ABSTRACT

This paper is focused on the development of new estimators propounding if someone statistical law could estimate or infer a system without damage knowing its reliability. This new measurement considers each experiment, and consequently, each projection to the PCA model as a random variable. An in-depth statistical analysis is performed for SHM. PCA projections are obtained from the undamaged structure (baseline projection). If these projections are considered as the set of possible results (population), then the new projections from the current structure (healthy or not) are defined as random samples. Therefore, the probability distribution of the baseline projection can be found. This new distribution can make an inference about the state of the structure and determine if there is damage in it. Consequently, the relative likelihood of each new projection is determined. If the new projection is strongly related with the population, then the structure is healthy. Otherwise, the relation indicates the damage.

INTRODUCTION

Structural Health Monitoring (SHM) and damage detection methods are nowadays playing a key role for improving the operational reliability of critical structures in several industrial sectors through the integration of novel materials, sensors, actuators and computational capabilities designed to operate automatically in real time. The essential paradigm is that a self-diagnosis and some level of detection and classification of damage is possible trough the comparison of the in-service dynamic time responses of a structure with respect to baseline reference responses recorded in ideal healthy operating conditions.

In previous works, the authors have presented a novel multi-actuator piezoelectric system for detection and localization of damages. The approach combines: (i) the dynamic response of the structure at different exciting and receiving points; (ii) the correlation of dynamical responses when some damage appears in the structure by

---

Department of Applied Mathematics III. Universitat Politècnica de Catalunya. Comte d'Urgell 187, 08036 Barcelona, Spain. {luis.eduardo.mujica, magda.ruiz, francesc.pozo, jose.rodellar}@upc.edu.

using Principal Components Analysis and statistical measures that are used as damage indices; (iii) the contribution of each sensor to the indices, what is used to localize the damage [1-4].

Damage indices, for detection, were calculated analyzing the variability of: (i) the projected data (scores) in the new space of the principal components, (ii) the residual data in the residual subspace and, (iii) a combination of them. In consequence, until now, does not exist only one technique for damage detection. These methodologies have been based in the combination of different techniques presenting on certain features several pros and cons.

Keeping the same paradigm for damage detection, a new methodology for damage detection is developed using statistical estimators. Firstly, from the undamaged structure (baseline projection), a PCA model is calculated. From the model, the score vectors (projection to principal components) are obtained and considered as the full set of results under study (population) [5]. Then, the probability distribution of the baseline projection is found. In this way, an inference about the current structure can be determinate [6]. If the new score is similar to the range obtained from baseline probability distribution then, the structure is healthy. Otherwise, the relation indicates damage. The projection vector will be the random variable that controls the property of the projection of data into a model and its distribution will depend on whether, or not, exists a defect. In this way, these inferences could be used as a new index to determine damage in structures.

The structure used to validate the index is an aluminium plate. Four piezoelectric transducers (PZT's) to generate lambwaves and to detect the time varying strain response data, were distributed over the surface.

STATISTICAL INFERENCE

The main idea in statistical inference is to obtain conclusions from the population (scores), based in observations of a simple sample, looking for the confidence. This sample from an unknown distribution will draw estimators that will describe features in undamaged systems. Before determining these statistical estimators, it is tested if the normal distribution will be satisfactory as a population model (null hypothesis - H_0). In this respect, the goodness-of-fit test procedure should be performed based on, for instance, the chi-square distribution. Thus, if the H_0 – null hypothesis that the distribution of the population is the normal distribution is not rejected, the population can be categorized as a normal distribution. Otherwise, there is the option that scores are not normally distributed (H_1). When the sampling distribution can be considered as normal, a wide variety of hypothesis tests can be performed and therefore some confidence intervals can be determined. An essential requirement is that the sample size has to be greater than or equal to 30.

The goodness-of-fit test procedure based on the chi-square distribution requires

1. A random sample of size n from the population whose probability distribution is unknown.
2. To arrange the n observations in k bins or class intervals.
3. To compute the observed frequency O_i in the ith class interval.
4. To compute the expected frequency E_i in the ith class interval.

Finally, the goodness-of-fit test statistic is:

$$\chi_0^2 = \sum_{i=1}^{k} \frac{(O_i - E_i)^2}{E_i} \tag{1}$$

If the population follows the normal distribution, χ_0^2 has, approximately, a chi-square distribution with $k-p-1$ degrees of freedom where $p=2$ represents the number of parameters of the normal distribution estimated (mean and standard deviation). The hypothesis that the distribution of the population is normal is rejected if the value in equation (1) is strictly greater than $\chi^2_{\alpha,k-p-1}$.

The following procedure can serve as a summary of the goodness-of-fit test based on the chi-square distribution:

1. The variable of interest is the distribution of the population of scores.
2. Null hypothesis: The form of the distribution is normal.
3. Alternative Hypothesis: The form of the distribution is not normal.
4. The confidence coefficient is α=0.05, for instance.
5. The test statistic is
$$\chi_0^2 = \sum_{i=1}^{k} \frac{(O_i - E_i)^2}{E_i} \tag{2}$$
6. Reject the null hypothesis if $\chi_0^2 > \chi^2_{\alpha,k-p-1}$

The goodness-of-fit test is performed for both: the population of scores from the baseline data and the population of scores from the current structure (damaged or not).

At this point, with the sample of undamaged results, the sample mean and the sample standard deviation is calculated, which are then considered as the population mean (μ) and the population variance (σ) , respectively. Let us consider $X_1, X_2, \ldots, X_n$ (scores from experiments using the current structure: damaged or not) as random sample from a population with normal distribution with unknown mean (μ) and unknown variance (σ^2). Considering the case of hypothesis testing on the mean, the random variable described by equation (2) has a *t*-student distribution with *n*-1 degrees of freedom.

$$T = \frac{\overline{X} - \mu}{S/\sqrt{n}} \tag{3}$$

where $\overline{X}$ is the sample mean and *S* is the sample standard deviation. In this way, the next hypotheses can be considered:

$$H_0 : \mu = \mu_0$$
$$H_1 : \mu \neq \mu_0 \tag{4}$$

From the Fisher theorem and as population follows a normal distribution then:

$$\frac{\sum_{i=1}^{n} \left(X_i - \overline{X}\right)^2}{\sigma^2} = (n-1)\frac{S^2}{\sigma^2} \mapsto \chi^2_{n-1} \tag{5}$$

Then,

$$\frac{S^2}{\sigma^2} \mapsto \frac{\chi^2_{n-1}}{(n-1)} \tag{6}$$

Replacing in equation (3):

$$T = \frac{\overline{X} - \mu}{\frac{\sqrt{\frac{\sum_{i=1}^{n}\left(X_i - \overline{X}\right)^2}{n-1}}}{\sqrt{n}}} = \frac{\frac{\frac{\overline{X} - \mu}{\sigma}}{\sqrt{n}}}{\sqrt{\frac{\sum_{i=1}^{n}\left(X_i - \overline{X}\right)^2}{\sigma^2 (n-1)}}} \mapsto t_{n-1} \tag{7}$$

Therefore, the random variable follows the *t*-student distribution subject to the constraint given by the confidence level with *n*-1 degrees of freedom. In this way, μ is estimated by means of the following confidence interval:

$$P\left(-\delta \le T \le \delta\right) = P\left(-\delta \le \frac{\overline{X} - \mu}{\frac{S}{\sqrt{n}}} \le \delta\right) = \gamma \tag{8}$$

Where γ corresponds to the confidence level and $\pm\delta$ are upper and lower limits. Solving for μ, the confidence interval is:

$$\left[\overline{X} - \delta\frac{S}{\sqrt{n}}, \overline{X} + \delta\frac{S}{\sqrt{n}}\right] \tag{9}$$

EXPERIMENTAL SET-UP

Methodology proposed here is tested in the small aluminium plate (25cm × 25cm × 0.2cm) shown in Figure 1. Four Piezoelectric transducer discs (PZT's) are attached on the surface. Each PZT is able to produce a mechanical vibration (lambwaves in a thin plate) if some electrical excitation is applied (actuator mode). Also, PZT's are able to detect time varying mechanical response data (sensor mode).

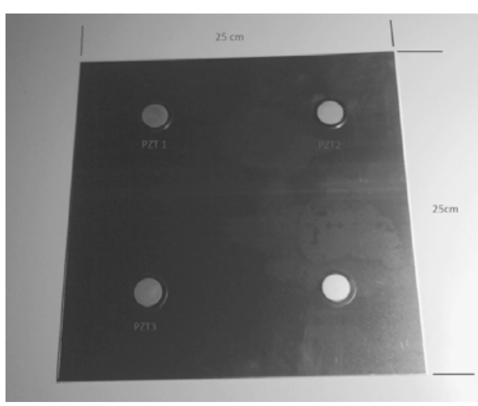

Figure 1: Aluminium plate: Dimensions and PZT's location.

In previous work [1-4] it was conclude that the farther is the damage from the actuator, the more difficult is its detection. Therefore, to detect damage on a larger area and being useful the fact that PZT's ca be used as much as actuators as sensors, the experiment to assess the structure is performed in several phases as such as a Multiactuator Piezoelectric System. In every phase, just one PZT is used as actuator (exciting the plate) and the others are used as sensors (recording the dynamical response). 100 experiments were performed using the healthy structure. Nine damages were simulated adding different masses at different locations. By each damaged structure, 50 experiments were achieved. The excitation is a sinusoidal signal of 112 KHz modulated by a Hamming window. An example of excitation signal and dynamical responses are sown in Figure 2

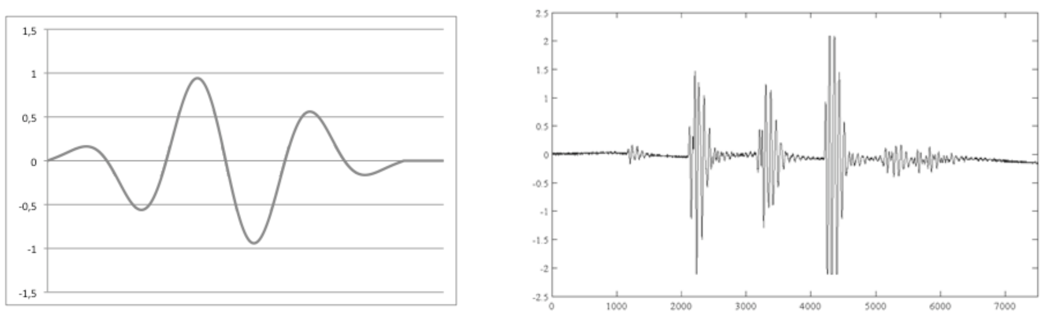

Figure 2: Excitation signal and dynamical response.

METHODOLOGY

The goal of the methodology proposed here is to find estimators that describe undamaged model using statistical inference. For this, PCA latent variables (specifically, scores vectors) are obtained representing the random experiment. The use of statistical inference is justified due to all possible scores obtained for a specific condition of the structure can be called as population. Each score has the same probability of occurs, and a new result does no influence the next one. Each score corresponds to an experiment, containing information from all sensor measures with different levels of contribution and, with an extra benefit, the noise is removed.

The PCA model is performed using the unfolded sensor matrix that contains data from 50 experiments with the undamaged specimen. The five scores obtained are used as simple sample baseline of the undamaged population. Of them, the sample mean ($\bar{X}$) and standard deviation (S) are calculated. With a 95% of confidence interval and the *t*-student distribution, the statistical estimator parameter is found. This estimator indicates the range where the population mean (μ) is located. In consequence, some hypothesis can be proposed, for example, is it the current structure that undamaged range (H_0)? For testing the hypothesis, 50 experiments by each nine different damaged specimens and 50 with the intact structure are projected into the PCA model. In this way, new scores are the current structure stage sets. From each set, the mean is calculated and projected onto the statistical estimator (population mean range). The whole methodology is showed in Figure 3.

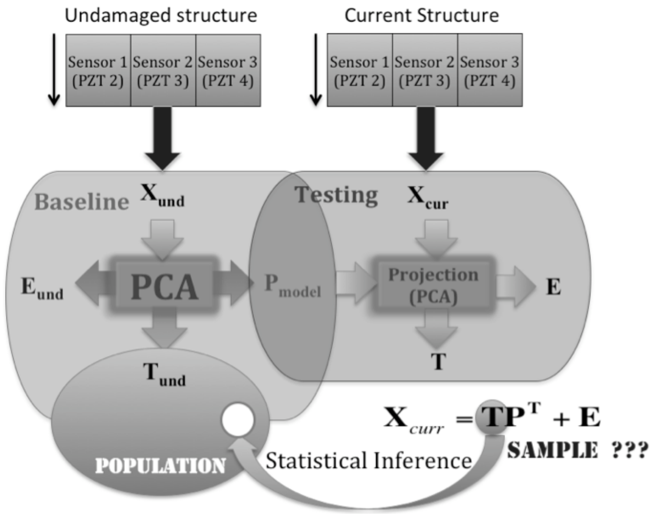

Figure 3: Methodology for statistical inference process

DAMAGE DETECTION

Unlike previous works, here, the scores and T^2 and Q indices (not shown) do not let differentiation between undamaged and damaged experiments, much less between the several damaged specimens due to many causes as: structural material, the damaged type, frequency and amplitude of the excitation signal, among others. To review what is happening, scores are agrupped. Figure 4 shows all the experiments for four scores, where again, the differentation between different types of experiments is not possible.

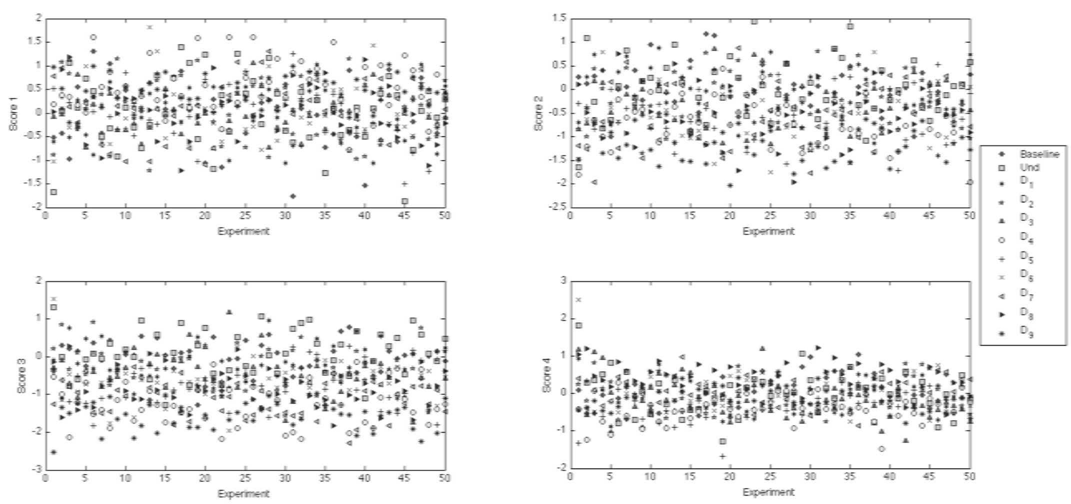

Figure 4. Experiments by each score

Before contrasting the hypothesis, behavior of each damaged and undamaged is tested using the box plots. From Figure 5 it can be seen that the mean of the first score of damages 1, 2, 5 and 7 could be confused with the baseline mean. Besides, if the quartile 1 (Q1) and quartile 3 (Q3) are used like limits, then, damages 3, 5, 6 and 8 could be detected as undamaged specimens. Better differences can be appreciated for the means of second, third and fourth score. But the error occurred again taking into account the Q1 and Q3.

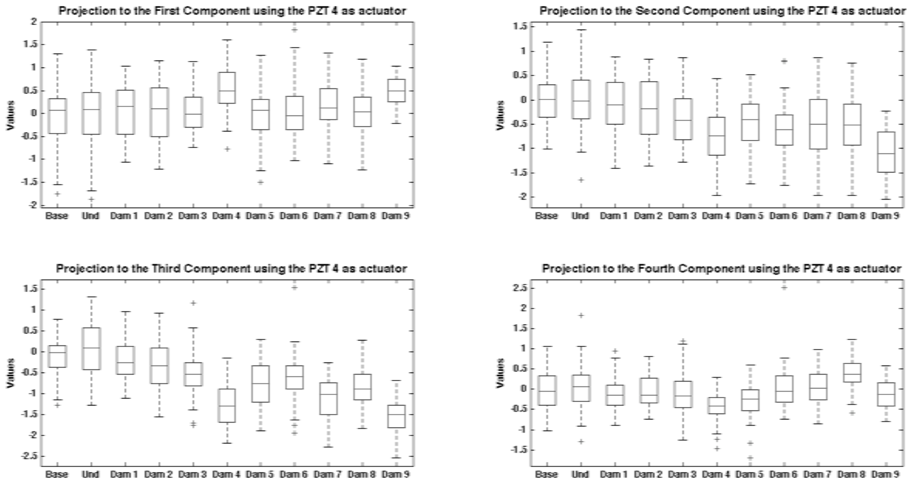

Figure 5: The box plot for the scores

Immediately after, the goodness-of-fit test based on the chi-square distribution is used to determine if the simple samples (all scores) are normally distributed. The p-value is the probability, under assumption of the null hypothesis, of observing the given statistic. In this way, H_0 is rejected with 5% of significance (therefore H_1 is accepted) if the *p*-values of the scores are equal or lesser than 0.05. Otherwise, the samples (scores) are considered with normal distribution. Figure 6 shows the *p*-value of simple samples from experiments when the PZT 2 is used as actuator. From them, it can be seen that the fourth score cannot be used as feature to detect damages because the baseline variables does not have normal distribution.

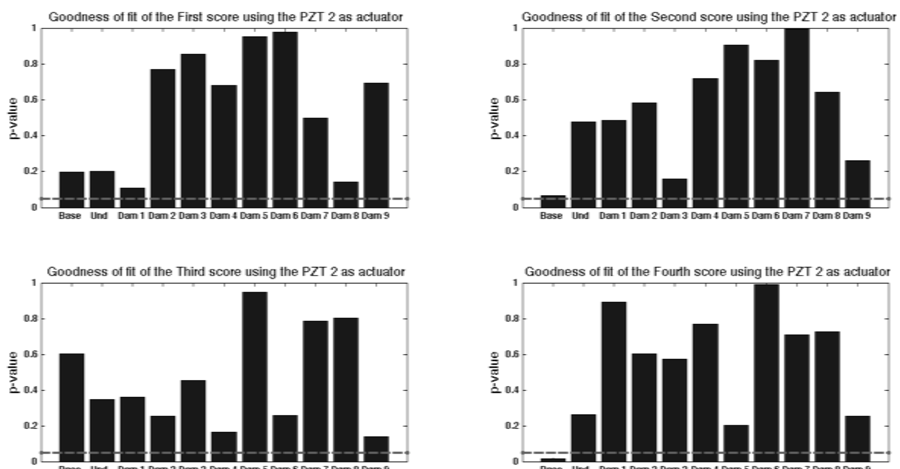

Figure 6: The goodness-of-fit test

This new distribution can make an inference about the state of the structure and determine whether there is damage in it. Consequently, the relative likelihood of each new projection is determined. If the new projection is strongly related with the population, then the structure is healthy (H_0). Otherwise, some damage is present in the structure (H_1). Table 1 shows a summary of the results for detection of damage. Error type I means that the hypotheses was wrongly rejected (a damage is detected in a healthy structure). Error type II means that the hypotheses was wrongly accepted (damage was not detected

	PZT 1 as actuator		PZT 2 as actuator		PZT 3 as actuator		PZT 4 as actuator	
	H_0	H_1	H_0	H_1	H_0	H_1	H_0	H_1
No reject H_0	4	4	5	2	4	3	4	21
Reject Ho and accept H_1	1	41	0	43	1	42	1	24

Error type I	(hatched)
Error type II	(shaded)

Table 1: Inference about the state of the structure

CONCLUSIONS

Although in SHM field, several researchers has focusing their works in applications of PCA and statistical tools as inference, authors have not knowledge of works that use PCA as statistical model combine with inference to calculate a damage index. Results show a high accuracy of the methodology for detecting defects. Besides, it help us to assess which experiments should not taken in account, and therefore must be repeated.

In this work, inference over statistical means were performed, however, results encourage to authors to keep studying inferences over standard deviations, difference between means, etc. On the other hand, the population (means of scores) has been considered by separately, it means that one diagnostic is performed analysing one score, another diagnostic by using other score. For the next step, authors are considering to do multivariate inference, in this way, one diagnostic can be obtained from all scores.

ACKNOWLEDGEMENTS

Authors thank to Ministerio de Ciencia e Innovación of Spain through the research projects DPI2008-06564-C02-01/02. We are also grateful also to Ph students Diego Tibaduiza, Fahit Gharibnezhad and Maribel Anaya who has collected the data.

REFERENCES

1. D.A. Tibaduiza, L.E. Mujica, A. Güemes, and J. Rodellar. Active piezoelectric system using PCA. In European Workshop on Structural Health Monitoring, 5th edition. Sorrento-Italy. 2010.
2. L.E. Mujica, D.A. Tibaduiza, and J. Rodellar. Data-driven multiactuator piezoelectric system for structural damage localization. In 5th World Conference on Structural Control and Monitoring. Tokio-Japan., July 2010.
3. L.E. Mujica, M. Ruiz, A. Güemes, and J. Rodellar. Contribution plots on PCA based indices for damage identification on structures. In Proc of the 4th. ECCOMAS thematic conference on smart structures and materials. Porto, Portugal, 2009.
4. L.E. Mujica, A. Fernández, A. Güemes, J. Rodellar. Q-statistic and T2-statistic PCA based measures for damage assessment in structures. Structural Health Monitoring, an International Journal. In press. 2011
5. Montgomery, D. C.; Runger, G. C.; "Applied Statistics and Probability for Engineers"; Editorial Wiley, 2006
6. F. Pozo, N. Parés, Y. Vidal, F. Mazaira; "Probabilitat i estadística matemàtica"; Iniciativa Digital Politècnica, (In catalan) 2010.