, J. Veh, M. Ruiz, M. Verleysen, W. Staszewski, and K. Worden, "Multivariate statistics process control for dimensionality reduction in structural assessment," *Mechanical Systems and Signal Processing* **22**, pp. 155–171, 2008.