Structural Health Monitoring (SHM) as a multivariate outlier detection problem

Tie-rods case study

Tommaso Bocchietti

Date to be defined

Politecnico di Milano

Agenda

1. Problem statement
2. Proposed solutions
   - Mahalanobis Square Distance (MSD)
   - Principal Component Analysis (PCA)
3. Results
4. Conclusions

Figure 1: Steel tie-rods connect the buttresses of the Cathedral of Saint Peter of Beauvais in France.

Problem statement

Disturbances effect of environmental condition on SHM

In case of axial-load beams (tie-rods), studies has highlighted that **temperature variations can cause greater changes to structural vibration than the presence of damage itself**.

The transverse vibration in a tensioned beam is described by the following equation[1]:

$$w(\xi, t) = [A \sin(\gamma_1 \xi) + B \cos(\gamma_1 \xi) + C \sin(\gamma_2 \xi) + D \cos(\gamma_2 \xi)]\, E \cos(\omega t + \phi) \tag{1}$$

Where:

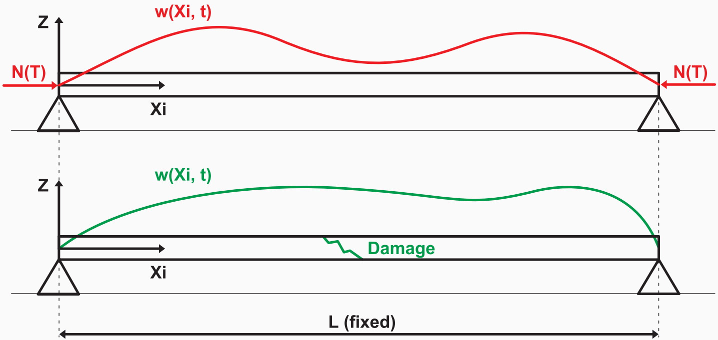
$$\gamma_1 = \sqrt{\frac{N - \sqrt{N^2 + 4EJ\rho A\omega^2}}{2EJ}} \qquad \gamma_2 = \sqrt{\frac{N + \sqrt{N^2 + 4EJ\rho A\omega^2}}{2EJ}} \tag{2}$$

Notice that $N = N(Temperature) = N_0 + k(T - T_0)$, with $k \approx -60 \frac{N}{^\circ C}$.

[1] A full derivation of the equation can be found in the appendix.

Formal definition of the problem

Both Temperature and Damage can affect the eigenfrequency and the mode shape of a structure.



After a proper OMA[1] analysis, **how to isolate the effect of environmental condition on the eigenfrequency from the effect of damage?**

[1] OMA: Operational Modal Analysis

Proposed solutions

Multivariate outlier detection

Two **methods** are presented, both **based on the concept of multivariate outlier detection in the frequency domain**:

- Mahalanobis Square Distance (MSD)
- Principal Component Analysis (PCA)

Mahalanobis Square Distance (MSD) approach

The Mahalanobis Square Distance (MSD) is a measure of the distance between a point and a distribution. It is defined as:

$$D^2_{MSD} = (\mathbf{x} - \mu)^T \mathbf{\Sigma}^{-1} (\mathbf{x} - \mu) \qquad (3)$$

Where:

- $\mathbf{x}_{(m \times 1)}$ is the vector of the observations
- $\mu_{(m \times 1)}$ is the mean of the observations
- $\mathbf{\Sigma}_{(m \times m)}$ is the covariance matrix of the observations

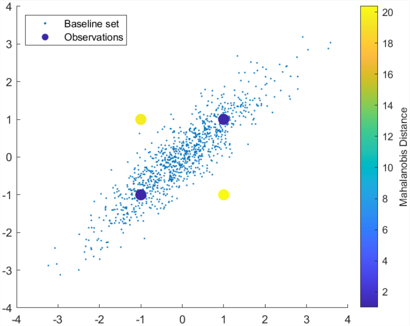


Figure 2: Application example of the MSD index. Outliers are clearly visible.

The MSD is used to detect outliers in the data, by comparing the distance of each observation from the mean of the distribution.

Problematic of the MSD approach

The MSD approach is **based on the assumption that the baseline data contains all the possible variations due to environmental effects** (e.g. temperature, vibrations noise, etc.).

To be effective then, the baseline data should be collected in a wide range of environmental conditions, in order to capture all the possible variations, **which imply a long and expensive data collection campaign** that is not always feasible.

The Principal Component Analysis (PCA) is a statistical method used to project the data onto a new coordinate system, where the new axes are the principal components of the data.

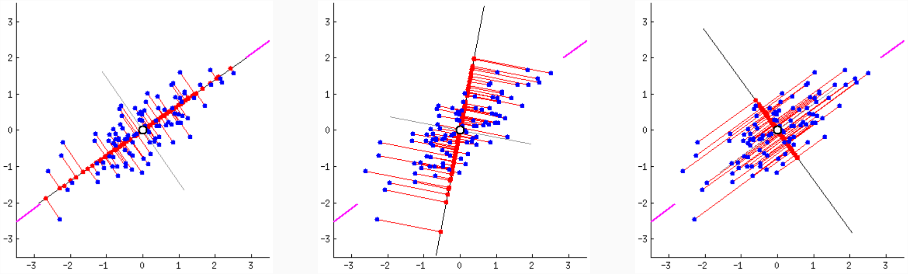


Figure 3: Example of PCA applied to a 2D dataset. Minimums of the distance function between the points and the set of new axes, determines the principal components. With refer to the figures, 1^{st} and 3^{rd} represent two different PCs configurations, while the 2^{nd} mimics the rotational transformation of the data.

In a broad, the result of the PCA can be interpreted ad the 'eigenvectors' of the cloud of data.

Singular Value Decomposition (SVD) in the PCA approach

In order to perform the PCA, a rotational transformation of the data is needed (i.e. the data is rotated in order to find the principal components).

The Singular Value Decomposition (SVD) is a mathematical technique used to compute this transformation.

By definition, SVD of a matrix $\mathbf{A}$ is defined as:

$$\mathbf{A} = \mathbf{U}\boldsymbol{\Sigma}\mathbf{V}^T \tag{4}$$

Where:

- $\mathbf{U}$ is the matrix of the left singular vectors of $\mathbf{A}$
- $\boldsymbol{\Sigma}$ is the diagonal matrix of the singular values of $\mathbf{A}$
- $\mathbf{V}$ is the matrix of the right singular vectors of $\mathbf{A}$

Analysis of signals via PCA

Given that just the first(s) principal component(s) are affected by environmental conditions, we can think of remove them from the data and apply the MSD approach to detect outliers of this dimensionally reduced dataset.

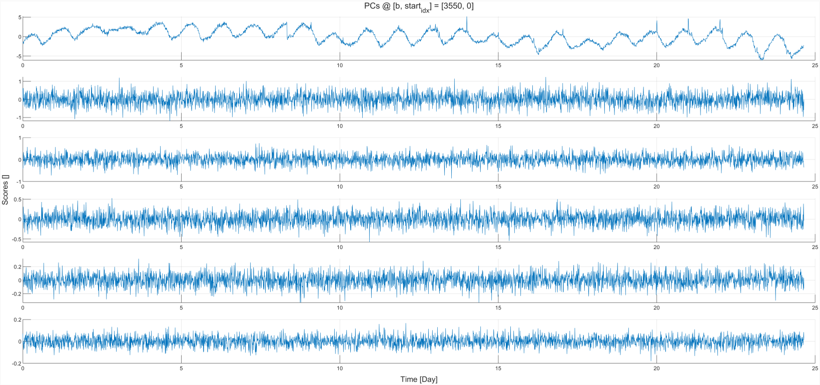


Figure 4: Scores of the eigenfrequencies of the structure projected into the principal components. Notice the clear decreasing trend of the deterministic amount as we consider higher principal components. Here, $b = 20\% \times data_{sampled} = 3550$.

Results

Here we observe the effect of the baseline set length b on the accuracy of the two methods.

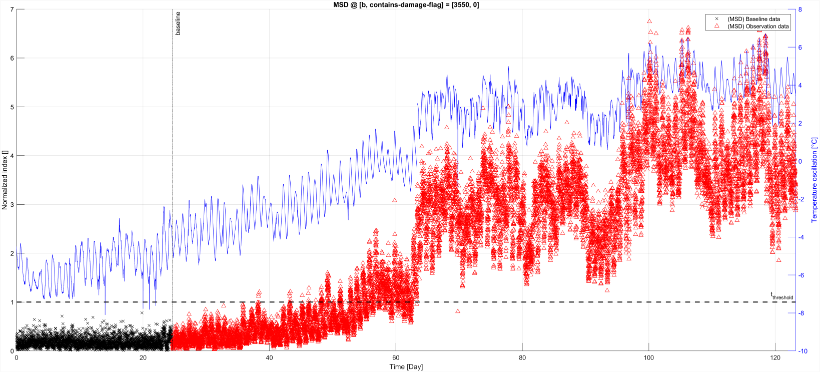


Figure 5: MSD method considering $b = 3550$

The MSD method is highly sensitive to b and if the baseline set doesn't contain a complete set of environmental conditions, it may lead to false positives/negative.

MSD vs. PCA - Baseline set length (b)

Here we observe the effect of the baseline set length b on the accuracy of the two methods.

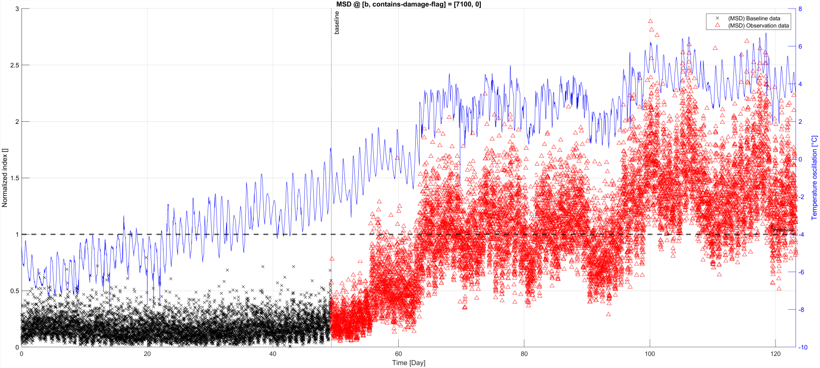


Figure 5: MSD method considering $b = 7000$

The MSD method is highly sensitive to b and if the baseline set doesn't contain a complete set of environmental conditions, it may lead to false positives/negative.

MSD vs. PCA - Baseline set length (b)

Here we observe the effect of the baseline set length b on the accuracy of the two methods.

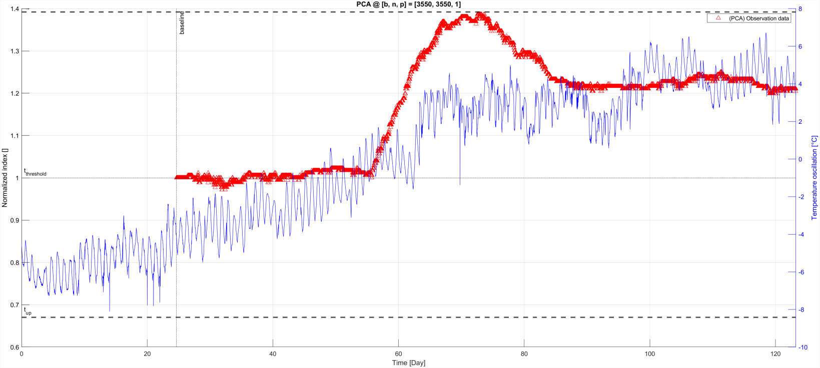


Figure 5: PCA method considering $b = n = 3550$

The MSD method is highly sensitive to b and if the baseline set doesn't contain a complete set of environmental conditions, it may lead to false positives/negative.

MSD vs. PCA - Baseline set length (b)

Here we observe the effect of the baseline set length b on the accuracy of the two methods.

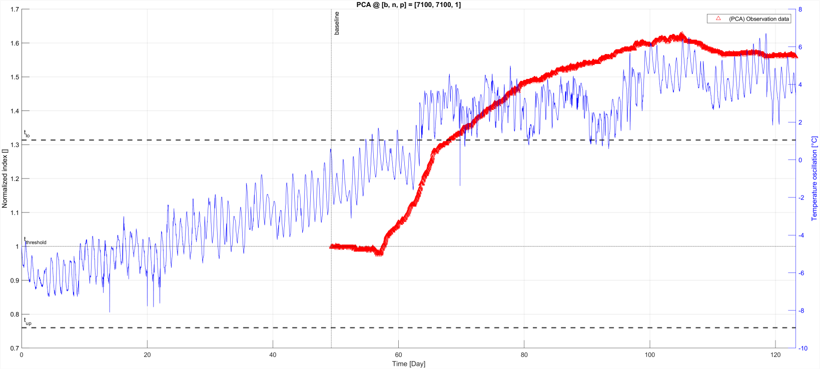


Figure 5: PCA method considering $b = n = 7000$

The MSD method is highly sensitive to b and if the baseline set doesn't contain a complete set of environmental conditions, it may lead to false positives/negative.

PCA - Observation window length (n)

A key difference between the MSD and PCA methods is that MSD compute the index for each observation record, while PCA computes the index for a set of records with length n.

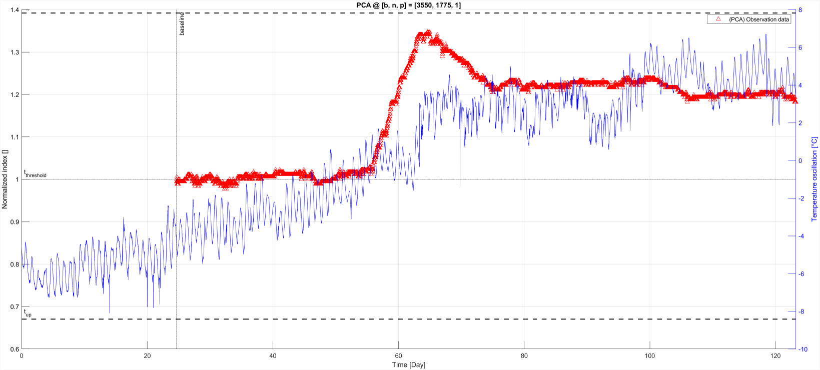


Figure 6: PCA method considering $b = 3550$ & $n = 1775$

It's clear how n affect the reactivity of the PCA method to detect outliers. In particular, the higher n, the longer the transient time before the detection of the feature.

PCA - Observation window length (n)

A key difference between the MSD and PCA methods is that MSD compute the index for each observation record, while PCA computes the index for a set of records with length n.

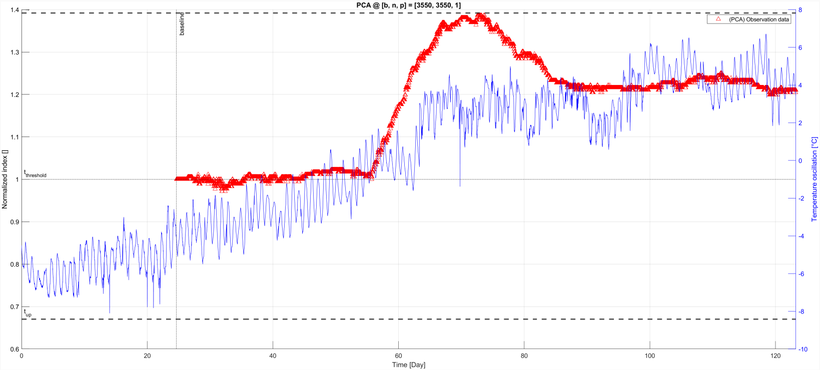


Figure 6: PCA method considering $b = 3550$ & $n = 3550$

It's clear how n affect the reactivity of the PCA method to detect outliers. In particular, the higher n, the longer the transient time before the detection of the feature.

PCA - Observation window length (n)

A key difference between the MSD and PCA methods is that MSD compute the index for each observation record, while PCA computes the index for a set of records with length n.

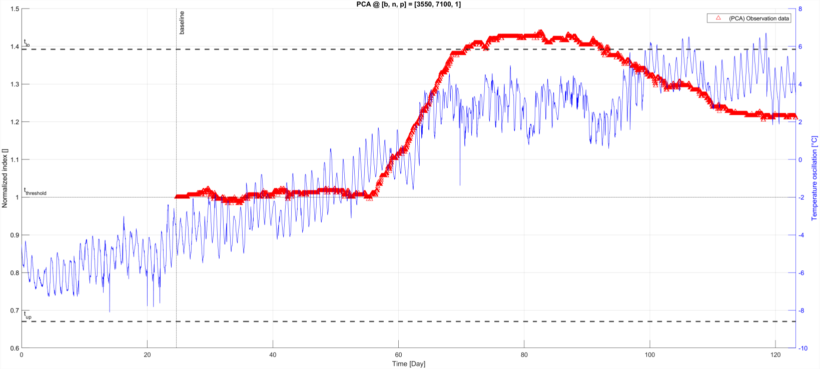


Figure 6: PCA method considering $b = 3550$ & $n = 7100$

It's clear how n affect the reactivity of the PCA method to detect outliers. In particular, the higher n, the longer the transient time before the detection of the feature.

PCA - Observation window length (n)

A key difference between the MSD and PCA methods is that MSD compute the index for each observation record, while PCA computes the index for a set of records with length n.

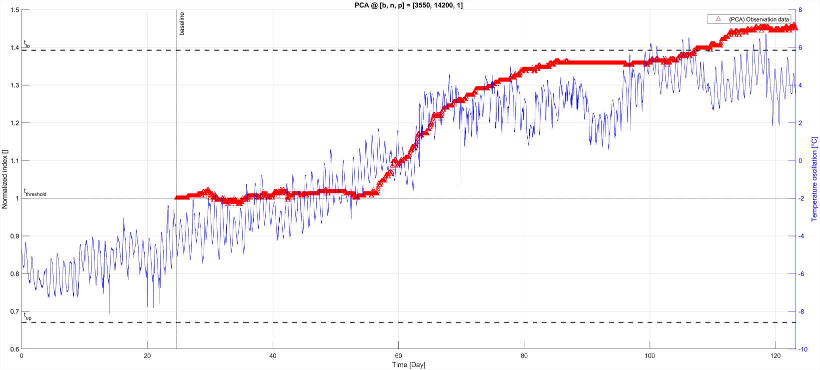


Figure 6: PCA method considering $b = 3550$ & $n = 14200$

It's clear how n affect the reactivity of the PCA method to detect outliers. In particular, the higher n, the longer the transient time before the detection of the feature.

Conclusions

Method comparison

Overall, both the proposed approaches has shown to be effective in detecting damage in the tie-rods.

However, the use of the PCA methods offers some non-negligible advantages:

- It's more robust in isolate the damage features from other sources of variability.
- It doesn't require a training set that includes all the possible environmental conditions, thus eliminating the need for a long data sampling campaign.

Future work and preliminary solutions

Regarding the PCA method, some possible future developments and solutions spotted during the research are:

Future work	**Preliminary solution**
Automatic PCs removal	Set a threshold based on an RMS of each PCs and remove the PC

M. Berardengo, F. Lucà, M. Vanali, and G. Annesi.
Short-training damage detection method for axially loaded beams subject to seasonal thermal variations.
Sensors, 23(3), 2023.

F. Lucà, S. Manzoni, A. Cigada, and L. Frate.
A vibration-based approach for health monitoring of tie-rods under uncertain environmental conditions.
Mechanical Systems and Signal Processing, 167:108547, 2022.

Questions?

Thank you!