

## **ROC-BASED PERFORMANCE EVALUATION OF DATA CLEANSING TECHNIQUES FOR FALSE ALARM RISK REDUCTION IN CONTINUOUSLY MONITORED BRIDGES**

**Valentina Giglioni<sup>1</sup>, Enrique García-Macías<sup>2</sup>, Ilaria Venanzi<sup>2</sup>, Laura Ierimonti<sup>2</sup> and  
Filippo Ubertini<sup>2</sup>**

<sup>1</sup>Department of Civil and Environmental Engineering, University of Perugia. Via G. Duranti 93,  
Perugia - 06125, Italy.  
e-mail: [valentina.giglioni1@studenti.unipg.it](mailto:valentina.giglioni1@studenti.unipg.it)

<sup>2</sup> Department of Civil and Environmental Engineering, University of Perugia. Via G. Duranti 93,  
Perugia - 06125, Italy.  
e-mail: {enrique.garciamacias, ilaria.venanzi, laura.ierimonti, filippo.ubertini}@unipg.it

---

**Abstract.** *Maintenance and prevention of roadway bridges is an important societal and economic challenge where structural health monitoring systems can play a major role. The main reason lies in the significant number of ageing infrastructures requiring a strategy to evaluate their structural integrity and durability as well as to enable structural prognosis. In the context of unsupervised learning processes, damage detection can be accomplished by tracking deviations of dynamic parameters from their normal conditions using specific tools, named control charts. However, the main drawback of dealing with changes in modal properties is represented by their high sensitivity to environmental and operational factors. It follows that these effects must be necessarily removed to obtain suitable damage sensitive features. In this regard, the selection of a particular statistical model for data cleansing inevitably influences the distribution of the residuals, the control chart and therefore the damage detection. To overcome this issue, in the present paper an original methodology is developed, based on Receiving Operating Characteristic (ROC) curves, which provides a new decision-support tool for the definition of the best environmental effects' removal technique. Residuals stemming from the application of different data cleansing methods, plotted in control charts are conceivably distinguished in False Positives, True Positives, False Negatives and True Negatives. Consequently, ROC curves are computed and compared, each one referred to a particular data cleansing procedure, among which Principal Component Analysis, Multiple Linear Regression and more and to a specific damage class. The proposed approach is applied in the illustrative case study of the Z24 Bridge, located in the canton Bern near Solothurn, Switzerland. The results of the application of the procedure demonstrated that ROC curves are a suitable tool for defining the optimal damage threshold and the most effective environmental effects' removal technique, in order to minimize the occurrence of false alarms.*

**Keywords:** Bridge monitoring, ROC curves, Structural Health Monitoring, Data cleansing, False alarm.

---

## 1 INTRODUCTION

Structural Health Monitoring (SHM) has become increasingly popular in the field of Civil Engineering. The main reason lies in the necessity to better manage the significant number of ageing structures and infrastructures. In particular, bridges represent critical elements in modern transport networks, hence the importance to develop strategies ensuring their safety as well as correct prevention [1]. Within this context, long-term vibration-based SHM systems represent a suitable solution [2]. Such methods provide modal parameters' estimate by exploiting OMA analysis and allow to identify damage-induced anomalies through statistical pattern recognition techniques.

However, a major challenge in application of SHM systems to civil structures and infrastructures is represented by the high sensitivity of dynamic properties to environmental and operational conditions [3]. For this reason, in order to minimize the significant risk of masking early stage damages, several techniques to remove environmental effects from monitoring data are continuously being developed in the SHM community. The basic idea is to create a proper statistical model able to reproduce the part of variance in frequency estimates that is associated with changes in environmental conditions. To this aim, common statistical tools such as Multiple Linear Regression (MLR) [3, 4] and Principal Component Analysis (PCA) [5] are broadly investigated in the literature.

Once a proper statistical model is selected, it is possible to compute quantities, called residuals, that are only minimally affected by environmental effects and therefore suitable to be used as damage-sensitive features [6]. In this framework, Novelty Analysis through control charts [7] represents a common statistical process control tool useful to identify the beginning of a damage mechanism. In particular, proper statistical distances between newly acquired data and a baseline population allow to detect any outlier consistently overpassing a fixed threshold. As a consequence, it is clear that the definition of the threshold in the classification (damage or non-damaged) represents one of the major issues, inevitably producing erroneous predictions as False Positives and False Negatives.

Therefore, with the aim to minimize the rate of damage detection errors in view of an effective monitoring, a strategy to select, for the specific case study, the most appropriate technique to remove environmental effects is highly required. Such purpose is accomplished in this paper by proposing a new decision-support tool to optimally solve the statistical model selection problem, which heavily affects the distribution of the residuals and hence, damage detection. The new methodology described in the present work is based on the use of ROC [8] curves as performance metrics, aiming at providing a quantification of damage classification errors. Despite the popularity of ROC in several application fields [9], their use for optimal statistical pattern recognition within the context of damage detection, is still lacking.

In order to bridge this research gap, the basic idea of this paper is to supply objective criteria allowing to define (i) which data cleansing technique would ensure the best damage identification and (ii) which cut-off value minimizes the number of false positives and false negatives.

Preliminarily, residuals are calculated and plotted in different control charts. Then, several ROC curves, each one referred to a particular statistical model and to a specific damage scenario, are computed and afterwards assessed by measuring the entire two-dimensional area underneath the curve (AUC). Hence, this strategy can be formulated as an optimization problem, aiming at selecting the curve with the higher AUC, which corresponds to the most suitable model for data cleansing purposes.

## 2 BACKGROUND

In order to discern damage-induced changes in the natural frequencies time-histories, several tools have been recently used to remove environmental and operational effects. This section briefly presents an overview on MLR, PCA and the Gaussian Mixture Model (GMM) and provides a description of the basic theory of Novelty Analysis, commonly utilized to detect the occurrence of damage by means of proper statistical distances.

### 2.1 Statistical models handling environmental and operational effects

Multiple Linear Regression (MLR) exploits linear correlations between one or more dependent variables, called estimators and a group of independent variables, defined as predictors. In this case, dependent variables are natural frequencies, while independent variables are typically environmental and operational factors. The linear statistical model should be able to study the influence of each predictor on the dependent variables and then, to predict future values of the natural frequencies when only the predictors are known [4].

Another well-established methodology, the Principal Component Analysis, has the primary purpose to reduce the dimensionality of a data set without producing significant losses of data variance. It consists in transforming a set of observations of possibly correlated variables into a group of uncorrelated variables, called principal components (PCs). Firstly, the original data are projected into the vectorial space generated by the PCs. Then, according to this approach, a subset of those PCs providing the largest contributions to the variance of the original data can be retained and assumed representative of the environmental and operational factors. As a result, it is possible to estimate natural frequencies by moving back the reduced subset of PCs onto the original space [5, 6].

In order to handle non-linear environmental/operational effects, Gaussian Mixture Model (GMM) represents an efficient tool, which is based on grouping damage sensitive features (e.g. natural frequencies) in the training period into different clusters. Considering that the data set in the training period is non-normally distributed, this approach consists in introducing  $K$  clusters, described by Gaussian distributions, and assuming a linear superposition of them equal to the probability density function of the data set in the training period [10, 11]. Once the  $K$  clusters have been defined, new acquired data can be assigned to the cluster with minimal Mahalanobis distance.

### 2.2 Novelty analysis

The selection of the proper statistical model, successful in reproducing environmental effects on dynamic signatures, allows to compute the residual error matrix  $\mathbf{E}$  as:

$$\mathbf{E} = \mathbf{Y} - \hat{\mathbf{Y}} \quad (1)$$

which contains quantities that are minimally affected by environmental factors and therefore assumed as damage sensitive features. It follows that any damage pattern can be inferred by observing anomalies in the distribution of  $\mathbf{E}$ .

To this aim, Novelty Analysis through the use of control charts represents a statistical tool able to associate significant deviations of natural frequencies from normal conditions, in terms of statistical outliers, with the occurrence of damage. Firstly, it is necessary to define a reference condition, containing data sets collected in a training period, in which the structure is assumed to remain healthy. Consequently, any new data point is compared with such baseline population by means of properly defined statistical distances. In this framework, the  $T^2$ -statistic is adopted,

defined as:

$$T^2 = r \cdot (\bar{\mathbf{E}} - \bar{\bar{\mathbf{E}}})^T \cdot \Sigma^{-1} \cdot (\bar{\mathbf{E}} - \bar{\bar{\mathbf{E}}}) \quad (2)$$

where  $r$  is an integer parameter referred to as subgroup size,  $\bar{\mathbf{E}}$  is the mean of the residuals computed in the subgroup of the last  $r$  observations, while  $\bar{\bar{\mathbf{E}}}$  and  $\Sigma$  are the mean value and the covariance matrix of the residuals statistically estimated in the training period, respectively. Once several values of the statistical distance steadily overcome the Upper Control Limit (UCL), a change in the distribution of the residuals may have been occurred. Therefore, damage-induced anomalies not encountered in the training period are conceivably identified. In this context, the limit threshold (UCL) is computed as a certain percentage of the standard deviation of the control chart in the training period.

### 3 METHODOLOGY

#### 3.1 ROC curves

A Receiver Operating Characteristic (ROC) curve represents a graphical tool which allows to quantify the performance of a process, varying the threshold position, and to statistically evaluate the erroneous predictions related to false detection.

Firstly, it is necessary to illustrate all the possible outcomes provided by the control chart (Fig. 1 a)):

- True Negative (TN): the data set point before damage lying under the UCL;
- False Positive (FP): the data set point before damage lying above the UCL;
- False Negative (FN): the data set point after damage lying under the UCL;
- True Positive (TP): the data set point after damage lying above the UCL;

The confusion matrix  $C$  can be constructed starting from the knowledge of TP, FP, TN and FN:

$$C = \begin{bmatrix} \text{TP} & \text{FP} \\ \text{FN} & \text{TN} \end{bmatrix}; \quad (3)$$

In order to characterize a ROC curve, some statistical measures, evaluating the performance of a binary classification test, are introduced hereafter.

The true positive rate TPr defines how many correct positive results occur among all positive samples. It is also known as *sensitivity* (SE) or *probability of detection*:

$$\text{TPr} = \frac{\text{TP}}{\text{TP} + \text{FN}} = \frac{\text{True Positives}}{\text{All positive cases}} \quad (4)$$

The false positive rate FPr, or *probability of false alarms*, counts how many incorrect positive results occur among all negative samples:

$$\text{FPr} = \frac{\text{FP}}{\text{FP} + \text{TN}} = 1 - \text{SP} = \frac{\text{False Positives}}{\text{All negative cases}} \quad (5)$$

where SP, called as *specificity* or true negative rate, describes how many correct negative results occur among all negatives samples.

A ROC curve plots the true positive rate (TPr) against false positive rate (FPr) at various classification thresholds. The basic theory is that a very low threshold produces 100% of false and true positives, classifying all the items as positive. On the contrary, a very high threshold could never indicate damage, resulting in 0% of false and true positives. It is important to highlight that the model's accuracy increases when the curve comes closer to the upper-left-hand corner of the ROC space and decreases, vice versa, when the curve comes closer to the 45° diagonal (Fig. 1 b).

The performance of a ROC curve is commonly evaluated by computing the Area Under the Curve (AUC), indicated in Fig. 1 c). Considering normalized units, AUC values go from 0 to 1. It is worth pointing out that this parameter is equivalent to the probability that a classifier will rank a randomly chosen positive instance higher than a randomly chosen negative one [8]. It can be defined as follows:

$$AUC = \int_0^1 ROC(f)df \quad (6)$$

where  $f$  and  $ROC(f)$  indicate, respectively, the false positive rate (FPr) and the corresponding true positive rate (TPr).

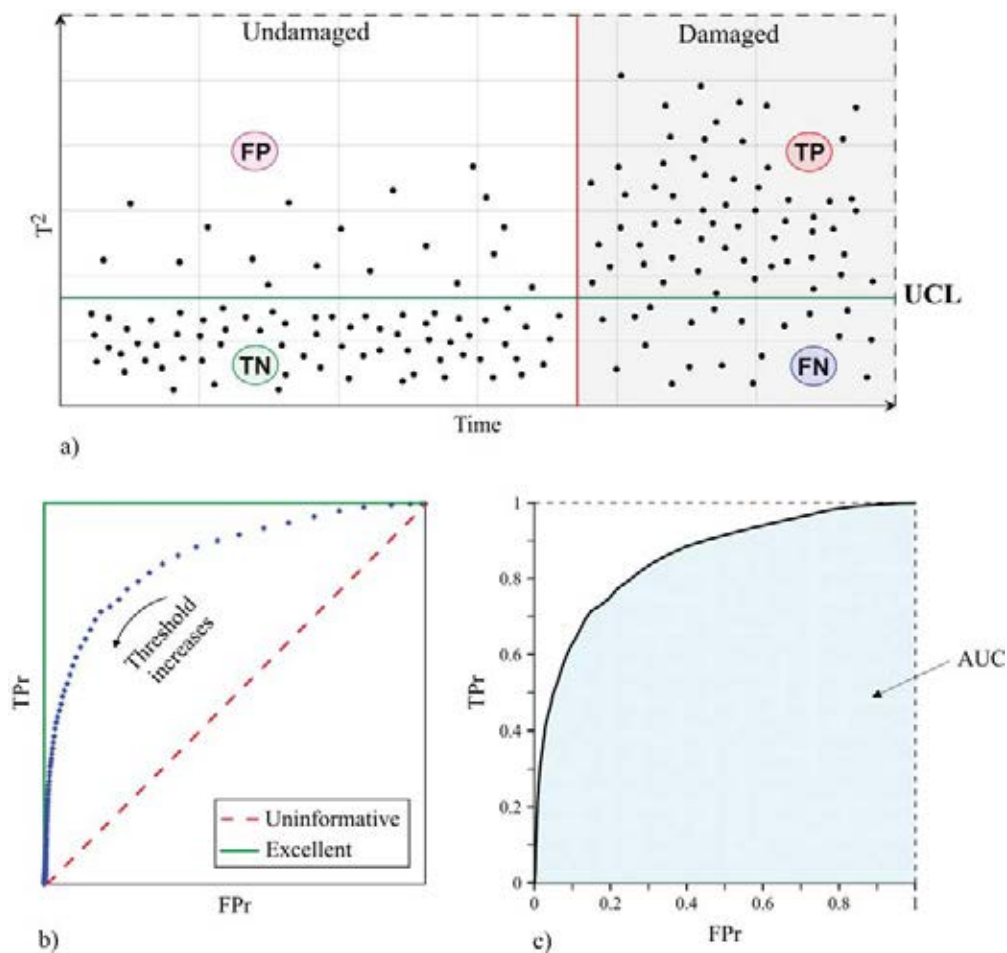


Figure 1: a) control chart items' classification: True Positive (TP), False Positive (FP), True Negative (TN) and False Negative (FN); b) different typologies of ROC curves; c) definition of AUC.



### 3.2 Selection of the best data cleansing technique

The methodology for selecting the best data cleansing technique is provided in Fig. 2 a). Preliminarily,  $m$  statistical models and  $d$  damage scenarios (real or simulated) are taken into account, with the purpose to evaluate and compare models' performance, varying the type of damage.

It is interesting to point out that damage data are rarely available in the field of civil engineering. Therefore, the use of suitable virtual models to simulate different damage scenarios becomes necessary. In these cases, damage-induced variations of modal properties predicted by the structural model can be artificially introduced into the monitoring data through constant shifts in the time series of identified modal frequencies.

By applying Eq. (1), residuals can be computed and then plotted in  $(m \times d)$  control charts. As a result, by counting the data set points belonging to each class, labelled as FP, TP, FN and TN,  $(m \times d)$  ROC curves are obtained, varying the classification threshold. Each curve is referred to a particular statistical model (or data cleansing technique), and to a specific damage scenario. In order to conceivably evaluate the behavior of different curves in a quantitative point of view, AUC is computed to provide a measure of performance across all possible classification thresholds. The basic theory of the proposed approach is to select the curve with the higher AUC, corresponding to the most performing model with an excellent capability in discerning between the positive and negative classes. This translates into an optimization problem:

$$i_{opt} = \arg \max_i (J_i) \quad (7)$$

$$J_i = \sum_{j=1}^d \text{AUC}_{ij} \quad (8)$$

where  $\text{AUC}_{ij}$  represents the area under the ROC curve for the  $i^{th}$  statistical model and the  $j^{th}$  damage scenario, with  $i = 1, 2, \dots, m$  and  $j = 1, 2, \dots, d$ .

### 3.3 Optimal threshold selection

In order to minimize false alarms and missing alarms, the selection of the optimal classification threshold has an extremely important role. To this aim, the proposed methodology, schematically presented in Fig. 2 b), involves the use of *Youden index* ( $Y$ ), which is a useful parameter allowing to provide a summary measure about model's performance. It is calculated in ROC space and conceptually represents the vertical distance between a point on the ROC curve and the  $45^\circ$  line. Typically, a high value of the *Youden index* corresponds to a well performing ROC curve (close to the upper-left-hand corner) leading to a high percentage of true positive rate over false positive rate. The index is calculated as follows:

$$Y = \text{SE} + \text{SP} - 1 \quad (9)$$

Once the most performing model is defined, the idea is to compute *Youden index* for each  $j^{th}$  damage scenario, varying classification thresholds. The maximum values,  $Y_{max}(j)$ , correspond to a specific damage threshold, which is denoted as  $T(j)$ .

However, it is important to underline that such procedure implies the inspection of several thresholds, each one related to a specific damage scenario. As a solution, the strategy proposed in this paper aims at fixing the optimal threshold as the minimum among the  $T(j)$ , so that the SHM system is able to detect even the smallest damage:

$$T_{opt} = \min \{T(j)\} \quad (10)$$

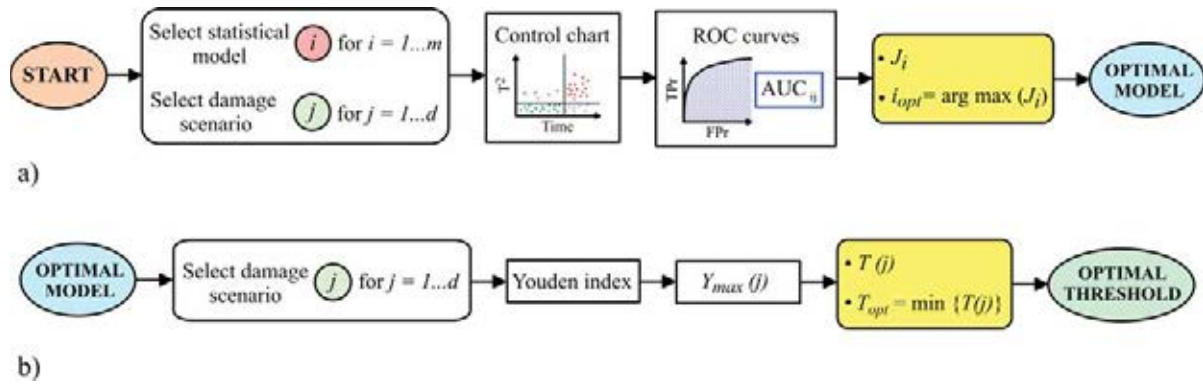


Figure 2: a) flow chart of the proposed methodology to select the most performing statistical model; b) flow chart of the proposed procedure for the optimal threshold selection.

## 4 APPLICATION CASE STUDY: Z24 BRIDGE

This section presents a general overview about the geometry and the monitoring system of the Z24 bridge and discuss, additionally, the results obtained by application of the proposed methodology on the investigated case study. To this aim, the available monitoring data are collected, processed and analysed through a specific MATLAB code, called MOSS, implemented by García-Macías et al. [12].

### 4.1 General description

The Z24 Bridge (Fig. 3) was located in the canton Bern, Switzerland, overpassing the national highway between Bern and Zurich. It was a post-tensioned concrete box girder bridge with a main span of 30 m and two side spans of 14 m. The two triplets of concrete columns at both ends were connected via concrete hinges to the girder, while two concrete piers, representing the intermediate supports of the bridge, were clamped into the girders and situated at the end points of the main span (Fig. 3 b)).

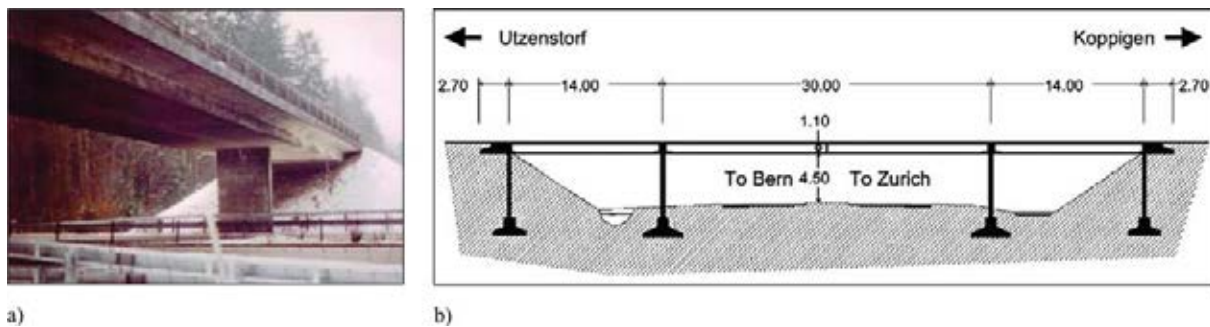


Figure 3: Z24 bridge: a) general view [13], b) longitudinal section [14].

The bridge was demolished at the end of 1998 to build a larger side span bridge. However, before demolition, it was continuously monitored from November 1997 till September 1998, with the aim to provide both environmental and vibration data for dynamic behavior's assessment. In particular, progressive damage tests were carried out during the summer of 1998 (shortly before the complete demolition of the bridge) in order to study the influence of different realistic damage scenarios on the modal properties.

Sensors to measure accelerations as well as environmental parameters were spread across the bridge. More specifically, among 16 accelerometers, only 8 were properly working. Data from all sensors were captured every hour with a sampling frequency of 100 Hz, resulting in 24 measurements per day.

#### 4.1.1 Definition of the damage scenarios

With the aim to evaluate modal parameters' variations in view of a prompt damage detection, frequency tracking has been performed over time, as depicted in Fig. 4.

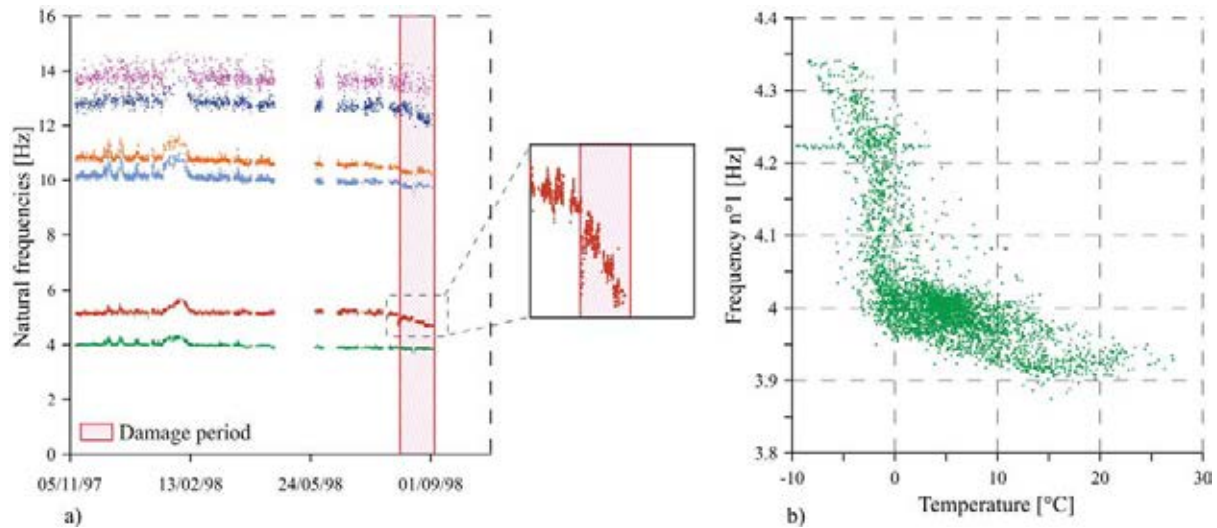


Figure 4: a) frequency tracking of the Z24 Bridge; b) relation between the first natural frequency and the temperature.

Between August 10<sup>th</sup> and September 4<sup>th</sup> 1998, the bridge was subjected to progressive damage tests. In particular, four damage scenarios ( $d = 4$ ) are taken into account in the present work:  $d_1$ ,  $d_2$ ,  $d_3$  and  $d_4$  indicate, respectively, the whole damage period (10<sup>th</sup> August - 4<sup>th</sup> September), the first portion of damage period (10<sup>th</sup> August - 18<sup>th</sup> August), the second portion (19<sup>th</sup> August - 26<sup>th</sup> August) and finally, the third one (27<sup>th</sup> August - 4<sup>th</sup> September).

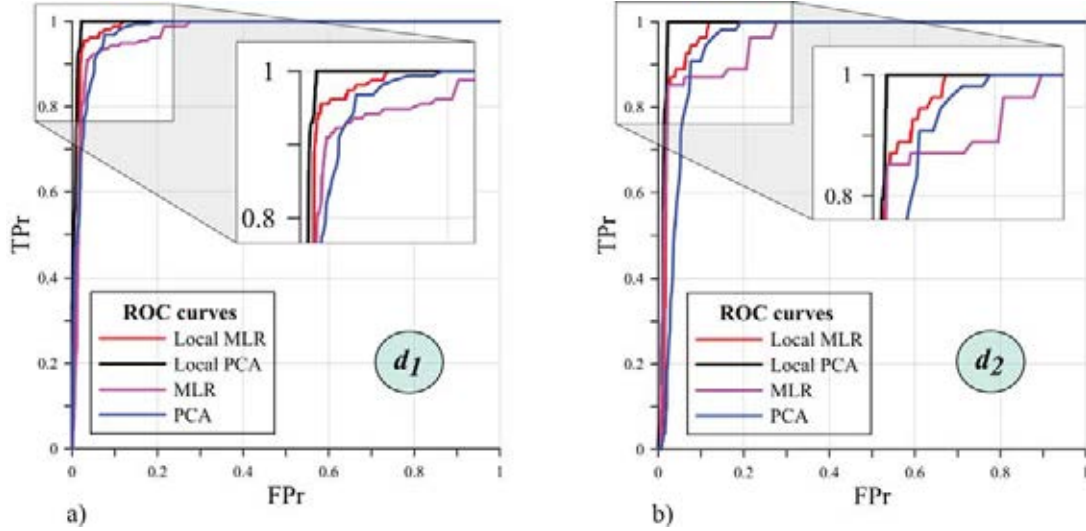
#### 4.1.2 Comparison between different statistical models

In order to get residuals, four data cleansing techniques ( $m = 4$ ) are considered, namely Multiple Linear Regression (MLR), Principal Component Analysis (PCA), Local MLR and Local PCA. The application of every statistical model, varying damage scenarios, leads to different control charts, each one referred to the  $i^{th}$  model and the  $j^{th}$  damage condition. By counting false positives, true positives, false negatives and true negatives, outgoing from every control chart, ROC curves are computed empirically by varying the classification threshold. Following the proposed methodology, AUC values are calculated and afterwards collected in Table 1.

It can be noticed that ROC curves exhibit a good behavior if the whole damage period ( $d_1$ ) is considered (Fig. 5 a)), even though slightly more performing results are visible when dealing with larger damage severity (e.g.  $d_4$ ), since the probability to discern true positive items highly increases. Conversely, when damage to be identified is not so marked (e.g.  $d_2$ ), AUC values get inevitably worse allowing, though, to clearly distinguish various performance levels associated to the  $m$  models, as shown in Fig. 5 b).



d	AUC			
	PCA	Local PCA	MLR	Local MLR
$d_1$	0.979	0.994	0.973	0.984
$d_2$	0.953	0.988	0.959	0.975
$d_3$	0.988	0.996	0.971	0.984
$d_4$	0.997	0.999	0.989	0.993

Table 1: *AUC* values for  $m$  statistical models and  $d$  damage scenarios.Figure 5: Comparison between the different statistical models in terms of ROC curves. a, b) ROC curves computed for the whole damage period ( $d_1$ ) and for the first portion of damage period ( $d_2$ ), respectively.

Since the damage inferred to the bridge was significant, it is necessary to point out that every model managed to yield extremely performing ROC curves, regardless of the selected damage scenario. However, in order to find out which is the best technique to remove environmental effects, the proposed optimization problem is applied. Following Eq. (7), the argument of the maximum of the objective function,  $i_{opt}$ , corresponds to the Local PCA model, the one providing the best performance in terms of AUC.

Due to the fact that the damage inferred to the structure progressively increases during the monitoring time, it could be interesting to identify which technique stands out in detecting damage first. To this aim, each control chart associated to the  $i^{th}$  model is taken into account and the portion corresponding to damage is afterwards divided into a certain number of steps. By computing and plotting AUC values at every step, as shown in Fig. 6, it comes to the light that Local PCA results to be the best model in a global view, while MLR turned out to be the fastest technique to detect damage.

#### 4.1.3 Selection of the optimal threshold

Once Local PCA is identified as the most performing model, it is crucial to minimize damage errors such as false positives and false negatives. This translates into the selection of the optimal threshold, leading to the higher TPr over the FPr within the ROC space. To this aim, *Youden index* is computed for every  $j^{th}$  damage scenario, varying the classification threshold. Table 2

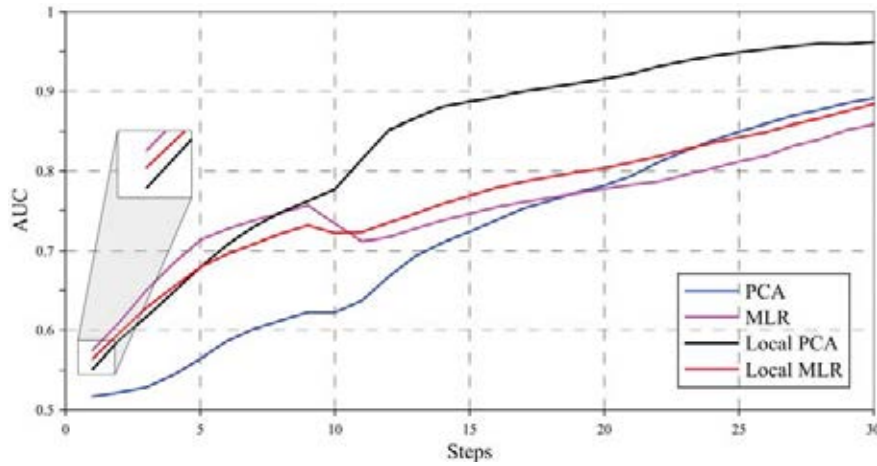


Figure 6: Comparison between  $m$  models in terms of AUC, computed at every step throughout the damage period.

summarizes the maximum values  $Y_{max}(j)$  with the corresponding thresholds  $T(j)$ .

d	Local PCA	
	$Y_{max}$	$T$
$d_1$	0.978	0.196
$d_2$	0.978	0.196
$d_3$	0.990	0.424
$d_4$	0.993	0.493

Table 2: Maximum values of *Youden index* ( $Y_{max}$ ) for every damage scenario and the corresponding thresholds ( $T$ ).

With the growing of damage severity, it is possible to highlight (i) an increase of the optimal threshold and (ii) an enlargement of the interval containing the maximum values of *Youden index*, as clearly visible in Fig. 7 a).

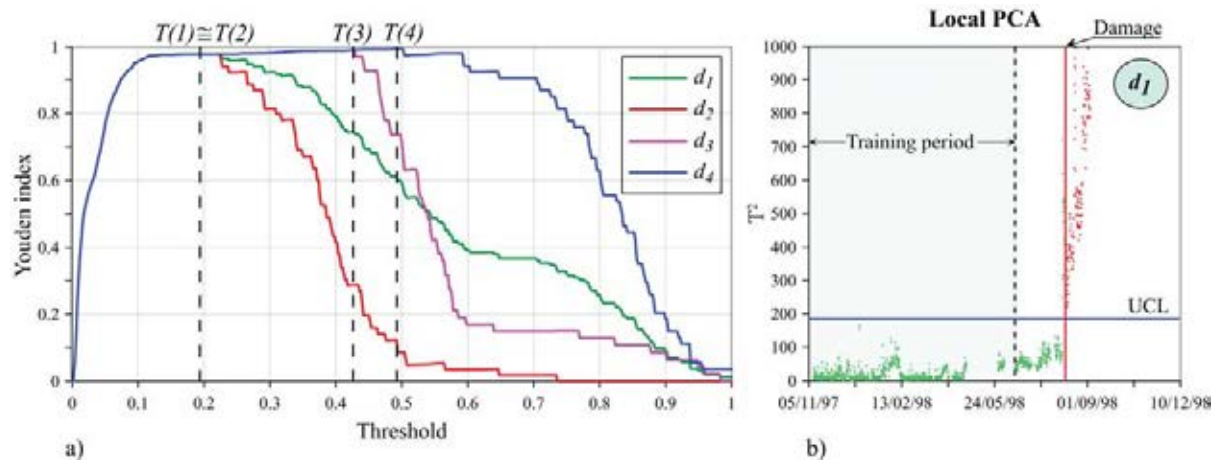


Figure 7: a) *Youden index* for  $d$  damage scenarios, varying the classification thresholds; b) control chart of the Z24 bridge obtained for a specific damage scenario ( $d_1$ ) by setting Local PCA as statistical model and  $T_{opt}$  as UCL.

However, as shown in Table 2, the methodology has provided several thresholds, each one related to a specific damage scenario. To overcome this issue, the optimal value  $T_{opt}$ , required to set the UCL of the control chart, is selected as the minimum:  $T_{opt} = \min \{T\} = 0.196$ . The control chart obtained by considering Local PCA as statistical model and  $d_1$  as damage scenario is depicted in Fig. 7 b), where the UCL is set equal to the optimal threshold computed using Eq. (10). It is easy to point out the excellent capability of this data cleansing technique (i.e. Local PCA) in correctly classifying the outcomes throughout the possible thresholds, notably avoiding the false negatives detection.

## 4.2 Discussion of the results

The relation between the identified natural frequencies and temperature may significantly affect the selection of the best environmental effects removal technique. Regarding the Z24 bridge, such relation is non-linear, as shown in Fig. 4 b). This could be the key for the interpretation of the results, as local approaches are typically useful to handle non-linear environmental/operational effects. In this particular case study, Local PCA (Principal Component Analysis applied to single clusters), in comparison with linear regression analysis (MLR), turned out to be the statistical model with the highest capability in minimizing erroneous predictions in view of a reliable damage detection.

## 5 CONCLUSIONS

The present paper provides a new decision-support tool for the definition of the best technique to remove environmental effects, as well as for the selection of the optimal classification threshold, minimizing false alarms and missing alarms. The proposed methodology is based on the use of ROC curves as performance metrics, which are computed and compared, in terms of AUC, for a variety of statistical models and different damage scenarios. Therefore, the selection of the best statistical model is carried out in the form of an optimization problem. Moreover, keeping fixed the most performing technique, the parameter *Youden index* is introduced to find out the optimal threshold for every damage scenario. Finally, in order to let the SHM system detect even the smallest damage, the strategy is to select the lowest value as optimal UCL.

The developed procedure has been exemplified by application in the Z24 bridge case study. The results illustrate that all the models managed to clearly detect the huge damage progressively inferred to the bridge. Local PCA, though, has showed a higher capability to correctly classify the outcomes of the control chart, proving to be the optimal model to remove environmental effects for an effective damage detection.

In conclusion, the provided methodology represents a valid tool to statistically evaluate the performance of different data cleansing techniques, by supplying reliable performance metrics within unsupervised damage detection under changing environment. The procedure is general and can be implemented in manifold SHM applications, considering different statistical models and different damage scenarios.

## Acknowledgements

The supports of the Italian Ministry of University and Research (MIUR) and the University of Perugia are acknowledged, within the Program “Dipartimenti di eccellenza 2018-2022”.

## REFERENCES

- [1] B.J.A. Costa, F. Magalhães, T. Cunha, J. Figueiras, Rehabilitation assessment of a century steel bridge based on modal analysis. *Engineering Structures*, **56**, 260–272, 2013.
- [2] O. Avci, O. Abdeljaber, S. Kiranyaz, M. Hussein, M. Gabbouj, D.J. Inman, A review of vibration-based damage detection in civil structures: From traditional methods to Machine Learning and Deep Learning applications. *Mechanical Systems and Signal Processing*, **147**, 2021.
- [3] A. Kita, N. Cavalagli, F. Ubertini, Temperature effects on static and dynamic behavior of Consoli Palace in Gubbio, Italy. *Mechanical Systems and Signal Processing*, **120**, 180–202, 2019.
- [4] F. Magalhães, A. Cunha, E. Caetano, Vibration based structural health monitoring of an arch bridge: From automated OMA to damage detection. *Mechanical Systems and Signal Processing*, **28**, 212–228, 2012.
- [5] A.-M. Yan, G. Kerschen, P. De Boe, J.-C. Golinval, Structural damage diagnosis under varying environmental conditions - Part I: A linear analysis. *Mechanical Systems and Signal Processing*, **19**, (4), 847–864, 2005.
- [6] E. García-Macías, I. Venanzi, F. Ubertini, Metamodel-based pattern recognition approach for real-time identification of earthquake-induced damage in historic masonry structures. *Automation in Construction*, **120**, 2020.
- [7] K. Worden, H. Sohn, C.R. Farrar, Novelty detection in a changing environment: Regression and interpolation approaches. *Journal of Sound and Vibration*, **258** (4), 741–761, 2002.
- [8] T. Fawcett, An introduction to ROC analysis. *Pattern Recognition Letters*, **27**, (8), 861–874, 2006.
- [9] L. Ierimonti, I. Venanzi, F. Ubertini, ROC analysis-based optimal design of a spatio-temporal online seismic monitoring system for precast industrial buildings. *Bulletin of Earthquake Engineering*, **19**, (3), 1441–1466, 2021.
- [10] J. Kullaa, Structural health monitoring under nonlinear environmental or operational influences. *Shock and Vibration*, **2014**, 2014.
- [11] C.M. Bishop, Pattern Recognition and Machine Learning. *Springer*, 2006.
- [12] E. García-Macías, F. Ubertini, MOVA/MOSS: Two integrated software solutions for comprehensive Structural Health Monitoring of structures. *Mechanical Systems and Signal Processing*, **143**, 2020.
- [13] G. Steenackers, P. Guillaume, Structural health monitoring of the Z-24 bridge in presence of environmental changes using modal analysis. *Conference Proceedings of the Society for Experimental Mechanics Series*, **18**, 2005.
- [14] KU Leuven, Z24 bridge benchmark, <https://bwk.kuleuven.be/bwm/z24>, accessed on 19th March 2021.