Preliminary robustness analysis of a Structural Health Monitoring
PCA-based algorithm

Francesca M. Pisano\textsuperscript{1}, Monica Ciminello\textsuperscript{1}
\textsuperscript{1} C.I.R.A. – Italian Aerospace Research Centre
Capua, Caserta, 81043, Italy
f.pisano@cira.it
m.ciminello@cira.it

Abstract

In the present work an application of the Principal Component Analysis (PCA) technique to detect and locate skin-stringer debonding, over stiffened composite panels, relaxing solicitation conditions for healthy and current structure states is presented. Herein the PCA-based algorithm has been implemented using R, the most widespread free programming environment for statistical computing. The algorithm calculates the most representative healthy system dynamics, then calculates the projections of the healthy and current observed data over a few of the healthy first principal components and compares the results to identify the most significant deviations. Outliers are sought using the box-and-whisker plots and are intended to be possibly failure warnings to be verified. Outliers are depicted on a graph to easily derive location information based on sensor position. The algorithm has been applied to experimental data collected from double stiffener composite panels. Basing on the skin-stringer debonding locations from NDI, algorithm effectiveness and robustness are shown to be satisfactory.

1. Introduction

The diagnostics of the state of the materials that make up a structure is mainly based on the application of statistical methods to the analysis of data collected repeatedly over time and in multiple experiments (1, 2). Assuming as reference the response to solicitations of the structure in healthy conditions, intended as absence of damage or effective operational conditions, any "anomalous" response is sought. Recently, the use of distributed optical fibers has been investigated: they offer advantages compared to traditional sensors such as weight and size reduction, non-electrical nature. Moreover, high density of sensors is supplied by optic fibers, recording high volume of data even for short monitoring periods and reducing the risk of neglecting a damage in a position not sufficiently covered by traditional sensors. However, new algorithms to get local damages information from strain measurements gathered by fiber optic sensors are needed (3).

In order to increase the probability of detection of any damage, all the answers recorded by the fiber optic sensors should be analyzed at the same time, therefore methods and tools for multivariate statistics (4) can be applied.

One of the statistical techniques applied to the Structural Health Monitoring (SHM) of aeronautical structures is the Strain field Pattern Recognition based on Principal Component Analysis (PCA) (5). PCA is a multivariate statistical technique aimed at
identifying the main dynamics of a system, removing redundant or noisy data. The PCA is used to reduce the size and complexity of data samples in order to discover patterns and trends that would otherwise remain hidden in the data. In this sense, an effective data-driven model can be built when the adequate volume of information is available. PCA theoretical background is briefly outlined in (6, 7). PCA is based on the linear transformation of observed physical variables, eventually correlated, in a set of uncorrelated variables, called Principal Components.

First PCA applications to the aeronautical structure health state analysis are discussed in (6, 7, 8), based on strain data collected by four piezoelectric transducer discs, respectively on an aluminum plate and an aircraft turbine blade. In (8), variability of the data projection within the residual subspace and variability of the projected data in the new space of the principal components are evaluated using Q-index and Hotelling’s $T^2$ - statistic statistical indices for fault detection; some localization solutions basing on contribution methods are presented. In (6) a damage classification approach based on Self-Organizing Maps (SOM) is added to previous work (8). In (7) a data driven model based on PCA using data strain projections onto principal component space – so called scores – and hypothesis testing of scores for fault detection is presented, aiming at accepting or rejecting the null hypothesis “the multivariate random sample of the structure to be diagnosed is distributed as the baseline projection”.

In (3) Q-index and Hotelling’s $T^2$ - statistic statistical indices are used for fault detection based on the PCA, analyzing strain data gathered through 32 fiber optics sensors, bonded on an aluminum beam solicited under some load conditions and magnitudes; also some strain difference-based algorithms are presented. In (9) strain data is collected from 24 up to 36 fiber optic sensors applied to wind turbine blades and an Unmanned Aircraft Vehicle (UAV) wing suitably solicited and is firstly processed using SOM and then Q-index and $T^2$ – statistic are calculated to detect damages. Finally, in (10, 11) optical fibers are bonded within the wing spars on a High Altitude Long Endurance Unmanned Aerial Vehicle fleet and a graphical comparison of PCA scores related to only the first principal component is used to detect structural degradations.

In the present paper a PCA-based algorithm is proposed for skin-stringer debonding detection and location and is tested over experimental data related to two impacted carbon-fiber panels. Strain data collected from solicited panels before impacts is intended as panel healthy conditions and is used as reference. In the experimental set-up, solicited panel strains are recorded by hundred of fiber optic sensors. In this work, box-and whisker plot are proposed as statistical methods for potential strain “anomalies” detection, intended as damage warnings to be verified. Furthermore, a graphical approach for damage localization is presented. Moreover, within the cited papers, aeronautical structures in both healthy and current conditions are subjected to the same loading solicitations.

In the present work, for both panels the identical loading solicitation constrains for healthy and current structure states have been softened. Available experimental data have allowed to test the proposed algorithm verifying robustness and efficacy when it is applied to real data, likely collected within experimental trials, that rarely can satisfy with such theoretical bonds. Finally, SHM analysis results have been compared to Non Destructive Inspections (NDI) carried out by means of the C-scan.
2. Proposed methodology

The proposed PCA-based algorithm detects statistically significant differences between the healthy and current states of each panel. It looks for possible statistical outliers among the score differences by means of the box-and-whisker plots. These graphs are used to effectively represent the information related to the positional distribution of projection differences of recorded strain onto a reduced selection of the principal components. The values of such differences that are statistically "anomalous" can potentially represent warnings of the presence of failures. In (Figure 1) the work-flow exemplifies the proposed approach.

Moreover the robustness of the algorithm has been investigated softening constraints on loading solicitations. A time-variable load has been used to solicit the first panel, but the related time law is not known for healthy or current conditions; instead, a single static load was used to solicit the second panel in both healthy and current conditions. Finally, the score differences on a few of selected principal components are depicted according to the position of the fiber optic sensors, in order to simplify the location of possible damages.

3. Experimental set-up

Specimen have been subjected to a single 50 J impact in two different experiments: the impact is located, respectively, at the center of the skin and under a stringer semi-cap. Moreover, two different configurations for solicitation and sensoring have been accomplished for the two carbon-fiber panels.

Panel 1 has been clamped on all sides and has been flexing stressed by means of a load located at the center of the skin (Figure 2 (a)). The load has been increased linearly in time up to 800 N. The optic fiber has been bonded along the stringer perimeters. In (Figure 2 (b)), the proposed fiber layout is yellow-highlighted and the fiber direction is indicated by red arrows. Fiber portions bonded on the panel are those included between the points indicated as distance (cm) from the fiber starting point.
Panel 2 has been clamped on just one side and has been flexing stressed by means of a 2000 gr static load located at the skin free edge (Figure 3 (a)). In this case, the optic fiber has been bonded along the stringer inner semi-cap and in the between of the two stringers. At each bending of the fiber, the tangency at the stringer edges has been assured. In (Figure 3 (b)), the proposed fiber layout is yellow-highlighted and the fiber direction is indicated by red arrows. Fiber portions bonded on the panel have width corresponding to indicated distances (cm) from the fiber starting point.

The fiber optic, 1.2 m long, has been interrogated by using LUNA OdisiB (Figure 4).

The acquisition rate of 250 Hz allows for a sensing gauge of 5 mm, providing 468 sensing points.

4. Data analysis

In the present work, CRISP-DM (CRoss-Industry Standard Process for Data Mining) methodology (12) has been used as reference to carry out the data analysis activities. CRISP-DM supplies six phases and just the modeling phase is specifically defined for Data Mining-based solutions. In authors’ experience, all the other phases are equally effective when applied to more general data analysis contexts.
CRISP-DM phases applied at the present work have been:
• Data Understanding: CRISP-DM second phase is focused on data gathering and data exploration in order to identify data quality issues and to formulate hypothesis about the hidden information.
• Data Preparation: CRISP-DM third phase is focused on data selection and processing to provide the dataset to be analyzed by the algorithm. It includes tasks such as record/table/attribute selection, data cleaning and data transformation.
• Evaluation: CRISP-DM fifth phase is focused on built model and obtained results evaluation.

4.1 Data Understanding

Experimental strain data coming from the two panel solicitations are supplied as pairs of files in .txt format: the former file is related to the panel healthy state, the latter to the panel current state. Each input textual file includes four header lines that specify registration starting date and time, strain unit of measurement (microstrain) and position of sensors as distance with respect to the starting point of the optical fiber (meters). Strain records recorded by all the sensors at an instant are aligned along rows, strain data recorded by a sensor during the whole registration are ranged along columns. A synthetic overview of the number of sensors, the registration duration and the percentage of missing data is provided in (Table 1) for the couple of source data of the two panels.

<table>
<thead>
<tr>
<th>Panel</th>
<th>Total sensors</th>
<th>Registration duration (s)</th>
<th>Missing data (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Panel 1 – healthy state</td>
<td>437</td>
<td>00:58</td>
<td>4.8 %</td>
</tr>
<tr>
<td>Panel 1 – current state</td>
<td>437</td>
<td>01:43</td>
<td>0.008%</td>
</tr>
<tr>
<td>Panel 2 – healthy state</td>
<td>444</td>
<td>00:13</td>
<td>0.017%</td>
</tr>
<tr>
<td>Panel 2 – current state</td>
<td>444</td>
<td>00:06</td>
<td>0.16 %</td>
</tr>
</tbody>
</table>
4.2 Data Preparation

The input files have been appropriately prepared for subsequent PCA-based analysis activities. Format transformations have been performed, non-relevant information has been removed, missing data have been replaced and derived fields have been added. In each input file of the two panels, the missing data have been processed by applying the replacement approach with average deterministic imputation. Moreover, some columns in panel 1 corresponding to sensors indicated by the domain expert have been removed from the input files. That is because optical fiber, about 1 m long, has been bonded on the stringer semi cap making a one-way alternating path (it is a 1D sensor). So, after the fiber optic sensors were bonded on the first semicap, the fiber was curved in order to reach the end of the other semicap, and so on. Since the sections of transition from one structural element to the other one have not been bonded, they do not measure any structure stress, and it has been considered possible to neglect the data coming from sensors belonging to these fiber portion. Finally, algorithm implementation has required to prepare data so that the number of random vectors (strain records) do not exceed the number of statistical units (sensors); for this reason an approach for subdividing the input datasets has been applied (Table 2).

Table 2. Two panel input datasets subdivision. Total number of subsets, dimension of each subset (number of rows x number of columns)

<table>
<thead>
<tr>
<th>Panel</th>
<th>Total subsets</th>
<th>Subset dimension</th>
</tr>
</thead>
<tbody>
<tr>
<td>Panel 1</td>
<td>9</td>
<td>272x272</td>
</tr>
<tr>
<td>Panel 2</td>
<td>2</td>
<td>444x212</td>
</tr>
</tbody>
</table>

5. Proposed algorithm

In this paper, an approach to identify potential “anomalous” deformations is proposed based on the analysis of the scores, i.e. the contributions of the projections onto the principal component space along the statistical units of the sample. Strain data collected by all the sensors at the same time stand for random vectors and the sensors represent the statistical units of the PCA model. In order to effectively reduce the problem complexity, just a few of the first principal components are selected as those able to explain at least the 80% of the total variance. Then the score differences among the healthy and current state strain data onto the selected principal components are calculated. Diff represents the score difference related to the i-th principal component. Then possibly outliers are sought and therefore, higher is the outlier score difference corresponding to a statistical unit of the sample, higher is the contribution of the corresponding sensor to healthy and current strain data discrepancy. In end, being the sensor positions fixed, hypothesis about the damage location can be derived from “anomalous” differences among the scores for strain data related to, respectively, the structure healthy and current states. In (Figure 5), the main steps of the proposed algorithm are sketched.
The algorithm has been implemented using R (13), a widespread programming environment for statistical computing.

5.2 Algorithm outputs

The algorithm provides the following outputs:
- Standard Deviation and Percentages of the Total Variance for selected healthy state principal components.
- Statistical indices (average, minimum, maximum) of the score differences among the healthy and current state strain data projections onto the selected principal components.
- Possible outliers identified by means of the score difference box-and-whisker plot.
- Plot of the score differences, highlighting the outliers, with respect to the corresponding sensor positions.

6. Results evaluation

Applying the proposed algorithm to the experimental data described in section 3, prepared as discussed in section 4.2, just up to three principal components have been needed to take into account over than the 80% of total variability of the deformation dynamics. In most cases only the first principal component variance has exceeded the 90% of the total variance.

For all cases, the score difference has average of magnitude $10^{-16}$, while the maximum-minimum range is of the order of some tens of microstrain, suggesting the presence of some anomalous strain data deviations.

In order to evaluate algorithm results, NDI outcomes, obtained by means of the C-scan, have been used. In the following the method of comparison among algorithm and NDI results is outlined.

In (Figure 6) and (Figure 8), in order to facilitate algorithm result evaluations, different plot portions are highlighted as follows:
- Red Boxes represent the stringer four semi-cap extensions.
- Green Boxes represent skin-stringer debonding extensions as revealed by the C-scan.
- Red circles represent the PCA score deviation outliers.

In (Figure 6) and (Figure 8), the x axis represent the sensor position with respect to the optic fiber starting point expressed in millimeters (mm), the y axis represent the score...
differences for the selected principal component expressed in microstrain (µε). It is worth to underline that PCA scores are no longer physical variables.

6.3.1 Panel 1

In (Figure 6) the score deviation outliers related to the first three principal components are compared to the NDI outcomes.

![Figure 6](image)

FIGURE 6. Comparison among algorithm and NDI results. Score deviation outliers related to the first three principal components are shown.

In (Figure 6), some warning zones of the panel can be noticed, possibly related to structural damages. Results are outlined in (Table 3).

<table>
<thead>
<tr>
<th>Table 3. Panel 1 PCA-based algorithm results</th>
</tr>
</thead>
<tbody>
<tr>
<td>Damage extension prediction [start; end] [mm]</td>
</tr>
<tr>
<td>Left Stringer</td>
</tr>
<tr>
<td>Right Stringer</td>
</tr>
</tbody>
</table>

C-Scan analysis have been carried out to validate the proposed algorithm and related outcomes are represented in (Figure 7), revealing panel damages as indicated in (Table 4).

<table>
<thead>
<tr>
<th>Table 4. Panel 1 C-scan analysis results</th>
</tr>
</thead>
<tbody>
<tr>
<td>Debonding extension [start; end] [mm]</td>
</tr>
<tr>
<td>Left Stringer</td>
</tr>
<tr>
<td>Right Stringer</td>
</tr>
</tbody>
</table>
6.2 Panel 2

In (Figure 8) the score deviation outliers related to the first principal component for the subsets introduced in section 4.2 (Table 2) are compared to the NDI outcomes. In (Figure 8), some warning zones of the panel can be noticed, possibly related to structural damages. Results are outlined in (Table 5).

<table>
<thead>
<tr>
<th>Damage extension prediction [start; end] [mm]</th>
<th>Left Semicap</th>
<th>Right Semicap</th>
</tr>
</thead>
<tbody>
<tr>
<td>Left Stringer</td>
<td>-</td>
<td>[60.9; 80], [125.5; 146.4]</td>
</tr>
<tr>
<td>Right Stringer</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>

C-Scan analysis (Figure 9) reveal panel damages as indicated in (Table 6). The left stringer suffers from debonding for almost half of the its whole length, while the right stringer is fully healthy.

<table>
<thead>
<tr>
<th>Debonding extension [start; end] [mm]</th>
<th>Left Semicap</th>
<th>Right Semicap</th>
</tr>
</thead>
<tbody>
<tr>
<td>Left Stringer</td>
<td>-</td>
<td>[50; 160]</td>
</tr>
<tr>
<td>Right Stringer</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>
Figure 8. Top: first subset score deviation outliers; Down: last subset score deviation outliers.

Figure 9. C-scan: (a) left stringer; (b) right stringer.
7 Conclusions

In the present work a new approach to skin-stringer debonding detection and localization has been proposed based on healthy and current state strain data statistically significant deviations by means of PCA and box-and whisker plots. Those possible deviations can be intended as damage warnings to be verified. The proposed score difference analysis allows to locate possible damages investigating which sensors mostly contribute to the outliers among healthy and current state strain data projection differences on a few of selected principal components. Moreover, the identical loading solicitation constraints to healthy and current structure state, applied in literature, have been softened and preliminary algorithm robustness has been tested.

The PCA-based algorithm is successfully able to identify sensors featured by high strain variability, typically corresponding to start or end points of debonding zones. Outliers outside of debonding zones still correspond to strain high variability sensors, but related to fiber bending or transition zones from the bonded fiber to the free fiber where a stress concentration is generated.

In the next future, the proposed algorithm is planned to be improved for full characterization to up 20 aircraft panels.

Acknowledgements

The authors wish to thank their colleagues Antonio Concilio, Umberto Mercurio, Fulvio Romano, and Gianpaolo Romano for their management and technical assistance in supporting the presented research. This work has been carried out inside the project Smart Airframe (SMAF) within the National Aerospace Research Programme (PRORA).

References

7. Pozo F., Arruga I., Mujica L.E., “A structural damage detection indicator based on principal component analysis and multivariate hypothesis testing over scores”,

11


