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

Tie-rods case study

Tommaso Bocchietti 10740309

June 24, 2024

Politecnico di Milano

Agenda

1. Problem statement
2. Proposed solutions
   Mahalanobis Squared Distance (MSD)
   Principal Components Analysis (PCA)
3. Results
4. Conclusions

Figure 1: Steel tie-rods connect the buttresses of the Cathedral of Saint Peter of Beauvais in France. Credit to: *P. Dillmann*.

Problem statement

Disturbances effect of environmental conditions 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**.

By looking at the transverse vibration in a tensioned beam we can observe these dependencies:

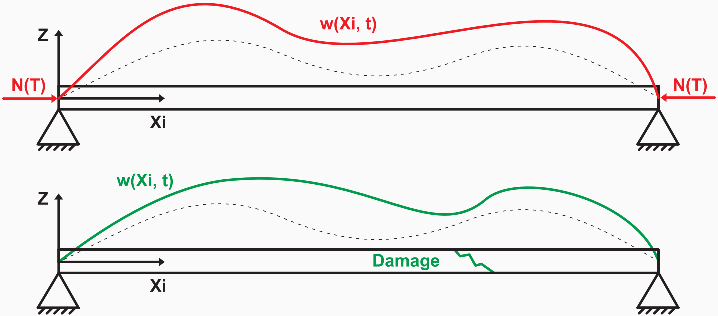
$$w(\xi, t) = [A \sin(\gamma_1 \xi) + B \cos(\gamma_1 \xi) + C \sinh(\gamma_2 \xi) + D \cosh(\gamma_2 \xi)] \, E \cos(\omega t + \phi) \quad (1)$$

Where:

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

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

Both Temperature and Damage can affect the eigenfrequencies and the mode shapes of a structure.



After a proper modal analysis, **how to detect the presence of damage in a structure that might be simultaneously affected by environmental effects?**

Proposed solutions

Multivariate outlier detection

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

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

Mahalanobis Squared Distance (MSD) approach

The Mahalanobis Squared Distance (MSD) is a measure of the distance between a point and a distribution.

It's 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

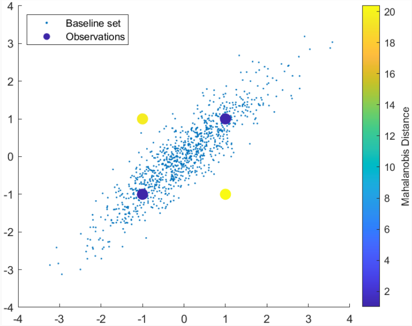


Figure 2: Application example of the MSD index. Outliers are clearly visible. Credit to: *MathWorks*.

The MSD is used to detect outliers in the data, by computing the distance between observations and the distribution of the data.

Issues of the MSD approach related to SHM

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.

Principal Components Analysis (PCA) approach

The Principal Components Analysis (PCA) is a statistical method used to project the data onto a new set of coordinate, where the new axes are the principal components of the data. It can also be used to reduce the dimensionality of the problem.

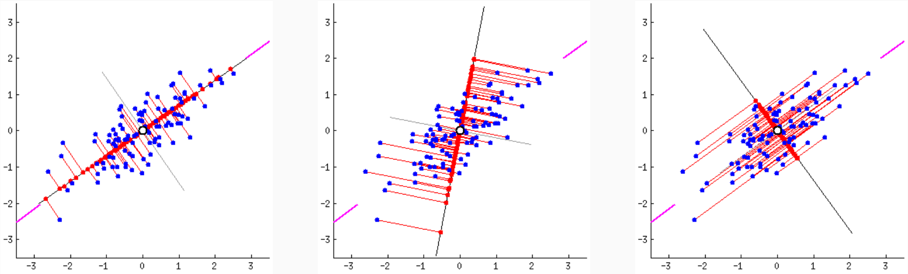


Figure 3: Example of PCA applied to a 2D dataset. PCs are identified as the directions that maximize the variance of the projected cloud of data. With refer to the figures, 1^{st} and 3^{rd} represent two different PCs configurations, while the 2^{nd} mimics the rotational transformation of the data. Credit to: *Z. Jaadi*.

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

Singular Value Decomposition (SVD) in the PCA approach

The Singular Value Decomposition (SVD) is a mathematical technique used to compute the rotational transformation needed to project the data onto the principal components.

By definition, SVD of a matrix $\mathbf{A}_{(n \times m)}$ is defined as:

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

Where:

- $\mathbf{U}_{(n \times n)}$ is the matrix of the left singular vectors of $\mathbf{A}$
- $\mathbf{\Sigma}_{(n \times m)}$ is the diagonal matrix of the singular values of $\mathbf{A}$
- $\mathbf{V}_{(m \times m)}$ is the matrix of the right singular vectors of $\mathbf{A}$

Finally, the original data can be projected onto the principal components directions by means of the following transformation:

$$\hat{\mathbf{A}} = \mathbf{A}\mathbf{V} \tag{5}$$

The clear decreasing trend of the deterministic amount, suggests that only the first(s) principal component(s) are affected by environmental conditions. By removing them, **we can analyze the remaining components that are likely to be strictly related to the damage itself**.

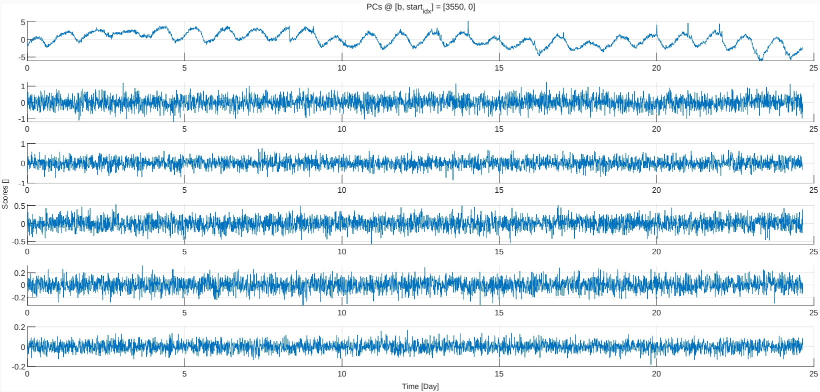


Figure 4: Scores of the eigenfrequencies of the structure projected onto the principal components. Here, $b = 20\% \times data_{length} = 3550$.

Results

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.

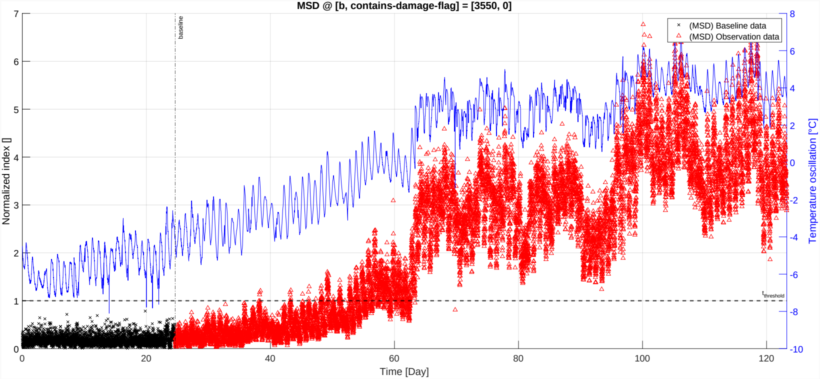


Figure 5: MSD method considering $b = 3550$

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/negatives.

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.

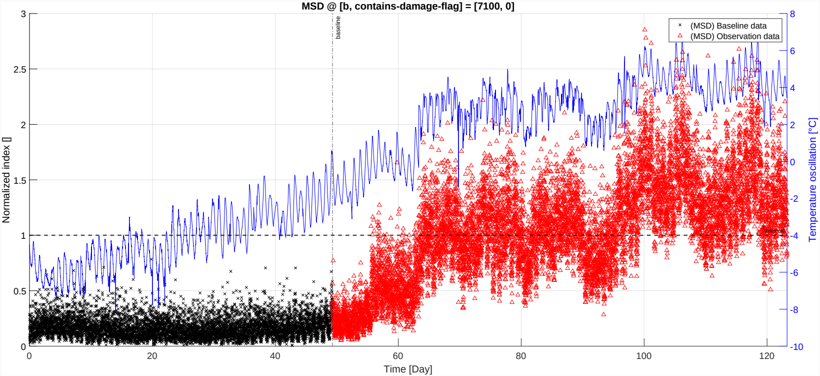


Figure 5: MSD method considering $b = 7100$

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/negatives.

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.

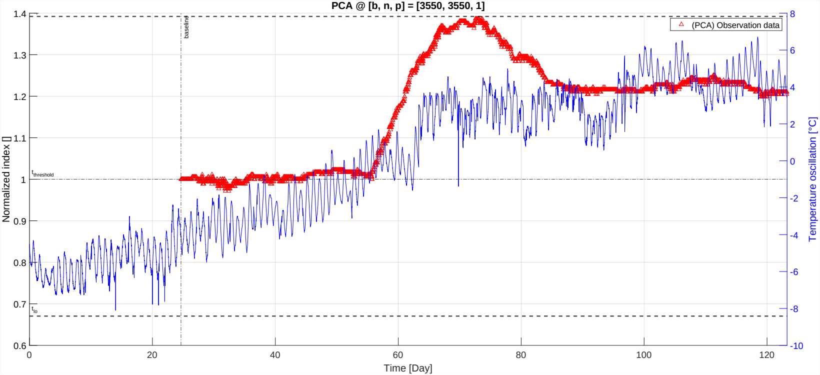


Figure 5: PCA method considering $b = n = 3550$

The PCA method, instead, is able to detect outliers (almost) independently of the baseline set length, thus shorter campaigns can be performed without losing accuracy.

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.

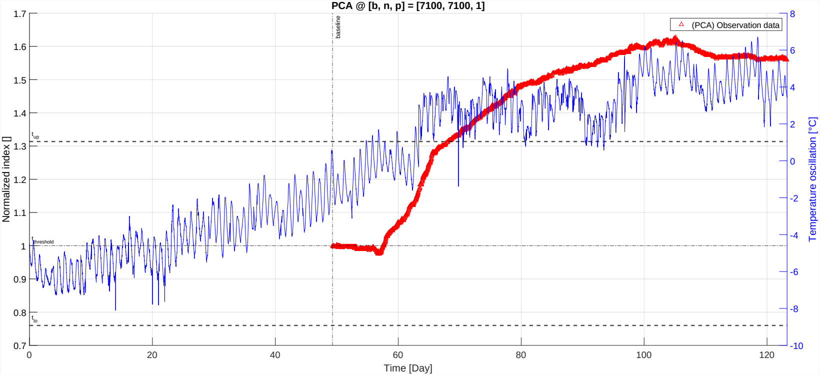


Figure 5: PCA method considering $b = n = 7100$

The PCA method, instead, is able to detect outliers (almost) independently of the baseline set length, thus shorter campaigns can be performed without losing accuracy.

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.

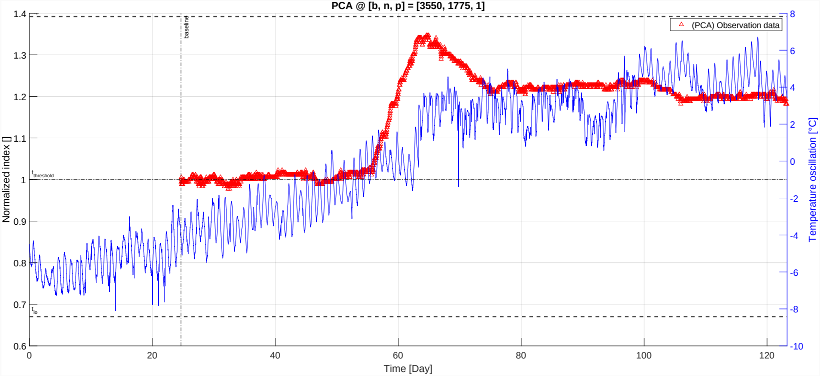


Figure 6: PCA method considering $b = 3550$ & $n = 1775$

It's clear how n affect the reactivity of the PCA method to detect outliers. In particular, the higher n, the slower the transient response of the index.

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.

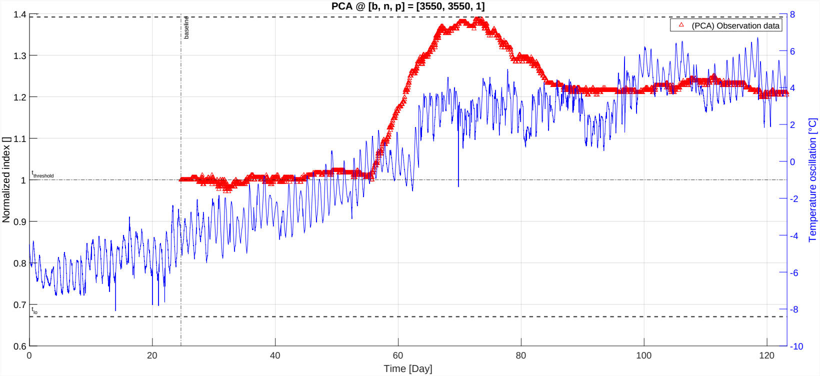


Figure 6: PCA method considering $b = 3550$ & $n = 3550$

It's clear how n affect the reactivity of the PCA method to detect outliers. In particular, the higher n, the slower the transient response of the index.

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.

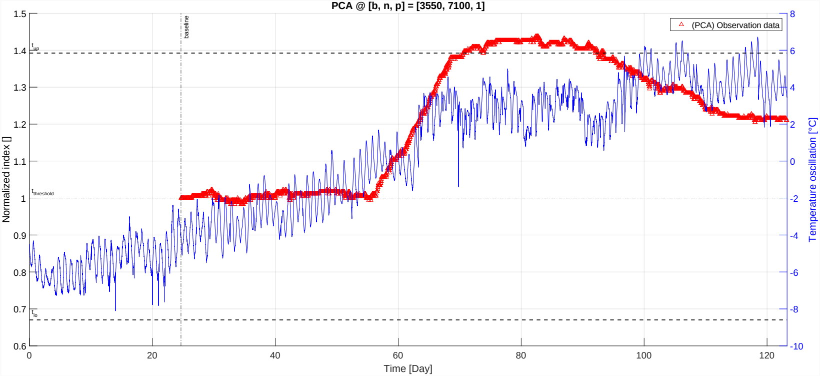


Figure 6: PCA method considering $b = 3550$ & $n = 7100$

It's clear how n affect the reactivity of the PCA method to detect outliers. In particular, the higher n, the slower the transient response of the index.

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.

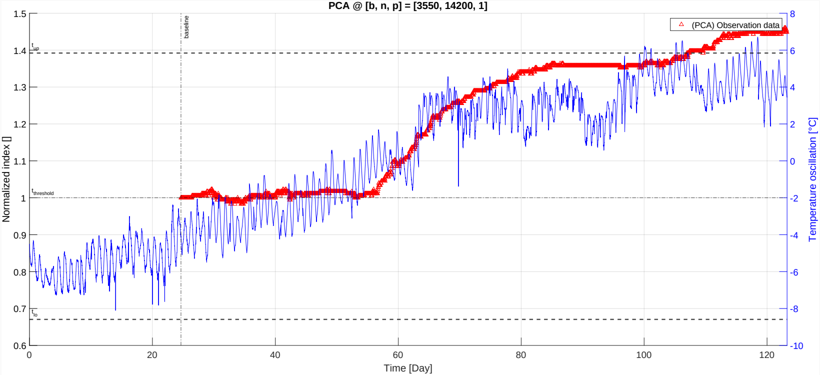


Figure 6: PCA method considering $b = 3550$ & $n = 14200$

It's clear how n affect the reactivity of the PCA method to detect outliers. In particular, the higher n, the slower the transient response of the index.

Conclusions

Conclusions

Overall, both the proposed approaches have shown to be effective in performing SHM over simple structure like tie-rods, with the capability of detecting damage even in the presence of environmental variability.

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.

Moreover, parameters like the baseline set length b and the observation window length n must be carefully chosen to obtain an optimal performance of the method.

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!