Least Angle Regression for early-stage identification of earthquake-induced damage in a monumental masonry palace: Palazzo dei Consoli

Enrique García-Macías[a,b,*], Filippo Ubertini[a]

[a]Department of Civil and Environmental Engineering, University of Perugia. Via G. Duranti, 93 - 06125 Perugia, Italy.
[b]Department of Structural Mechanics and Hydraulic Engineering, University of Granada, Av. Fuentenueva sn, 18002 Granada, Spain.

Abstract

This paper presents a novel methodology for earthquake-induced damage identification of historical constructions through sparse mutivariate regression. The proposed methodology comprises a first data cleansing stage using the minimum covariance determinant (MCD) method to mitigate the adverse effects related to the existence of outliers in the training feature dataset. Afterwards, a sparse multiple linear regression model (SMLR) is trained using the least-angle regression (LAR) model to eliminate the influence of environmental effects upon the selected features set. The proposed SMLR model allows to identify the optimal set of predictors in a fully automated way, minimizing the need for expert judgement in the process. The effectiveness of the proposed approach is demonstrated with an application case study of a monumental masonry palace, the Consoli Palace in Gubbio (Italy). The palace has been monitored with an aggregated static/dynamic/environmental SHM system since July 14th 2020. A recent seismic sequence of small intensity hit the palace on May 15th 2021 with a main earthquake of magnitude Mw 4.0. The epicenters of the main seismic event and the following aftershocks were located at a distance of 2-3 km far from the palace, making this case study a prominent example of a monumental construction subjected to near-field ground motion. The presented results demonstrate that a new damage condition arises in the Consoli Palace after the seismic sequence, although its severity remains at an early stage not detectable by visual inspections.

Keywords: Earthquake, Control charts, Damage detection, Historic buildings, Operational Modal Analysis, Structural health monitoring, Statistical Pattern Recognition.

1. Introduction

Cultural heritage buildings constitute especially sensitive assets in the built stock due to their strategic role in the tourism industry and their invaluable historical and social value. Indeed, the turnover generated in industries closely linked to cultural heritage represented 10.3% of the European Union GPD in 2018 [1] and, despite the Covid-19 recession, experts forecast the recovery to pre-pandemic levels by 2024 [2]. The maintenance of historic constructions is often troublesome due to their complex distribution of volumes and heterogenity, uncertainties in

*Corresponding author. Department of Civil and Environmental Engineering, University of Perugia. Via G. Duranti, 93 - 06125 Perugia, Italy. phone: +39 075 585 3908; fax: +39 075 585 3897

Email addresses: `enrique.garciamacias@unipg.it` (Enrique García-Macías), `filippo.ubertini@unipg.it` (Filippo Ubertini)

their materials and inner structure, as well as the presence of historical series of damage with uncertain origin and extension. Masonry historical buildings are particularly vulnerable to seismic actions due to their often low tensile strength, massive weight, poor connection between vertical and horizontal structural elements, and irregularities in plan and height [3, 4]. An example of this is the 2016-2017 Central Italy seismic sequence which caused complete destruction or heavy damage of important historical centres in four regions along the Apennines (Abruzzo, Lazio, Marche and Umbria) and 300 fatalities [5]. Apart from their vulnerability to natural material degradation and seismic actions, heritage structures must also face new challenges including increasing usage demands and visit flows, growing presence of corrosive pollutants, and climate change-induced more frequent utmost weather events [6]. Examples of sudden collapses such as the civic tower of Pavia in 1989 [7] and the Albiano Magra bridge in 2020 have evidenced the large risks associated with ageing degradation and poor maintenance. In this light, large financial efforts have been dedicated to R&D actions in the field of SHM of historical constructions since the seventies, although their extensive application to engineering practice remains marginal [8, 9]. Such a slow technological transfer is in part due to the lack of performance validation of damage identification techniques on full-scale structures under real operating conditions and, consequently, uncertain return on investments [10].

The management of long-term SHM systems falls within the pattern recognition paradigm formalized by Farrar *et al.* [11]. The basic idea is to establish relationships between damage states or classes and certain features extracted from the monitoring data by seeking for patterns in the response of the monitored structure. Within this paradigm, the stages of data cleansing, normalization, and damage classification are pivotal elements to attain effective damage identification. Data cleansing regards the process of filtering out uninformative or corrupted data (outliers), while data normalization relates the ability of separating the variability in the selected features induced by damage from those caused by environmental/operational conditions (EOC). Finally, damage classification concerns the inference of mappings between the extracted features and diagnosis classes. In the realm of historic structures, there is broad consensus on the importance of implementing aggregated SHM systems exploiting features extracted from diverse sensing technologies to so achieve a comprehensive damage identification. As a global damage identification technique, ambient vibration-based monitoring has become particularly widespread owing to their non-destructive nature and minimum invasiveness upon the normal fruition of the structure under study [12]. These techniques exploit experimentally identified modal signatures (i.e. natural frequencies, mode shapes, and damping ratios) as damage sensitive features (DSFs) [13, 14]. Nonetheless, their ability to detect local defects is rather limited (e.g. freezing/thawing cycles, chemical attack, corrosion) [15], whereby it is convenient to complement them with static monitoring such as the assessment of crack amplitudes, displacements or tilts [16]. The management of such aggregated long-term systems requires to handle large heterogeneous databases, framing the SHM problem into a Big Data and Machine Learning context [17–19]. While the statistical pattern recognition paradigm of SHM is formulated in broad general terms, the diverse steps involved in the process are generally highly case dependent.

The effectiveness of the damage identification is critically determined by the quality of the removal of EOC in the data normalisation stage. There are numerous works in the literature reporting the striking influence of manifold environmental/operational factors (e.g. temperature, humidity, traffic, wind) upon the dynamic/static response of civil engineering structures. Such factors provoke variations in the boundary conditions and the stiffness/mass properties of structures [20], resulting in fluctuations in their behaviour with different space- and time-scales. Among the extensive literature works reporting on these effects, it is worth noting the one by Zonno *et al.* [21], who investigated during one year the correlations between environmental factors and the modal properties of an adobe historic building, the San Pedro Apostol church in Peru. Their results reported daily and seasonal variations in the resonant frequencies of up to 1.5% and 8%, respectively, and identified the variations in the environmental humidity as the main driving mechanism. A recent work by Ceravolo *et al.* [22] reported the analysis of the environmental effects on the static and dynamic behaviour of the 17$^{\text{th}}$ century Sactuary of Vicoforte in Italy. Their study covered the analysis of monitoring data from a dense sensor network including wire gauges, pressure and load cells, crack meters, temperature sensors, and accelerometers, as well as climatic data from a meteorological station close to the monitoring site. The reported results evidenced positive correlations between environmental temperature and resonant frequencies, with average annual fluctuations around 5%. Such positive correlations are often observed in masonry structures, which is usually ascribed to thermal-induced crack closure phenomena (see e.g. [23–27]). Nevertheless, completely different correlations can appear in practice depending on the specific material, structural typology and mass distribution, solar radiation, thermal capacitance, etc. A noticeable example was provided by Gentile *et al.* [25] who reported the SHM of the Milan Cathedral in Italy. Their results showed negative correlations between resonant frequencies and temperature, which was ascribed to the constraints exerted by metallic tie-rods located in the cathedral. These harmless and reversible variations are often markedly larger than permanent changes induced by structural defects, resulting in a masking effect in the damage identification. To achieve an early-stage damage identification, it is thus indispensable to identify the main driving EOC and to remove their influence through proper statistical models.

Statistical models for pattern recognition in SHM can be generally classified as output-only or input-output models. Output-only models directly operate on the selected features to be normalised, without requiring monitoring data from EOC. Common approaches are Principal Component Analysis (PCA) [28], Factor Analysis [29], Autoassociative Neural Networks [30], time-series models [31] or Cointegration [32], to mention a few. These models exploit correlations between the selected features, in such a way that structural defects affecting such correlations will rise an anomaly. Nevertheless, since these models do not rely on predictor variables independent from structural damage, the physical interpretation of anomalies may be cumbersome and some structural defects may go unnoticed. Input-output models instead exploit correlations between damage-sensitive features and EOC. Examples of this approach are multiple linear regression (MLR) models [23], AutoRegressive with eXogeneous input models (ARX) [33], artificial neural networks [34], or support vector regression [35]. Although this approach requires monitoring data from EOC with the subsequent larger archive storage, the physical interpretation of anomalies is straightforward since model predictions are built on variables that are intrinsically independent from any structural damage. Nonetheless, a major difficulty in the definition of these models regards the selection of suitable sets of predictors which, again, is eminently case dependent and usually requires the intervention of expert judgement.

Alongside the multiple challenges reviewed above, one of the major obstacles for the extensive implementation of long-term SHM is the scarce number of research works in the literature reporting the successful damage identification of in-service full-scale structures. The Z24-Bridge in Switzerland firstly studied in 2001 by Maeck and co-authors [36] represents the most iconic case study in the field. Before its demolition in 1998, this bridge was instrumented with a dynamic SHM system, and subjected to a series of controlled damage scenarios for research purposes. Peeters and De Roeck [37] reported the identification of the damage scenarios through statistical pattern recognition of the time series of the bridge's resonant frequencies using ARX. The monitoring records were later made available to the scientific community, becoming a benchmark case study to test new damage identification techniques (see e.g. [38–40]). In the realm of historic constructions, since it is infeasible to induce controlled damage to any structure, most reported case studies in the literature focus on the application of SHM to assess and control restoration interventions [41, 42]. In this light, a noteworthy contribution was made by Masciotta *et al.* [43], who implemented a static/dynamic monitoring system to assess the interventions carried out in 2014-2015 to the Saint Torcato church (Portugal) with the aim of correcting structural damage induced by differential soil settlements. The reported post-rehabilitation results evidenced persistent shifts in the natural frequencies after one of the interventions. Mesquita *et al.* [44] reported the 1-year static/environmental monitoring of the 16$^{\text{th}}$ century Foz Côa Church (Portugal), an historical building affected by a series of old crack patterns originated by earthquakes occurred in 1755 and 1969. Correlation analyses between static and environmental data allowed to conclude that the behaviour of the church was stable and that no interventions were required, the variability in the response being only attributable to EOC with no risk to the structural integrity. The number of case studies reporting about the damage identification of historic constructions under in-service conditions is considerably lower. Amongst the few contributions in the literature, it is worth noting the work by Saisi *et al.* [45], who reported the damage identification of the 13$^{\text{th}}$ Century Gabbia Tower in Italy. After removing the environmental effects from the resonant frequencies of the tower by MLR, those authors identified permanent frequency decays after a far-field earthquake occurred in June 2013. Another noticeable contribution was reported by Ubertini and co-authors [46] on the damage identification of the San Pietro bell-tower in Perugia (Italy) after the 2016 Central Italy seismic sequence. Through a combination of MLR and PCA to filter out environmental effects, their results reported persistent decays in the resonant frequencies of the tower right after the main shocks of the seismic sequence. Interestingly, although the developed damage was not detectable by regular visual inspections, an independent non-linear time-history analysis of a numerical model of the tower reported similar frequency decays to the experimental ones with damage concentration in the base of the belfry.

With the aim of addressing the development of unsupervised damage identification of aggregated SHM systems with minimal support of expert judgement, this paper presents a novel methodology combining data cleansing and sparse MLR. The proposed approach comprises a first data cleansing stage using the MCD method to minimize the prejudicial effects related to the presence of outliers in the training dataset. Subsequently, a SMLR model is trained in an unsupervised fashion using the LAR method to eliminate the influence of environmental effects upon the dataset of damage-sensitive features. The proposed SMLR model automatically identifies the optimal set of EOC predictors, including both static and dynamic (time-delayed) predictors to accommodate environmental capacitance effects. The effectiveness of the proposed approach is demonstrated with an application case study of a monumental masonry palace, the Consoli Palace in Gubbio (Italy). The Consoli Palace has been instrumented since July 14$^{\text{th}}$ 2020 with an aggregated static/dynamic/environmental SHM system. A seismic sequence of small intensity recently hit the palace between May 15$^{\text{th}}$ and May 27$^{\text{th}}$ 2021. The sequence included a main earthquake of magnitude Mw 4.0 and peak ground acceleration (PGA) of 102.4 cm/s^2, followed by five 2.9 <Mw< 3.6 aftershocks in the following days. The most remarkable aspect of this case study regards the extremely closeness of the epicenters of the events, only 2-3 km far from the palace, making the investigated case study a unique example of a massive masonry building subjected to near-field strong motions. The presented results demonstrate the effectiveness of the proposed statistical pattern recognition approach to identify the earthquake-induced effects upon the resonant frequencies and the amplitudes of two major cracks of the Consoli Palace. Specifically, the conducted analyses report decays of up to 2% of the average resonant frequencies of the main bending and torsional modes of the palace. Concerning the analysis of the static data, the reported results evidence the appearance of persistent earthquake-induced closure of a major crack possibly related to an initial activation of an overturning mechanism of one of the façades of the building. Interestingly, no significant effects are observed upon the time series of the mode shapes of the palace nor new structural pathologies are found by preliminary in-situ inspections, indicating that the developed damaged condition remains at an early state level not observable by visual inspections. Given the singularity of the case study and the gap observed in the literature on the availability of field SHM data of heritage structures for damage identification, the time series of modal signatures, static and environmental monitoring data are made available for free use of the scientific community as part of the supplementary material.

The remainder of the paper is organized as follows. Section 2 introduces the proposed damage identification approach. Within this section, Subsection 2.1 overviews the general framework of anomaly detection through statistical pattern recognition, and Subsections 2.2 and 2.3 present the proposed data cleansing and data normalisation approaches, respectively. Section 3 presents the numerical results and discussion of the case study and, finally, Section 4 concludes the paper.

2. Removal of Environmental effects using LAR

2.1. Anomaly detection through statistical pattern recognition

Let us consider a SHM system tracking n different DSFs collected in an observation matrix $\mathbf{Y} = [\mathbf{y}_1, \ldots, \mathbf{y}_n] \in \mathbb{R}^{N\times n}$ containing N observations. As anticipated above, data normalization constitutes the process of subtracting the reversible variability in the selected features in $\mathbf{Y}$ induced by variations in EOC. This is typically achieved by training a certain statistical model using a set of t_p feature samples from $\mathbf{Y}$ defining a baseline in-control population, $\mathbf{Y}_{tp} \in \mathbb{R}^{t_p \times n}$, often referred to as the *training period* (see Fig. 1 (a)). This baseline dataset must statistically represent the healthy state of the structure under all possible EOC, being a one-year period often adopted. Once trained, the predictions of the model $\hat{\mathbf{Y}}$ can be used to phase out the variance due to EOC from $\mathbf{Y}$ forming the so-called residual error matrix $\mathbf{E} \in \mathbb{R}^{N \times n}$, that is:

$$\mathbf{E} = \mathbf{Y} - \hat{\mathbf{Y}}. \tag{1}$$

When the system remains healthy, matrix $\hat{\mathbf{Y}}$ reproduces the part of the variance of the features driven by EOC, while $\mathbf{E}$ only contains the residual variance stemming from modelling errors. Conversely, if a certain damage develops, this only affects the data contained in $\mathbf{Y}$ while matrix $\hat{\mathbf{Y}}$ remains unaltered. Therefore, matrix $\mathbf{E}$ concentrates the damage-induced variance and is apt for being used for damage identification.

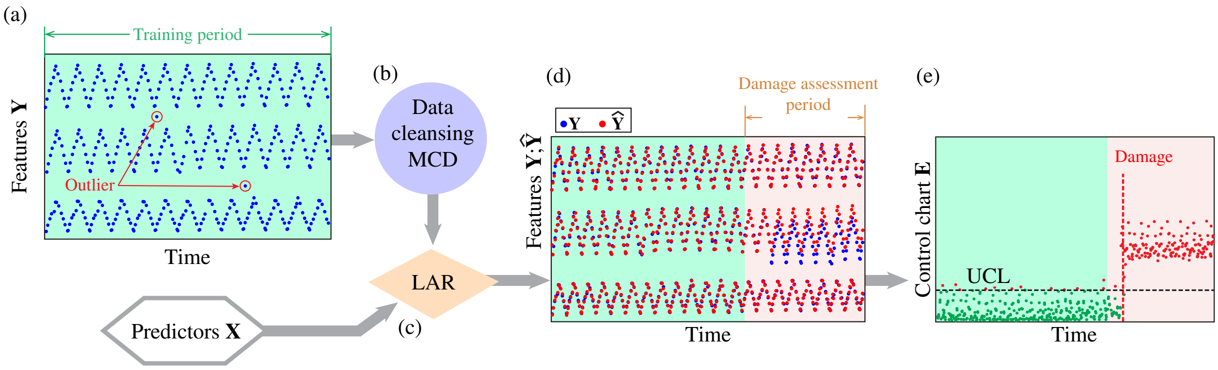

Figure 1: Flowchart of the proposed LAR-based data normalization approach for damage identification.

Damage classification can be performed by analysing the residuals in $\mathbf{E}$ through three different approaches: unsupervised learning, supervised learning, and semi-supervised learning as an intermediate solution. Supervised learning is often impractical in the context of historic constructions due to the serious difficulties to generate tagged damage data. Unsupervised learning instead simplifies the classification by tagging newly acquired data as damaged or non-damaged by analysing their discrepancies with respect to training population dataset (tagged as non-damaged or healthy). Unsupervised classification thus limits to damage detection or level 1 diagnostic (i.e. verify whether certain damage developed) and, to some qualitative extent, to damage quantification or level 2 diagnostic (i.e. a measure of the damage extension). This framework is unable to perform prognosis of the damage and to offer an estimate of the remaining life of the structure (level 3 diagnostic), being imperative to develop numerical models to such purposes. Nevertheless, level 1 diagnostic often suffices for the maintenance of heritage assets, whose criticality justifies the execution of in-situ inspections every time any fault is detected.

Novelty analysis and statistical process control charts are common tools to identify the presence of damage-induced anomalies in the time series of residuals in $\mathbf{E}$ in an unsupervised fashion. As sketched in Fig. 1 (e), control charts furnish in time a certain statistical distance accounting for nonconformities in the distribution of the residuals with respect to the training period. This allows to identify out-of-control processes as data points violating certain thresholds or in-control regions. A wide variety of control charts are available in the literature, although the The Hotelling's T^2 control chart [47] is possibly the most commonly used one in the realm of SHM. The plotted statistic T^2 (squared Mahalanobis distance) is defined as:

$$T_i^2 = r\left(\overline{\mathbf{E}} - \overline{\overline{\mathbf{E}}}\right)^{\mathrm{T}} \mathbf{\Sigma}_0^{-1} \left(\overline{\mathbf{E}} - \overline{\overline{\mathbf{E}}}\right),\ i = 1, 2, \ldots, N/r, \tag{2}$$

and the upper control limit (UCL) related to a $1-\alpha$ confidence level when residuals are ideally normally distributed reads:

$$UCL = \frac{krn - kr - rn + n}{kr - k - n + 1} F_{\alpha;n,kr-k-n+1}, \tag{3}$$

with parameter r in Eqs. (2) and (3) being an integer referred to as subgroup size, $\overline{\mathbf{E}}$ the mean of the residuals in the subgroup of the last r observations, and $\overline{\overline{\mathbf{E}}}$ and $\mathbf{\Sigma}_0$ the mean values and the covariance matrix of the residuals empirically estimated in the training period. Term $F_{\alpha;n,kr-k-n+1}$ denotes value of the cumulative F distribution with n and $kr - k - n + 1$ degrees of freedom for a $1 - \alpha$ confidence level.

The sensitivity of the control chart to detect small damage is highly influenced by the quality of the data normalization model. As previously discussed, the use of input-output statistical models facilitates the interpretation of nonconformities, although these models heavily rely on the suitable selection of the predictors set. To optimize this process and minimize the need for expert judgement, an automated procedure based on the LAR method is implemented in this work. Furthermore, the quality of classification may be also considerably affected by the presence of outliers in the monitoring data. Outliers are always present to a certain degree in every feature set in SHM, stemming from manifold sources like noise, identification errors, faulty sensors, imperfect mounting, etc. Their presence in the training period has a twofold effect: (i) outliers bias the computation of the parameters of the data normalization model; and (ii) hinder the proper definition of the UCL. Note in Eq. (3) that the definition of UCL depends upon the statistical moments of the residuals in the training population. Therefore, the presence of outliers will bias such moments, reducing the damage sensitivity of the classification. To minimize such effects, an outlier elimination approach based on the MCD method is proposed in this work. These procedures are assembled in a new methodology for aggregated long-term SHM sketched in Fig. 1, which comprises the following five sequential steps:

(a) Definition of the baseline population (training period) of damage-sensitive features and potential predictors describing the EOC variability.

(b) Data cleansing using MCD.

(c) Construction of the statistical model for data normalization using MLR and LAR (SMLR).

(d) Computation of the residual matrix $\mathbf{E}$ by subtracting the predictions by the statistical model $\hat{\mathbf{Y}}$ from the observation matrix $\mathbf{Y}$.

(e) Damage detection through novelty analysis of the Hotelling's control chart.

In the remainder of this section, the theoretical fundamentals of MCD and SMLR using LAR are presented in Sections 2.2 and 2.3, respectively.

2.2. Data cleansing using MCD

The MCD method introduced by Rousseau [48] is a robust estimator of multivariate location and scatter commonly used for outlier detection. Assuming Gaussian-distributed data, the MCD method seeks a subset of given size with lowest sample covariance. In the context of this work, the presence of outliers ought to be minimized in the training population $\mathbf{Y}_{tp}$. Let $H_1 \subset \left\{1,\ldots,t_p\right\}$ be an h-subset with $|H_1| = h$, and $\boldsymbol{\mu}_1 = (1/h)\sum_{i\in H_1}\mathbf{y}_i$ and $\boldsymbol{\Sigma}_1 = [1/(h-1)]\sum_{i\in H_1}(\mathbf{y}_i-\boldsymbol{\mu}_1)(\mathbf{y}_i-\boldsymbol{\mu}_1)^{\mathrm{T}}$ being the empirical mean and covariance matrix of the data in H_1, respectively. The Mahalanobis distances of all the data samples in the training population read:

$$d_1(\mathbf{y}_i) = \sqrt{(\mathbf{y}_i-\boldsymbol{\mu}_1)^{\mathrm{T}}\boldsymbol{\Sigma}_1^{-1}(\mathbf{y}_i-\boldsymbol{\mu}_1)} \quad \text{for} \quad i = 1,\ldots,t_p. \tag{4}$$

Now take H_2 another h-subset such that $\{d_1(i)\,;\, i\in H_2\} := \left\{(d_1)_{1:t_p},\ldots,(d_1)_{h:t_p}\right\}$ where $(d_1)_{1:t_p} \leq (d_1)_{2:t_p} \leq \ldots \leq (d_1)_{t_p:t_p}$ are the ordered distances, and compute $\boldsymbol{\mu}_2$ and $\boldsymbol{\Sigma}_2$ based on H_2. Then $\det(\boldsymbol{\Sigma}_2) \leq \det(\boldsymbol{\Sigma}_1)$ holds with equality if and only if $\boldsymbol{\mu}_2 = \boldsymbol{\mu}_1$ and $\boldsymbol{\Sigma}_2 = \boldsymbol{\Sigma}_1$. This process, also known as the concentration step (C-step), can be iteratively repeated as follows:

1. Select h observations from the training dataset $\mathbf{Y}_{tp}$ conforming H_s.
2. Compute the empirical covariance $\boldsymbol{\mu}_s$ and $\boldsymbol{\Sigma}_s$.
3. Compute the Mahalanobis distances $d_s(\mathbf{y}_i)$, $i = 1,\ldots,t_p$.
4. Sort the Mahalanobis distances, and select the h observations having the smallest distances to form H_{s+1}.
5. Stop if $\det(\boldsymbol{\Sigma}_{s+1}) = 0$ or $\det(\boldsymbol{\Sigma}_{s+1}) = \det(\boldsymbol{\Sigma}_s)$, otherwise go to step 2.

The sequence $\det(\boldsymbol{\Sigma}_1) \geq \det(\boldsymbol{\Sigma}_2) \geq \det(\boldsymbol{\Sigma}_3) \geq \det(\boldsymbol{\Sigma}_4) \geq \ldots$ is non-negative, so the algorithm always converges in finite steps as there is a finite number of h-subsets [49]. Nevertheless, the final calculation of the covariance matrix may not converge to the global minimum since it highly depends upon the definition of the initial subset H_1. The evaluation of all $\binom{t_p}{h}$ subsets of size h may lead to prohibitive computational costs as the number of data samples in the training period is usually large. As an alternative solution, Rousseau and Driessen [50] proposed a Fast-MCD algorithm based upon the application of the raw MCD to a large number of initial candidates for H_1. Specifically, the algorithm comprises three sequential stages when the number of samples is considerably large ($t_p > 600$ [50]): Firstly, several disjoint subsets are drawn from the dataset (a recommended number of $n+1$ subsets [49]) and several C-steps are applied to each subset keeping the solutions with lowest determinants. Secondly, the subsets are pooled together forming a merged set, and the previously obtained best h-subsets are used as the initial subset H_1. For every initial subset, several C-steps are applied and the solutions with lowest determinants are kept. Finally, the raw MCD method is applied to the full dataset keeping the solution with lowest determinant obtained by considering the previously obtained solutions as initial subsets H_1. The algorithm is given in detail in references [49–52].

For illustration purposes, Fig. 2 shows a toy example of outlier detection using the MCD method. The Fast-MCD algorithm is applied to a dataset of $t_p = 1250$ observations and $n = 2$ variables, in which 1000 and 250 (outliers) observations were drawn from two bivariate normal distributions $\mathcal{N}_1$ and $\mathcal{N}_2$:

$$\mathcal{N}_1\left(\begin{bmatrix}0\\0\end{bmatrix},\begin{bmatrix}1 & 1.5\\1.5 & 3\end{bmatrix}\right),\quad \mathcal{N}_2\left(\begin{bmatrix}2\\-2\end{bmatrix},\begin{bmatrix}1 & -0.5\\-0.5 & 1\end{bmatrix}\right). \tag{5}$$

The dimension h of the subsets has been selected according to the recommendation by Rousseau and Driessen [50] as $h \approx (n + p + 1)/2 = 626$. The Fast-MCD algorithm has been applied to the synthetic dataset starting from five subsets with 300 samples, and the obtained analysis results are shown in Fig. 2. The scatter plot in Fig. 2 (a) shows the optimal h-set and the remaining $t_p - h$ samples with blue and red solid points, respectively. In this figure, the 99% tolerance ellipses are also shown. Figure 2 (b) depicts the distance-distance plot, which represents the robust Mahalanobis distances (based upon the mean and covariance estimates after applying the MCD method) versus the distances computed from the complete dataset. On both axes, threshold limits corresponding to a 99% confidence level and defined as $\sqrt{\chi^2_{2,0.99}} = 3.0349$ are also indicated. It is clear in this figure that the MCD concentrates the data samples drawn from $\mathcal{N}_1$ in the h-subset, while isolating most of the samples from $\mathcal{N}_2$ as outliers. From the analysis of Fig. 2 (b), it is found that the classification using the MCD method identifies 231 outliers, while the direct analysis of the dataset only leads to 26 outliers.

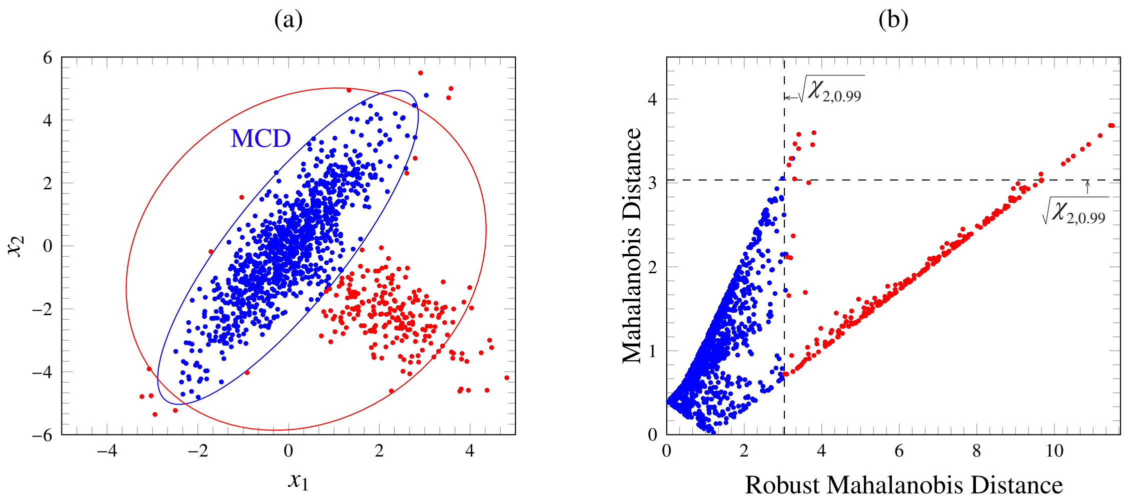

Figure 2: Toy example of two overlapping bivariate Gaussian distributions: scatter plot with 99% tolerance ellipses before and after the application of the MCD method (a), and distance-distance plot (b).

2.3. Optimal sparse MLR model using LAR (SMLR)

MLR models exploit linear correlations between the n selected features (estimators or dependent variables) and a set of p independent exploratory variables (predictors or independent variables), which are typically taken from monitoring data of EOC. The predictions by MLR of the observation matrix $\mathbf{Y}$ are obtained as:

$$\hat{\mathbf{Y}} = [\mathbf{1}_{N\times 1}, \mathbf{X}] \begin{bmatrix} \boldsymbol{\beta}_0 \\ \boldsymbol{\beta} \end{bmatrix}, \tag{6}$$

where $\mathbf{1}_{N\times 1}$ is a column vector of ones and $\mathbf{X} = \left[\mathbf{x}_1, \ldots, \mathbf{x}_p\right] \in \mathbb{R}^{N\times p}$ is an observation matrix with columns containing the time series of the p selected predictors. Term $\boldsymbol{\beta}_0 \in \mathbb{R}^{1\times n}$ is a row vector of intercept terms and $\boldsymbol{\beta} = \left[\boldsymbol{\beta}_1, \ldots, \boldsymbol{\beta}_p\right] \in \mathbb{R}^{p\times n}$ is a matrix of linear regression coefficients. As anticipated above, the quality of the MLR model for feature normalization is highly dependent upon the quality of the predictor selection. This motivates the interest of sparse linear regression methods to select the optimal set of predictors among a large database of potential variables. To this aim, the LAR algorithm is proposed in this work. LAR is an efficient algorithm for model selection of sparse linear models [53]. Let us first consider centred and normalized versions of the predictors $\mathbf{x}_j$ in $\mathbf{X}$ arranged in a normalized predictor matrix $\mathbf{X}_n$, as well a centred version of an arbitrary i-th estimator $\mathbf{y}_i$ in $\mathbf{Y}$:

$$\mathbf{x}_j^n = \frac{\mathbf{x}_j - \overline{\mathbf{x}}_j}{\sigma_{x_j}},\ j = 1, \ldots, p, \quad \mathbf{y}_i^n = \mathbf{y}_i - \overline{\mathbf{y}}_i,\ i = 1, \ldots, n, \tag{7}$$

with

$$\sigma_{x_j} = \sqrt{\frac{1}{N-1}\left(\mathbf{x}_j - \overline{\mathbf{x}}_j\right)\left(\mathbf{x}_j - \overline{\mathbf{x}}_j\right)^{\mathrm{T}}}. \tag{8}$$

A regression method estimates the coefficients vector $\boldsymbol{\beta}_i^*$ relating the normalized predictor matrix $\mathbf{X}_n$ and the i-th normalized output $\mathbf{y}_i^n$ as:

$$\mathbf{y}_i^n = \mathbf{X}_n \boldsymbol{\beta}_i^*. \tag{9}$$

Once determined, the coefficients in $\boldsymbol{\beta}_i^*$ can be readily converted to the original scaled model as $\boldsymbol{\beta}_i = \sigma_{x_j}\boldsymbol{\beta}_i^*$, and the i-th intercept term can be computed as $\boldsymbol{\beta}_{0,i} = \bar{\mathbf{y}}_i - \bar{\mathbf{X}}\boldsymbol{\beta}_i^*$.

For simplicity of the notation, indexes i, j, n and $*$ in Eq. (9) are dropped in the subsequent derivations. In general, let us consider a univariate linear model defined as $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$, with $\boldsymbol{\varepsilon}$ being a zero mean error term. In the context of SHM, the predictors in $\mathbf{X} \in \mathbb{R}^{N\times p}$ may contain a large number of variables (e.g. environmental temperatures, humidity, wind intensity, etc., as well as delayed variations to account for capacitance effects), some of which may not have significant effects upon the variance of $\mathbf{y}$. In order extract a subset of the most representative predictors in $\mathbf{X}$, the linear model in Eq. (9) can be assumed to be sparse. In this light, a sparse regression method will estimate the coefficients in $\boldsymbol{\beta} \in \mathbb{R}^p$ corresponding to the most representative predictors, while allocating zeroes to the least influential ones. Let use denote the active set $\mathcal{A}$ as the indices in $\boldsymbol{\beta}$ corresponding to non-zero elements, and the inactive set $\mathcal{I}$ as the complementary set of $\mathcal{A}$. Also, let us note $\mathbf{X}_{\mathcal{A}}$ consisting of a subset of predictors obtained by extracting from $\mathbf{X}$ the columns corresponding to the indices in $\mathcal{A}$. Among the variety of sparse regression methods available in the literature (see e.g. [54]), the LAR model is adopted in this work. The LAR algorithm is a forward stepwise regression approach set out by Efron *et al.* [53]. The active set is initialized to be empty, $\mathcal{A} = \emptyset$, and the indexes of all the predictors are included in the inactive set, i.e. $\mathcal{I} = \{1, \ldots, p\}$. The algorithm starts by assuming the coefficient vector $\boldsymbol{\beta}^{(0)} = \mathbf{0}$ and, thus, the residual $\boldsymbol{\varepsilon}_0 = \mathbf{y} - \hat{\mathbf{y}}^{(0)}$, with $\hat{\mathbf{y}}^{(0)} = \mathbf{0}$ being the initial prediction of the linear model. The first predictor to be included in the active set is the one which has the largest correlation with the current residual, that is:

$$c = \max_{i \in \mathcal{I}} \left| \mathbf{x}_i^{\mathrm{T}} \boldsymbol{\varepsilon}_0 \right| . \tag{10}$$

Let us assume the index j is the one corresponding to c, and thus the index to be added to the active set $\mathcal{A}$. Then, the regression coefficients are moved towards their least-square value, until some other predictor has as much correlation with the current residual. This corresponds to an updating of the form:

$$\boldsymbol{\beta}^{(1)} = \boldsymbol{\beta}^{(0)} + \gamma \left(\boldsymbol{\beta}_{OLS}^{(1)} - \boldsymbol{\beta}^{(0)} \right), \tag{11}$$

with $\boldsymbol{\beta}_{OLS}^{(1)}$ being the ordinary least-squares (OLS) solution:

$$\boldsymbol{\beta}_{OLS}^{(1)} = \left(\mathbf{X}_{\mathcal{A}}^{\mathrm{T}} \mathbf{X}_{\mathcal{A}} \right)^{-1} \mathbf{X}_{\mathcal{A}}^{\mathrm{T}} \mathbf{y}, \tag{12}$$

and γ the step length $0 < \gamma \leq 1$. Accordingly, the prediction by the linear model and the residual are updated as $\hat{\mathbf{y}}^{(1)} = \hat{\mathbf{y}}^{(0)} + \gamma \left(\hat{\mathbf{y}}_{OLS}^{(1)} - \hat{\mathbf{y}}^{(0)} \right)$ and $\boldsymbol{\varepsilon}_1 = \mathbf{y} - \hat{\mathbf{y}}^{(1)}$, respectively, with $\hat{\mathbf{y}}_{OLS}^{(1)}$ being the least squares solution, i.e. $\hat{\mathbf{y}}_{OLS}^{(1)} = \mathbf{X}_{\mathcal{A}} \boldsymbol{\beta}_{OLS}^{(1)}$. In order to determine the value of γ, one seeks the smallest positive value where correlations with the current residual become equal, i.e. $\mathbf{x}_{i \in \mathcal{I}}^{\mathrm{T}} \boldsymbol{\varepsilon}_1 = \mathbf{x}_j^{\mathrm{T}} \boldsymbol{\varepsilon}_1$, leading to:

$$\mathbf{x}_{i \in \mathcal{I}}^{\mathrm{T}} \left[\mathbf{y} - \hat{\mathbf{y}}^{(0)} - \gamma \left(\hat{\mathbf{y}}_{OLS}^{(1)} - \hat{\mathbf{y}}^{(0)} \right) \right] = \mathbf{x}_j \left[\mathbf{y} - \hat{\mathbf{y}}^{(0)} - \gamma \left(\hat{\mathbf{y}}_{OLS}^{(1)} - \hat{\mathbf{y}}^{(0)} \right) \right]. \tag{13}$$

Solving the expression in Eq. (13) for γ, one gets:

$$\gamma = \frac{\left(\mathbf{x}_i - \mathbf{x}_j \right)^{\mathrm{T}} \left(\mathbf{y} - \hat{\mathbf{y}}^{(0)} \right)}{\left(\mathbf{x}_i - \mathbf{x}_j \right)^{\mathrm{T}} \left(\hat{\mathbf{y}}_{OLS}^{(1)} - \hat{\mathbf{y}}^{(0)} \right)} = \frac{\left(\mathbf{x}_i - \mathbf{x}_j \right)^{\mathrm{T}} \boldsymbol{\varepsilon}_0}{\left(\mathbf{x}_i - \mathbf{x}_j \right)^{\mathrm{T}} \mathbf{d}}, \tag{14}$$

where $\mathbf{d} = \hat{\mathbf{y}}_{OLS}^{(1)} - \hat{\mathbf{y}}^{(0)}$ is the direction of the walk. Note that $\mathbf{d}$ is orthogonal to $\boldsymbol{\varepsilon}_0$, therefore we have $\mathbf{x}_i \boldsymbol{\varepsilon}_0 = \mathbf{x}_i \mathbf{d} \equiv c$. Since the predictors in $\mathbf{X}$ are assumed to be normalized, i.e. $|\mathbf{x}_i| = 1$, the condition in Eq. (13) may be interpreted in terms of dot products as $\mathbf{x}_i^{\mathrm{T}} \boldsymbol{\varepsilon}_1 = \cos \theta_i = \mathbf{x}_j^{\mathrm{T}} \boldsymbol{\varepsilon}_1 = \cos \theta_j$, i.e. $\theta_i = \theta_j$. This bisection condition is equivalent to imposing the movement of the predictor coefficients along the equiangular direction between the predictors $\mathbf{x}_i$ and $\mathbf{x}_j$ and the current residual $\boldsymbol{\varepsilon}_1$, i.e. the least angle direction. Furthermore, since the sign of the correlation between variables is irrelevant, Eq. (14) can be in general written as:

$$\gamma = \min_{i \in \mathcal{I}} \left\{ \frac{\mathbf{x}_i^{\mathrm{T}} \boldsymbol{\varepsilon}_0 - c}{\mathbf{x}_i^{\mathrm{T}} \mathbf{d} - c}, \; \frac{\mathbf{x}_i^{\mathrm{T}} \boldsymbol{\varepsilon}_0 + c}{\mathbf{x}_i^{\mathrm{T}} \mathbf{d} + c} \right\}, \quad 0 < \gamma \leq 1. \tag{15}$$

This process can be performed iteratively $p-1$ times according to the following steps:

1. Initialize the coefficient vector $\boldsymbol{\beta}^{(0)} = \mathbf{0}$, the fitted vector $\hat{\mathbf{y}}^{(0)} = \mathbf{0}$, the active set $\mathcal{A} = \emptyset$ and the inactive set $\mathcal{I} = \{1, \dots, p\}$.
2. **for** k=0 to p-2 **do**
3. Update the residual $\boldsymbol{\varepsilon}_k = \mathbf{y} - \hat{\mathbf{y}}^{(k)}$.
4. Find the maximum correlation $c = \max_{i \in \mathcal{I}} \left| \mathbf{x}_i^{\mathrm{T}} \boldsymbol{\varepsilon}_k \right|$.
5. Move variable corresponding to c from $\mathcal{I}$ to $\mathcal{A}$.
6. Compute the least squares solutions $\boldsymbol{\beta}_{OLS}^{(k+1)} = \left(\mathbf{X}_{\mathcal{A}}^{\mathrm{T}} \mathbf{X}_{\mathcal{A}} \right)^{-1} \mathbf{X}_{\mathcal{A}}^{\mathrm{T}} \mathbf{y}$ and $\mathbf{y}_{OLS}^{(k+1)} = \mathbf{X}_{\mathcal{A}} \boldsymbol{\beta}_{OLS}^{(k+1)}$.
7. Compute the direction of the walk $\mathbf{d} = \hat{\mathbf{y}}_{OLS}^{(k+1)} - \hat{\mathbf{y}}^{(k)}$.
8. Compute the step length $\gamma = \left\{ \frac{\mathbf{x}_i^{\mathrm{T}} \boldsymbol{\varepsilon}_k - c}{\mathbf{x}_i^{\mathrm{T}} \mathbf{d} - c}, \frac{\mathbf{x}_i^{\mathrm{T}} \boldsymbol{\varepsilon}_k + c}{\mathbf{x}_i^{\mathrm{T}} \mathbf{d} + c} \right\}, \quad 0 < \gamma \leq 1.$
9. Update the regression coefficients: $\boldsymbol{\beta}^{(k+1)} = \boldsymbol{\beta}^{(k)} + \gamma \left(\boldsymbol{\beta}_{OLS}^{(k+1)} - \boldsymbol{\beta}^{(k)} \right)$.
10. Update the fitted vector: $\hat{\mathbf{y}}^{(k+1)} = \hat{\mathbf{y}}^{(k)} + \gamma \left(\hat{\mathbf{y}}_{OLS}^{(k+1)} - \hat{\mathbf{y}}^{(k)} \right)$.
11. **end for**

The algorithm at step p is completed with the full OLS solution, i.e. $\boldsymbol{\beta}^{(p)} = \left(\mathbf{X}^{\mathrm{T}} \mathbf{X} \right)^{-1} \mathbf{X}^{\mathrm{T}} \mathbf{y}$. The main output is the series of coefficients $\mathcal{B} = \left\{ \boldsymbol{\beta}^{(0)}, \dots, \boldsymbol{\beta}^{(p)} \right\}$, which represent different linear models with decreasing level of sparsity. Finally, the best regression model in $\mathcal{B}$ can be selected according to certain quality criteria such as the maximum number of selected predictors or the minimum residual sum-of-squares (RSS), or information based statistical criteria like the Bayesian Information Criterion (BIC) [55] or the Akaike Information Criterion (AIC) [56]. To better illustrate the working mechanism of the LAR algorithm, Fig. 3 presents the geometrical interpretation of the determination of the coefficient parameters in the case of 3 covariates $\mathbf{x}_1$, $\mathbf{x}_2$, and $\mathbf{x}_3$. In the initial step $k = 0$ in Fig. 3 (a), $\mathbf{x}_1$ is selected as the first predictor since it has the largest correlation with the initial residue $\boldsymbol{\varepsilon}_0 = \mathbf{y}$. Therefore, $\boldsymbol{\beta}^{(1)} = \gamma \boldsymbol{\beta}_{OLS}^{(1)}$, and we need to determine the step length γ. To do so, we need to apply the equiangular condition in Eq. (14). In this initial case ($k = 0$), this equation reduces to:

$$\mathbf{x}_1 \underbrace{\left(\mathbf{y} - \gamma \boldsymbol{\beta}_{OLS}^{(1)} \right)}_{\boldsymbol{\varepsilon}_{12}} = \mathbf{x}_2 \underbrace{\left(\mathbf{y} - \gamma \boldsymbol{\beta}_{OLS}^{(1)} \right)}_{\boldsymbol{\varepsilon}_{12}}, \tag{16}$$

and

$$\mathbf{x}_1 \underbrace{\left(\mathbf{y} - \gamma \boldsymbol{\beta}^{(1)}_{OLS}\right)}_{\boldsymbol{\varepsilon}_{13}} = \mathbf{x}_3 \underbrace{\left(\mathbf{y} - \gamma \boldsymbol{\beta}^{(1)}_{OLS}\right)}_{\boldsymbol{\varepsilon}_{13}}. \tag{17}$$

The bisection condition in Eq. (16) leads to a solution where the residue vector $\boldsymbol{\varepsilon}_{12}$ has the same angle α_{12} with the inactive predictor $\mathbf{x}_2$ and the active predictor $\mathbf{x}_1$. Similarly, Eq. (17) leads to a different solution where the residue vector $\boldsymbol{\varepsilon}_{13}$ has the same angle α_{13} with $\mathbf{x}_3$ and $\mathbf{x}_1$. In this example, $\mathbf{x}_3$ has the least angle ($\boldsymbol{\varepsilon}_{13} < \boldsymbol{\varepsilon}_{12}$) and, therefore, Eq. (17) determines the step length γ. In the second step ($k = 1$), the direction of the walk is given by the OLS projection of $\mathbf{y}$ onto the active set defined by $\mathbf{x}_1$ and $\mathbf{x}_3$, i.e. $\hat{\mathbf{y}}^{(2)}_{OLS}$. This procedure is repeated until reaching the full OLS as shown in Fig. 3 (b), where covariates $\mathbf{x}_1$, $\mathbf{x}_3$, and $\mathbf{x}_2$ are added sequentially to the regression. Variables $\hat{\mathbf{y}}^{(1)}_{OLS}$ and $\hat{\mathbf{y}}^{(2)}_{OLS}$ represent the partial OLS solutions on $\mathbf{x}_1$ and $\{\mathbf{x}_1, \mathbf{x}_3\}$, respectively, while $\hat{\mathbf{y}}^{(3)}_{OLS} = \hat{\mathbf{y}}^{(3)}$ represents the full OLS solution.

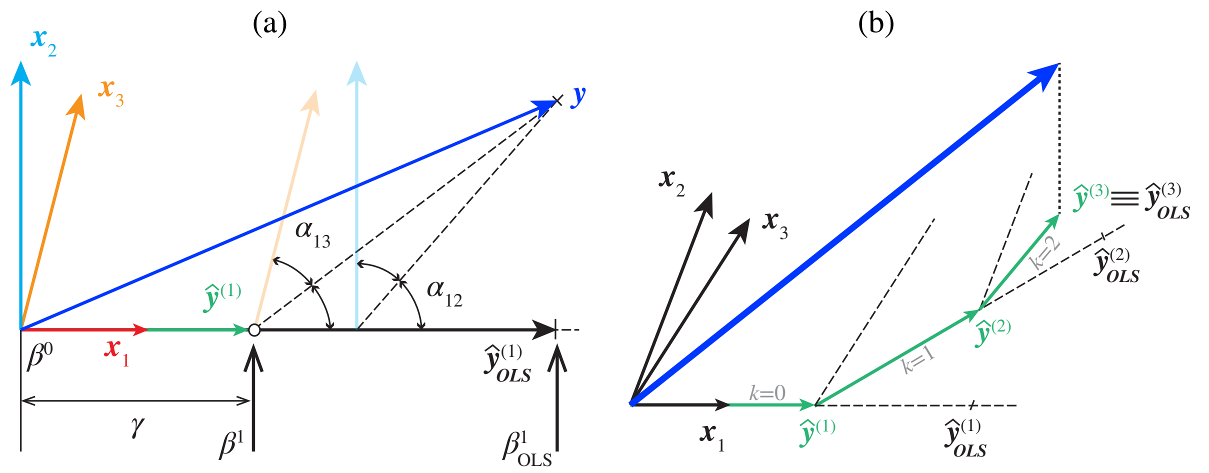

Figure 3: Geometric representation of the LAR algorithm in the case of 3 covariates. (a) Initial step $k = 0$, and (b) determination of the complete solution path.

3. Application case study: the Consoli Palace

3.1. Description of the structure and monitoring layout

The Consoli Palace is the most emblematic building in the medieval town of Gubbio in central Italy. The palace forms part of a monumental ensemble built in the 14th century together with the Podestà Palace and a vaulted hanging square, named “Piazza Grande” (see Fig. 4 (a)). Although originally dedicated to host the legislative/executive and judicial courts, the Consoli and the Podestà palaces respectively house the Civic Museum and the municipality headquarters of Gubbio since the early nineties. The Consoli Palace presents a 40×20 m rectangular plan and it is structurally constituted by calcareous stone masonry thick bearing walls and vaulted ceilings. The foundations of the building are placed on two levels with a drop of approximately 10 m to accommodate the steep slope of the terrain (see Fig. 4 (b)), giving the building an irregular distribution in height. The south façade includes a panoramic loggia and stands 60 m from the ground level until a 13 m high bell-tower rising from the roof level, while the north façade has a height of about 30 m between the square's level and the roof. As an extraordinary example of a monumental masonry structure, the Consoli Palace has been the case study of a number of research projects. Interested readers are referred to references [57–60] for further details on the architecture and some of the investigations carried out in the palace.

Gubbio is located on the Umbria-Marche Apennine Mountains, an area of almost continuous seismicity and catalogued as a natural laboratory for seismic studies (TABOO - Alto Tiberina Near Fault Observatory). The seismic activity in this area is dominated by the Gubbio fault on which the city rises. The Gubbio fault is a 22-km-long normal fault pertaining to a set of active SW-dipping sub-parallel normal faults known as the seismogenic Umbria Fault System (UFS) [61]. The UFS faults are antithetic splays located in the hanging wall of the regional Alto Tiberina Fault (ATF), a major east-dipping low-angle (20°) normal fault. Geophysical data and seismological studies characterized the geometry of the Gubbio fault, revealing a listric trend and the intersection with the ATF at a depth of 6 km [62]. The strongest registered earthquake to date was on April 29th 1984 (Mw 5.6) with epicenter located ≈10 km south of the town of Gubbio [63] and causing important damage to the Consoli Palace. Later on, an intense seismic activity started on August 26th 2013 with a Mw 3.8 event and followed by several aftershocks with 3<Mw<4.9 through the entire 2014. Since then, the seismic activity has been quite constant, with one single major event occurred on October 2016 with a magnitude Mw 3.0 [63].

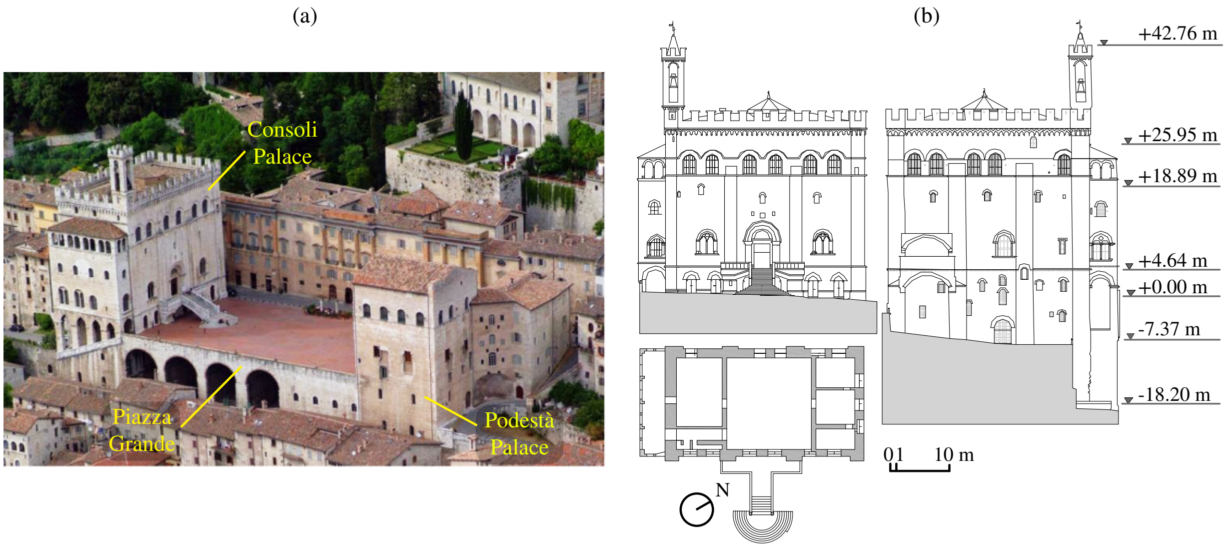

Figure 4: Aerial view of the monumental ensemble of the Consoli Palace, the Podestà Palace and the Piazza grande (a); Plan and elevation views of the Consoli Palace (b).

Within the framework of a national research project on the surveillance and identification of ageing deterioration of historical constructions, an aggregated static/dynamic/environmental SHM system was installed in the Consoli Palace in July 2017 and remained active until July 2020. The monitoring system comprised three uniaxial accelerometers, two Linear Variable Displacement Transducers (LVDTs), and two temperature sensors. The analysis of the monitoring data acquired during this first phase of the SHM system was reported by Kita and co-authors [59]. Nonetheless, an important upgrade of the system was carried out in July 2020 with a considerable increase in the number of sensors, which remains active up to date and is subject of study in this work. The layout of the system is sketched in Fig. 5 (b) and comprises: twelve accelerometers, four LVDTs, six temperature sensors, and a data acquisition system (DAQ). The accelerometers, labelled with A1 to A12 in Fig. 5 (b) and shown in Fig. 5 (a), are high-sensitivity uni-axial piezoelectric accelerometers model PCB393B12 (10 V/g, broadband resolution 800 µg, and ±0.5 g measurement range). They are deployed in the two main floors of the building and on the roof level, namely at heights of +4.64 m (Arengo Hall), +18.89 m (Nobili Hall), and +29.77 m. Three accelerometers are located in each floor, with a biaxial station in the south façade and a mono-axial station in the north façade monitoring ambient accelerations along the y-direction. Such a configuration is aimed at characterizing rigid diaphragm motions of the floors and global torsional rotations of the building. A similar scheme has been also considered on the roof level but with the consideration of two accelerometers (A11 and A12) located at the centre of the east and west façades, respectively, with the aim of monitoring out-of-plane bending movements. Four S-series LVDTs (50 mm measurement range and < 0.3 µm resolution), labelled with D1 to D4 in Fig. 5 (b), are also installed monitoring the opening/cracking of two major cracks previously identified in reference [59]. Specifically, LVDTs D1 and D3 monitor two cracks in the second level of the palace, whose origin is possibly related to the overturning mechanism of the loggia in the south façade. Instead, LVDTs D2 and D4 monitor the movements at two levels (second and third floors) of a second major crack located in the north façade of the palace and propagating downwards until reaching the west façade. The origin of the latter may be indicative of the initiation of a failure mechanism of overturning of the northern part of the west façade. Finally, six K-type thermocouples, labelled with T1 to T6, are also deployed in the palace. Thermocouples T1 to T4 are located aside LVDTs D1 to D4 measuring the surface temperature of the masonry, while thermocouples T5 and T6 monitor the ambient temperature at the roof level and the third level, respectively. The accelerometers, thermocouples T1 and T2, and LVDTs D1 and D2 are connected to a DAQ system located in the third level and powered from an uninterruptible power supply. (Fig. 5 (a.4)). The DAQ, model NI Compact DQ-9132 (1.33 GHz Dual-Core Atom, 4 slots, Windows Embedded Standard 7, 16 GB SD storage), is equipped with three NI 9234 acceleration acquisition modules (4 channels, 24-bit resolution, 102-dB dynamic range and anti-aliasing filters) and a NI 9219 acquisition module for LVDTs and thermo-couples (4 channels, 24-bit resolution, ±60 V range, 100 S/s). Conversely, the monitoring records of crack-meters D3 to D4 and thermo-couples T3 to T6 are transferred through wireless communication to a wifi router (Fig. 5 (a.3)).

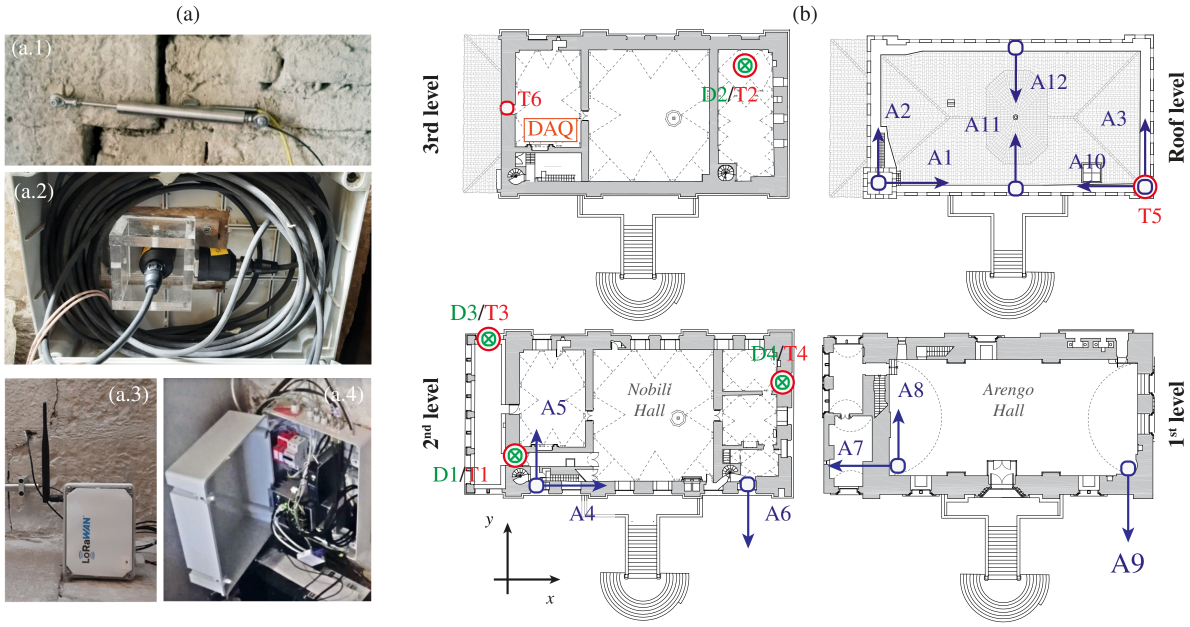

Figure 5: Views of the monitoring and acquisition equipment (a; a.1 - LVDT, a.2 - Accelerometer, a.3 - wireless data transmission, a.4 - DAQ), and layout of the continuous monitoring system (b).

Ambient vibrations are sampled at 40 Hz, and crack amplitudes and temperatures from channels D1-D2 and T1-T2 are sampled at 0.1 Hz. The monitoring records are stored in separate binary data files containing 30-min-long recordings. A Labview script is implemented and used for data acquisition and quality control from remote, including amplitude and spectral plots. Single acquisitions from sensors D3-D4 and T3-T6 are taken every 30 minutes and collected in common text files on a daily basis. The recorded data are sent through the internet to a cloud archive, where they are accessed by a remote server computer in the Laboratory of Structural Dynamics of the University of Perugia. The monitoring data are collected and processed in an in-house software code named MOSS, the Italian acronym of SHM. The software code, whose first release was reported in reference [60], implements all the steps involved in SHM as a statistical pattern recognition, including signal pre-processing, automated dynamic identification, feature extraction, data cleansing, normalisation, and novelty analysis.

3.2. Dynamic Identification and continuous monitoring

A dense ambient vibration test (AVT) was conducted at 13:00 pm CEST on May 7th 2021 in order to characterize the modal signatures of the Consoli Palace. The test comprised 19 uni-axial piezoelectric accelerometers (same technical specifications as those used for the continuous monitoring) with positions sketched in Fig. 6. Specifically, the accelerometers layout used in the continuous monitoring system (Fig. 6 (a)) was complemented with nine accelerometers covering the two orthogonal directions of the palace at the roof level, and three stations at the top level of the bell-tower to monitor rigid-diaphragm motions. Such a configuration was designed with a twofold purpose: firstly to identify the interaction of the bell-tower with the main body of the palace and so to distinguish between local, global and mixed modes; secondly, to assess the degree of rigidity of the roof level of the palace and the possible appearance of out-of-plane bending modes in the main façades of the palace. Recordings from channels A7, A8 and A9 are omitted in the dynamic identification, both in the AVT and the continuous monitoring, and reserved for monitoring ground motions. The reason is that the excitation level in the first floor of the palace under normal conditions is extremely low, so these accelerometers simply record noise when no seismic actions are present. Two asynchronous 30 minutes long acquisitions were carried out and ambient vibrations were recorded at a sampling rate of 10652.89 Hz (the maximum rate allowed by the DAQ). The test was conducted under normal operating conditions, with micro-tremors induced by traffic in the neighbouring roads and wind forces as the main sources of excitation. The mean environmental temperature during the test was 17.2° and the average wind speed was equal to 6.4 km/h as measured from the meteorological observatory of Gubbio centre, only 500 m from the palace. Such moderately strong wind speeds favoured the dynamic identification of the palace, reaching maximum accelerations of about 0.8 cm/s^2, while average ambient vibrations during the continuous monitoring are typically around 0.2 cm/s^2. The ambient vibration recordings were processed in the in-house software code MOVA [60], a companion software of MOSS dedicated to AVT. The acceleration time series were pre-processed including: (i) removal of non-stationary excitations produced by swinging bells (with a frequency of 15-minutes all day and night long) and anomalous spikes through Hanning window filtering, (ii) elimination of spurious trends through moving average baseline correction; (iii) fourth order band-pass Butterworth filtering with cut-off frequencies of 0.5 Hz and 100 Hz; and (iv) decimation of the data to 200 Hz.

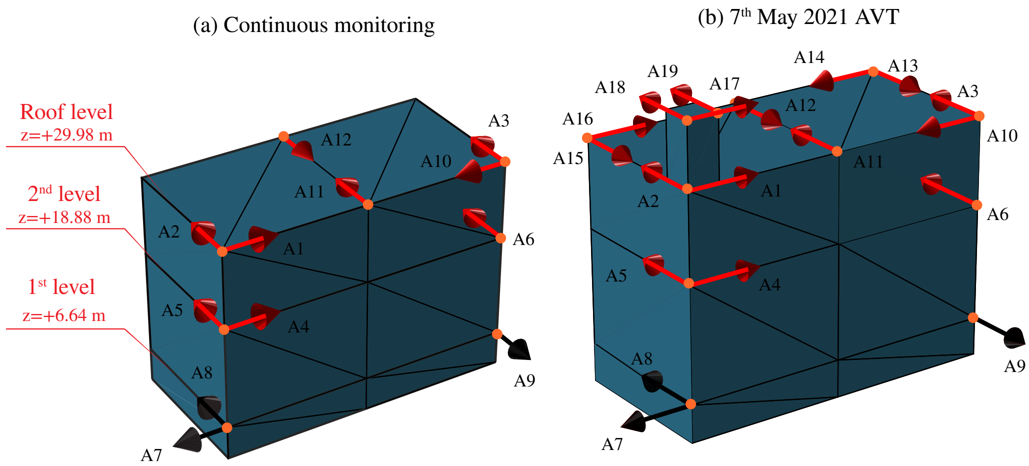

Figure 6: Comparison of the accelerometers layout during the continuous monitoring of the Consoli Palace (a) and the dense AVT performed on May 7th 2021 (b).

Figure 7 (a) furnishes one of the stabilization diagrams obtained using Covariance-based Stochastic Subspace Identification (COV-SSI) of the ambient accelerations recorded during the AVT considering a time lag of 6 s. The modal identification was performed using the automated procedure proposed in reference [64] for Data-based SSI and extended for COV-SSI in reference [60]. In general terms, the procedure started by defining a set of increasing time lags t_{lag}, or alternatively the number of block rows/columns j_b in the Toeplitz matrix of the output correlation matrix (i.e. $t_{lag} = (2j_b - 1)\,\Delta t$ with Δt being the time step of the acceleration series). In particular, we defined block rows/columns numbers j_b ranging from 301 ($t_{lag} = 6s$) to 401 ($t_{lag} = 8s$) with steps $\Delta j_b = 5$.

Then, for every value of j_b, the modal identification was performed considering model orders ranging from 40 to 80 with steps of 2. Afterwards, all the poles were collected and a first cleansing procedure is applied consisting of eliminating complex conjugate poles and poles with damping ratios above 10%. Finally, structural poles were distinguished from spurious ones by applying the hierarchical clustering approach reported in [64]. Threshold parameters to identify clusters of poles included relative variations of resonant frequencies $\Delta f < 1\%$, damping ratios $\Delta\zeta < 3\%$, and Modal Assurance Criterion (MAC) values $MAC > 0.99$. This approach allows the automated interpretation of the stabilization diagrams in the subsequent continuous OMA. Specifically, nine modes have been identified in the frequency range up to 10 Hz and highlighted with thick dashed lines in Fig. 7 (a), and the corresponding MAC matrix plot is furnished in Fig. 7 (b). The identified modal signatures (resonant frequencies, damping ratios, and Mode Phase Collinearities (MPC)) are collected in Table 1 and the first seven mode shapes are shown in Fig. 8. The modes have been classified as global (G), local (L), or high order models (HO) according to the interpretation of the mode shapes shown in Fig. 8. Specifically, four global modes have been identified and labelled with G-By1, G-T1, G-Bx1, and G-By2 in Fig. 8 and Table 1. Modes G-By1 and G-Bx1 correspond to first order bending modes along the *y*- and the *x*-directions of the building (refer to Fig. 5 (b)), respectively, mode G-By2 refers to a second-order bending mode along the *y*-direction, and mode G-T1 corresponds to the global torsional mode of the palace. Modes L-Bx1 and L-By1 refer to the first order bending modes of the bell-tower along the *x*- and the *y*- directions, while mode L-T1 corresponds to the first torsional mode of the bell-tower. Note that mode L-By1 involves certain torsion in the main body of the palace (see Fig. 8). Finally, modes HO1 and HO2 show complex interactions between the main body of the palace, including some out-of-plane deformation of the roof level. Specifically, modes HO1 and HO2 respectively show symmetric and anti-symmetric movements of the point locations of channels A11 and A12. Further analyses to correctly interpret these modes are left for future work, possibly with the aid of a numerical model, and, therefore, they have been omitted in Fig. 8. Nevertheless, for completeness, the mode shapes of these modes have been included as inserts in Fig. 9 reporting the results of the tracking of the resonant frequencies of the Palace through continuous monitoring. These modes identified by the automated OMA procedure clearly correspond to columns of stable poles in the stabilization diagram in Fig. 7 (a) as well as the peaks of the singular values of the spectral matrix, except for the column of poles at about 3.26 Hz. This mode has been omitted from the identification because its mode shape is almost identical to that of the mode at 3.54 Hz with a MAC value of 0.96. Therefore, we decided to omit this mode because of its slightly larger complexity and its poorer correlation with the identified poles during the continuous monitoring. As previously observed in the work by Kita *et al.* [59], this column of stable poles may indicate a possible splitting of mode L-T1, although more specific analyses should be addressed in this regard to confirm it. The MAC matrix plot in Fig. 7 (b) shows that most of the identified modes are highly independent, with MAC values close to zero in most of the off-diagonal terms. Only some correlation is observed between modes L-T1 and G-T1 with a MAC value of 0.78 but attributable to the common torsion of the main body of the palace, and modes G-By2 and HO2 with a MAC value of 0.62 also due to the common bending motion of the palace. Finally, let us remark that all the identified modes during the AVT are eminently real, with MPC values above 96%.

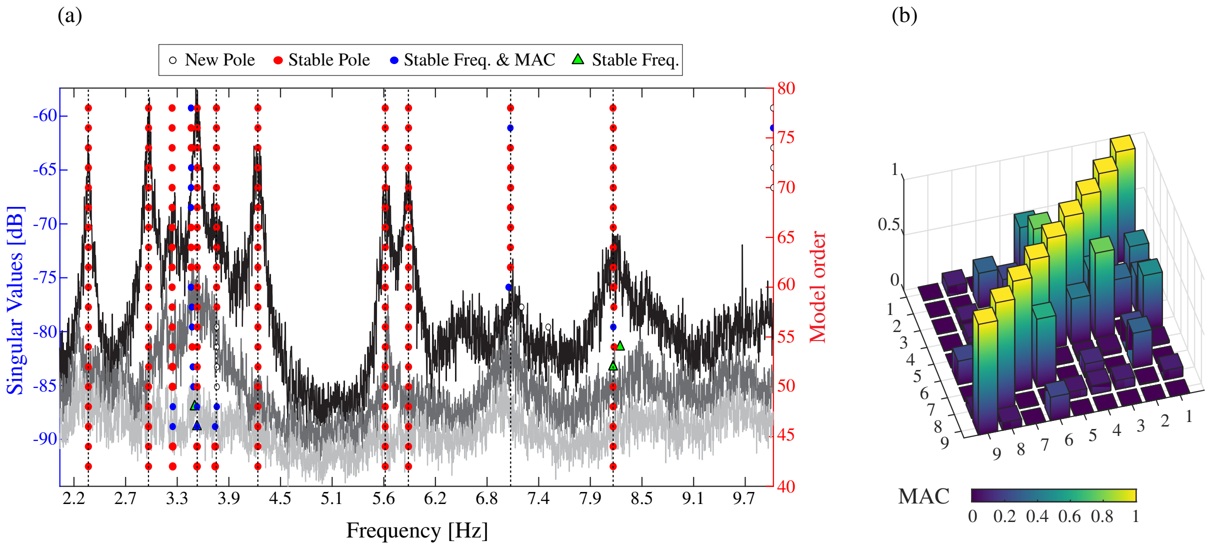

Figure 7: Stabilization diagram obtained using COV-SSI of the Consoli Palace during the AVT conducted on May 7th 2021 (j = 1200, t_{lag} = 6 s) (a), and MAC matrix plot (b).

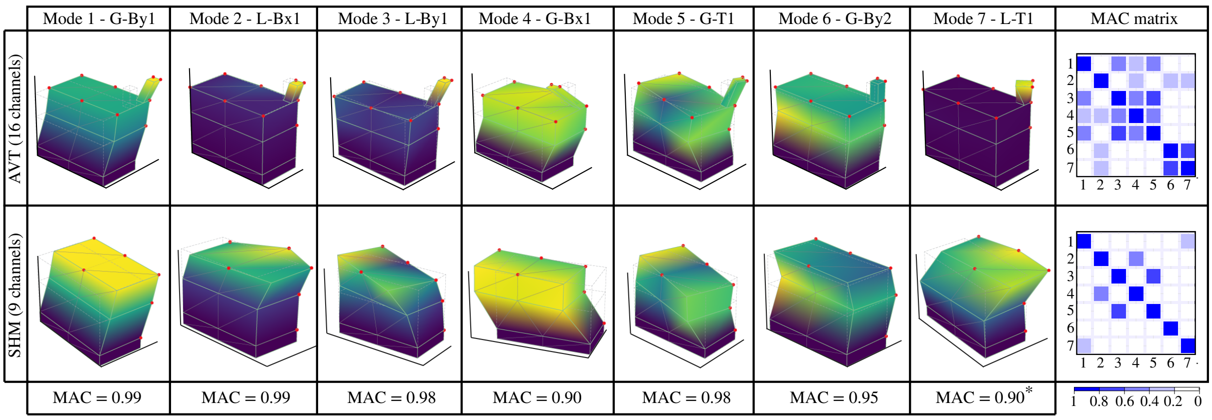

Figure 8: First seven mode shapes identified in the dense AVT and consistently tracked in the continuous monitoring of the Consoli Palace. The mode shapes representative of the continuous monitoring were extracted from the identification of the accelerations recorded on July 17th 2020 12:30 pm CEST. *In the comparison of Mode L-T1 between the AVT and the SHM campaign, only the modal displacements at the roof level were considered in the computation of the MAC value since, given the local nature of this mode, only marginal values were obtained at the first and second floors in the AVT.

Table 1: Comparison of the experimentally identified modal signatures of the Consoli Palace in the dense AVT and continuous monitoring using the COV-SSI method.

		AVT - 13:00 pm CEST May 7th 2021			Continuous monitoring - July 2020/May 2021		
Mode No.	Label	Frequency [Hz]	Damping ratio [%]	MPC*1 [%]	SR*2 [%]	Mean Frequency [Hz]	Variation range [Hz]
1	G-By1	2.32	0.98	98.3	85.5	2.32	2.18 (-5.80%) - 2.44 (+5.10%)
2	L-Bx1	2.99	0.92	100.0	36.25	3.02	2.75 (-8.65%) - 3.45 (+14.29%)
3	L-By1	3.54	0.78	99.9	49.57	3.53	3.32 (-8.54%) - 4.02 (+14.10%)
4	G-Bx1	3.75	2.76	99.0	56.08	3.75	3.51 (-6.36 %) - 3.97 (+5.77%)
5	G-T1	4.22	0.95	99.9	70.94	4.2	3.86 (-8.28%) - 4.51 (+7.34 %)
6	G-By2	5.65	0.72	99.8	57.3	5.53	5.10 (-7.78%) - 5.97 (+8.02 %)
7	L-T1	5.91	0.69	99.8	76.2	6.46	5.96 (-7.75 %) - 7.00 (+8.27%)
8	HO1	7.05	1.65	97.2	43.61	7.05	6.77 (-4.07%) - 7.41 (+5.00%)
9	HO2	8.20	1.67	96.3	76.61	7.97	7.23 (-9.29%) - 9.23 (15.89%)

*1 *Mode Phase Collinearity*
*2 *Success ratio in the identification*

The previous results were used to define the baseline modal features of the palace to be tracked during the continuous monitoring. To do so, a standard frequency tracking approach was implemented to trace the time series of the modal features of the palace. This approach consists of grouping the modal poles identified during the continuous monitoring by exploiting their similarities with the reference baseline features. Specifically, the implemented approach is semi-dynamic. This implies that the reference mode shapes (those identified in the AVT) are kept constant throughout all the monitoring period, while the reference resonant frequencies and damping ratios vary in time. The comparison between poles is performed in terms of relative variations in the resonant frequencies Δf and MAC values. In every step in the tracking procedure, all the poles complying with pre-defined thresholds are sorted according to a metric distance involving both Δf and MAC values MAC as d:

$$d = (1 - \eta)\,\Delta f + \eta\,(1 - MAC)\,, \tag{18}$$

with η being a weighing factor between the contributions of Δf and MAC. Once sorted, the pole with the lowest distance d is collected in the corresponding time series of the mode. On this basis, the same nine modes previously identified in the AVT have been tracked throughout all the monitoring period as shown in Fig. 9. To do so, the thresholds of maximum relative variations in the resonant frequencies have been selected after some manual tuning as 8% for modes 1 to 3, 10% for mode 4, 15% for mode 5, and 5% for modes 6 to 9. Minimum MAC values of 0.75 have been defined for modes 1 and 4, 0.8 for modes 3, 6 and 8, and 0.9 for modes 7 and 9. The weighing factor η has been selected as 0.5. Additionally, all the poles with MPC values below 80% are disregarded as complex or insufficiently excited modes. It is evident in Fig. 9 that all the resonant frequencies exhibit both daily and seasonal fluctuations, more exacerbated as the modes have a more local character with higher frequencies. The statistical properties of the tracked modes are collected in Table 1. In this table, the success ratios (SR) have been also included, that is, the percentage of times the modes have been identified during the monitoring period. In general, it is noted that local modes are poorly tracked with SRs of 36.25% and 49.57% for modes L-Bx1 and L-T1, respectively, which is expectable because no sensors are located in the bell-tower during the SHM campaign.

Surprisingly, this is not the case of mode L-T1 which is consistently tracked with a SR of 76.2%. Note in Fig. 8 that this mode combines certain bending movements concentrated at the roof floor of the palace, which explains the success in its identification given the considerable concentration of accelerometers in that level. Global modes are tracked with SRs between 56 and 85.5%, which agrees with our previous experience of dynamic SHM of stiff masonry structures. Interruptions in the frequency tracking intensify during night-time hours, when the palace remains closed to the public and the surrounding vehicle traffic reaches minimum levels. It is also noticeable in Table 1 the large environmental effects exhibited by most modes, reaching in most of them variations around 10% their average values. Such strong effects justify the need for implementing an effective statistical pattern recognition for performing damage detection. To this aim, the training period has been defined from July 14$^{\text{th}}$ 2020 until May 3$^{\text{rd}}$ 2021 (≈9 months), followed by the damage assessment period until July 12$^{\text{th}}$ 2021.

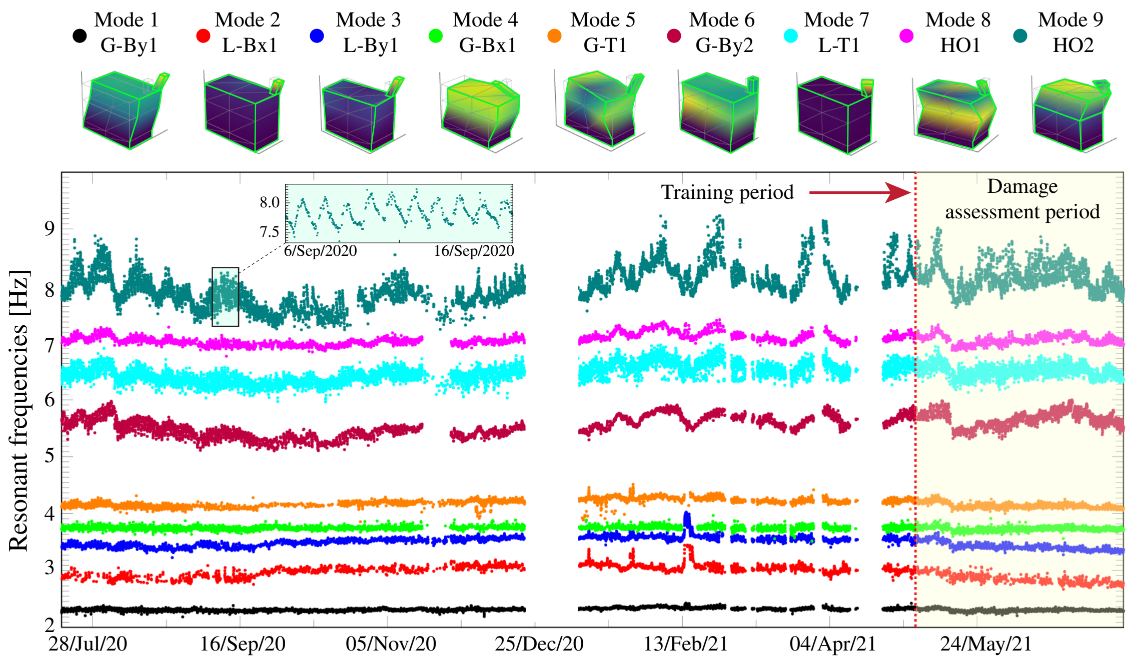

Figure 9: Tracking of the resonant frequencies of the Consoli Palace from July 2020 until July 2021.

The time series of the crack displacements recorded by LVDTs D1, D2, and the temperature readings by thermo-couples T1, T2, and T3 are shown in Figs. 10 (a) and (b), respectively, and some statistics are presented in Table 2. The monitoring data recorded by temperature sensors T4 and T5, and LVDTs D3 and D4 are omitted in this work due to difficulties in the data transmission, which made impossible to obtain consistent readings throughout the monitoring period. The recordings by LVDTs D1 and D2 exhibit similar behaviours, with amplitudes ranging between a closing of 0.112 mm to an opening of 0.25 mm with respect to the initial state of the monitored cracks. The analysis also evidences the strong effect of environmental temperature in Fig. 10 (b) upon the crack displacements in Fig. 10 (a), exhibiting both seasonal and daily fluctuations. The monitored cracks tend to open during the winter, while closing during the summer. Also the breathing behaviours of the cracks can be observed in the zoom inserts in Fig. 10 (a), with closing during the day-time and opening during the night-time. The monitored temperatures range from 0 to 35 °C. Note in Table 2 that sensor T3, which is almost located outdoor, shows significantly larger daily fluctuations compared to indoor sensors T1 and T2.

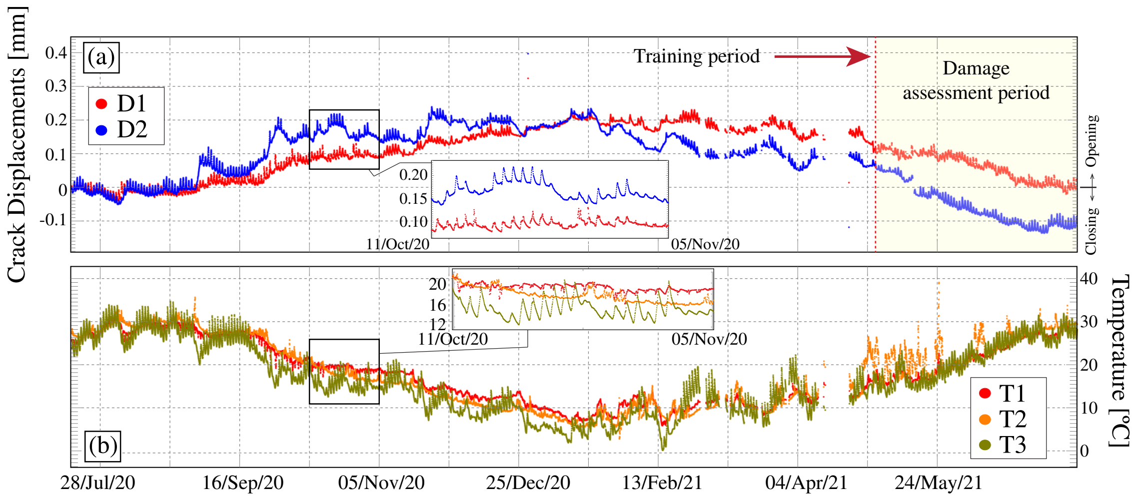

Figure 10: Time series of crack amplitudes of channels D1 and D2 (a), and time series of environmental temperatures of channels T1, T2, and T3 (b) from July 2020 to July 2021.

Table 2: Statistics of measured crack displacements (D1 and D2) and temperature data (T1, T2 and T3) from July 2020 to July 2021.

Var.	Mean Val. [mm]	Min Val. [mm]	Max Val. [mm]	σ [mm]
D1	1.20E-01	-0.035	0.24	7.53E-02
D2	1.70E-01	-0.112	0.25	7.19E-02
Var.	Mean Val. [°C]	Min Val. [°C]	Max Val. [°C]	σ [°C]
T1	17.3	5.78	31.3	7.32
T2	17.4	2.69	35.5	8.07
T3	16	-0.06	34.6	8

3.3. Statistical Pattern Recognition results

With the aim of addressing the elimination of EOCs from the resonant frequencies and the crack amplitudes, some preliminary correlation analyses have been conducted as reported in Figs. 11 and 12. In the subsequent results, only the resonant frequencies of modes with SRs above 50% are considered, namely G-By1 (Mode 1), L-T1 (Mode 3), G-Bx1 (Mode 5), G-T1 (Mode 5), L-T1 (Mode 7), and HO2 (Mode 9). In addition, because of space constraints, only the correlations with the temperature sensors yielding maximum coefficients of determination R^2 are presented herein.

Regarding the analysis of the resonant frequencies in Fig. 11, it is observed that negative frequency/temperature correlations are found in all the modes. This indicates that the global stiffness of the palace decreases as temperature rises. As previously discussed in the introduction, such a trend is quite uncommon in masonry structures such as slender towers or churches (see e.g. [25–27]). Most authors agree to hypothesize positive correlations to be driven by thermal-induced closure of micro-cracks in the mortar joints. In the case of the Consoli Palace, negative frequency/temperature correlations were previously found by Kita *et al.* [59]. Those authors attributed this trend to temperature-induced slackening of some metallic tie rods installed in the Arengo hall to restrain the lateral thrusts exerted by the barrel-vault ceiling, as well as the possible contribution of the existing macro cracks. Overall, considerably linear frequency-temperature correlations are observed in the natural frequencies of modes 3 and 4, whereas noticeably non-linear correlations are identified instead for modes 1, 5, 7 and 9, which may be ascribed to the existence of complex temperature-driven mechanisms or a strong dependence on unmonitored EOC. The large scatter in the correlation plots of modes G-By1, G-x1, and L-T1 may indicate the presence of thermal capacitance effects, that is some delays in heat transfer from the position of the thermocouples through the cross-sections of the masonry walls. The most significant correlations with temperature are found for modes L-By1 and G-T1 with coefficients of determination R^2 of 0.85 and 0.72, respectively. This strong correlation for global torsional mode G-T1 is expectable since it is dominated by the shear stiffness of the external walls of the palace with direct exposure to the outdoor environment. Considering the global bending modes of vibration G-By1 and G-Bx1, the higher degree of temperature correlation is observed for mode G-By1. This can be ascribed to the large exposed surface of the two longitudinal walls along the north-south façades of the palace, which contribute to the largest extent to mode G-By1. Conversely, two out of the four transversal walls along the East-West direction of the palace and activated by mode G-Bx1 are located indoor and, thus, are less affected by the external environment. Regarding the temperature correlations observed for local modes L-By1 and L-T1, it is noted that the local bending mode is highly affected by thermal variations (R^2=0.85), while the local torsional mode only exhibits a moderate correlation (R^2=0.32). The thermal sensitivity of the former is attributable to the circumstance that the bell-tower is directly exposed to the outdoor environment. In this case, the flexural restraint imposed by the roof floor is expected not to be sensibly affected by the environmental temperature. Conversely, the restraint of the roof floor to torsional rotations at the base of the bell-tower might be considerably affected by tower/floor differential thermal expansions. Indeed, this may explain the large scatter observed for this mode, which may be indicative of important thermal capacitance effects. With regard to the correlations between crack amplitudes D1 and D2 and temperature data, it is noted in Fig. 12 that crack amplitudes also exhibit negative correlations with the environmental temperature. Such a trend indicates that cracks tend close as the environmental temperature rises with the subsequent expansion of the masonry volumes, and vice versa. Almost perfect correlation is found for LVDT D1 (located in the South façade of the palace) with a coefficient of determination very close to 1. Conversely, a more complicated correlation is found for LVDT D2 with substantial scatter around the regression lines. This fact again evidences the potential existence of thermal capacitance effects.

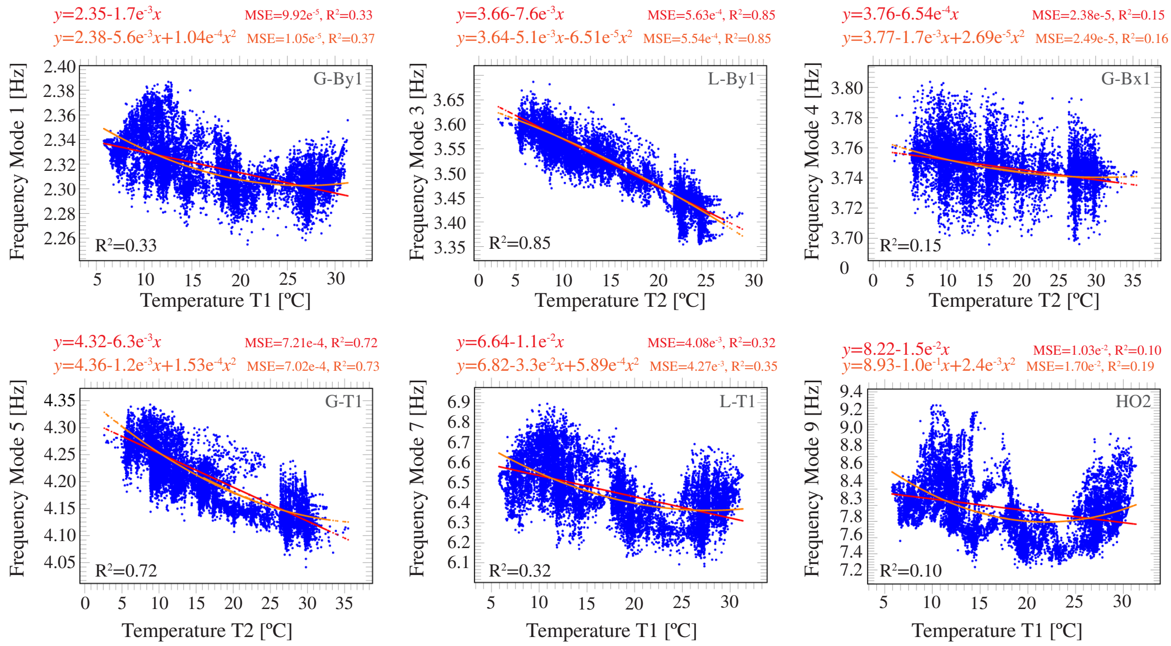

Figure 11: Correlations between the natural frequencies of the Consoli Palace and the environmental temperature in the training period from July 14th 2020 until May 3rd 2021 (12878 data samples).

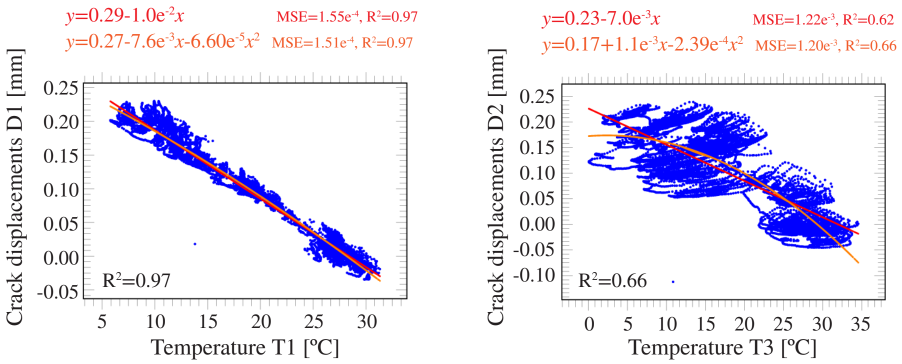

Figure 12: Correlation between the crack displacements monitored by LVDTs D1 and D2 and the environmental temperature in the training period from July 14th 2020 until May 3rd 2021 (12878 data samples).

In the subsequent analyses, an autoregressive (AR) time series model has been implemented to complete the time series of resonant frequencies. In general, an AR model conceives an arbitrary observation sequence $x[n]$ at instant n as a linear combination of p (model order) past observations:

$$x[n] = a_0 + \sum_{k=1}^{p} a_k x[n-k] + \varepsilon[n], \tag{19}$$

where a_k are prediction coefficients, a_0 is a constant value, and $\varepsilon[n]$ is a white noise process. The implemented algorithm (*fillgaps.m* Matlab function) determines local forward and reverse AR models using signal segments of certain length l around the missing data, and estimates the local prediction coefficients a_k using the Burg's method (for further details, readers may refer to reference [65]). To the purpose of this work, segments of 144 data points (corresponding to 3 days of monitoring data) and a model's order of $p = 3$ have been found suitable and selected henceforth. Based upon the reconstructed time series, the MCD-based outliers detection algorithm previously introduced in Section 2.2 has been applied to the time series of resonant frequencies in the training period as shown in Fig. 13. In particular, the dimension of the subsets has been selected according to the recommendation by Rousseau and Driessen [50] as $h = (n + p + 1)/2 = (12878 + 6 + 1)/2 \approx 6443$. Once the optimal h-subset is found, all the data samples are sorted according to their Mahalanobis distances from the optimal set. Afterwards, 20% of the data points with largest distances are considered as outliers and disregarded in the subsequent data normalisation. The MCD approach was not applied to the time series of monitoring data from the LVDTs because no significant outliers were observed, and only a few abnormal data points were manually eliminated. Note that no outliers elimination was conducted beyond the training period leaving the time series intact. This is crucial to prevent the erroneous elimination of nonconformities that may stem from any structural pathology.

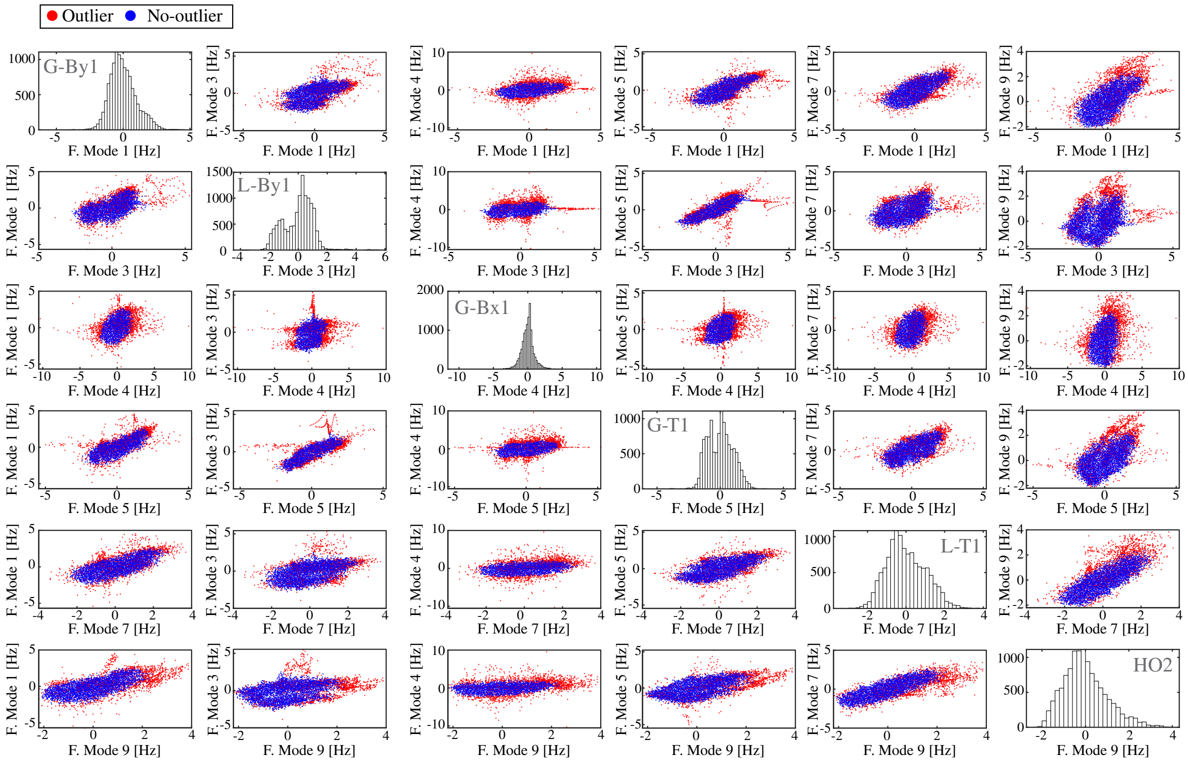

Figure 13: Correlation analysis between the resonant frequencies of the Consoli Palace (Modes G-By1/Mode 1, L-T1/Mode 3, G-Bx1/Mode 4, G-T1/Mode 5, L-T1/Mode 7, and HO2/Mode 9) and outliers detection results. Twenty percent of the data points with largest Mahalanobis distances to optimal h-subset selected by MCD are considered as outliers.

The intricate correlations between the dynamic/static features and the environmental conditions reported in Figs. 11 and 12 justify the implementation of the SMLR model previously presented in Section 2.3. Specifically, a total of 48 potential predictors are considered. These comprise linear and quadratic (denoted with the subscript 2) versions of the time series of environmental temperatures by thermo-couples T1, T2 and T3. Additionally, with the aim of accommodating potential thermal capacitance effects in the palace, time delayed versions of the previous time series are also accounted for. These include delays of 30 min (2 samples), 1 hour (4 samples), 2 hours (8 samples), 5 hours (20 samples), 12 hours (48 samples), 24 hours (96 samples), and 48 hours (192 samples). On this basis, Figs. 14 (a,b) and (c,d) show the coefficients of the LAR regressions obtained through the analysis of the cleansed training population of the resonant frequencies of the resonant frequency of Mode 1 (G-By1) and the crack amplitudes of LVDT D1, respectively. It is noted in Fig. 14 how the number of predictors with non-zero regression coefficients β_i increases as the LAR algorithm progresses. It is interesting to note in Fig. 14 how the proposed method is capable of automatically finding correlations with delayed predictors, making the statistical model dynamic in nature. In order to select the optimal model and, therefore, the optimal set of predictors, several metrics have been implemented, namely RSS, BIC, and AIC. In general, the BIC and AIC criteria yielded similar solutions in all the considered estimators, while the RSS led to less sparse solutions. This is the case of the LAR analysis of Mode G-By1 and LVDT D1 as shown in Figs. 14 (b) and (d). For the sake of minimizing overfitting limitations, the solutions obtained by minimizing the BIC criterion have been retained. The same procedure is applied to all the considered estimators, and the obtained results are summarized in Table 3, including the model sparsities, fitting mean squared errors (MSEs), and coefficients of determination R^2. In this table, the degree of sparsity s is indicated as the percentage ratio of regression coefficients shrunk to zero and the total number of potential predictors (i.e. 48).

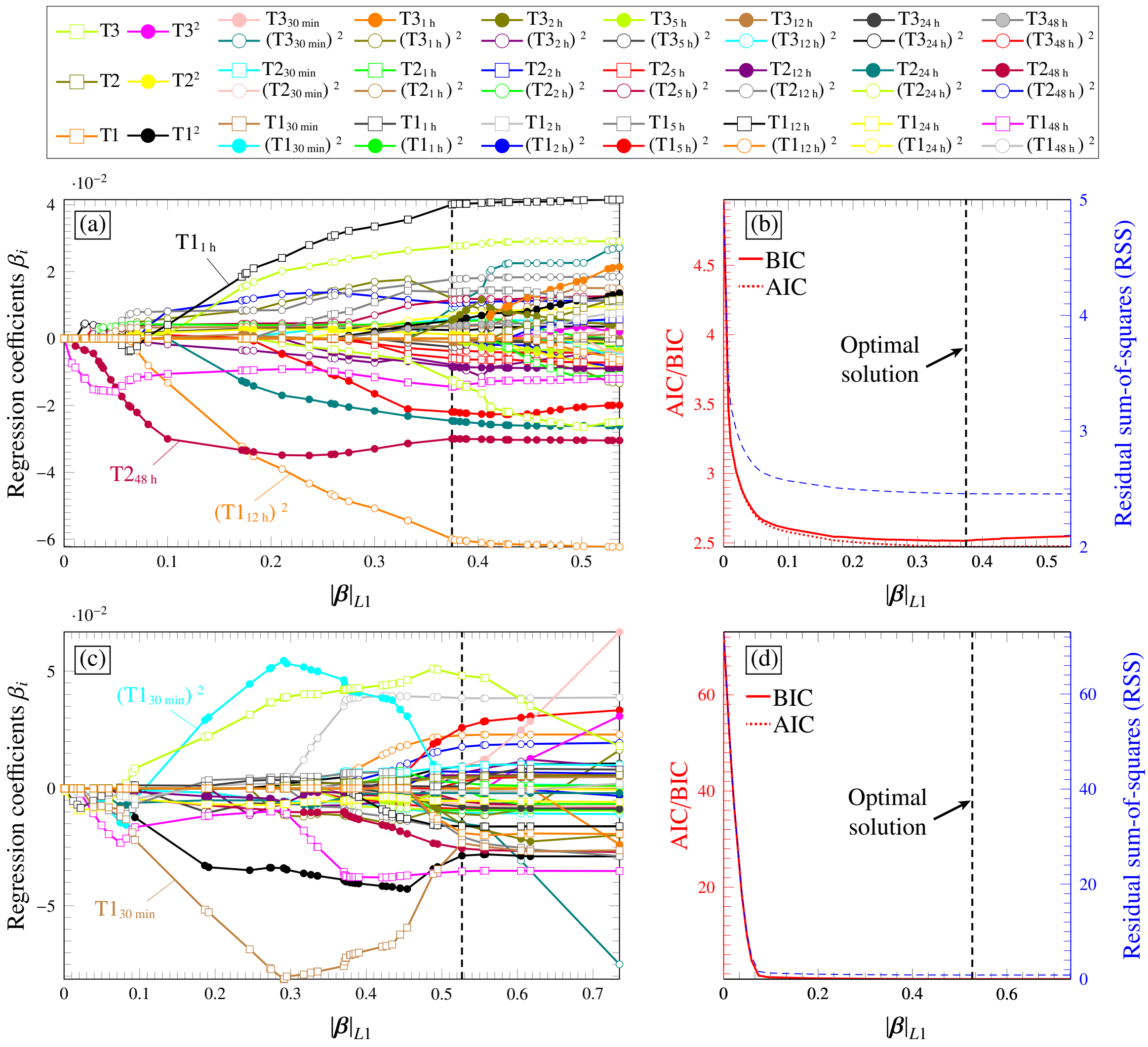

Figure 14: LAR analysis of the experimentally identified resonant frequencies of Mode G-By1 (a,b) and crack amplitudes D1 (c,d). Evolution of the regression coefficients β_i versus the L1 norm ($|\boldsymbol{\beta}|_{L1} = \sum_{i=1}^{p} |\beta_i|$) (a,c), and selection of optimal set of predictors based upon the residual sum-of-squares (RSS), and Bayesian Information Criterion (BIC) and Akaike Information Criterion (AIC).

Table 3: Degree of sparsity and fitting results of the SMLR analysis of the resonant frequencies and crack amplitudes of the Consoli Palace.

Estimator	Sparsity (s)	MSE	R^2	K-S p-value*
Mode 1 (G-By1)	32.65	1.05E-4	0.48	0.127
Mode 3 (L-T1)	38.78	4.17E-4	0.9	0.571
Mode 4 (G-Bx1)	22.45	2.13E-5	0.18	0.004
Mode 5 (G-T1)	36.73	4.95E-4	0.82	0.013
Mode 7 (L-T1)	28.57	3.89E-3	0.61	0.316
Mode 9 (HO2)	14.29	1.63E-2	0.67	0.060
Crack-meter D1	10.00	7.70E-5	0.99	0.014
Crack-meter D2	20.00	6.13E-4	0.87	0.125

**Kolmogorov–Smirnov (K-S) normality test*

The quality of the pattern recognition has been assessed by the inspection of the statistical distribution of the residuals. Ideally, the residuals in the training period should only contain normally distributed errors stemming from limitations in the identification of the healthy database of the monitored structure and marginal EOC effects.

For comparison purposes, the results obtained by SMLR are benchmarked against those obtained by using standard PCA (refer to e.g. [28] for further details on its implementation). In the latter, three principal components (PCs) have been kept explaining more than 90% of the variance in the selected resonant frequencies. Figure 15 shows an example of the analysed residuals corresponding to the fundamental frequency of the palace. In this figure, it is noted that very similar results are obtained using PCA and SMLR in statistical terms. Interestingly, the results obtained by PCA seem to slightly outperform those achieved by SMLR. In particular, both the mean (μ) and the standard deviation (σ) of the residuals obtained by PCA are lower to those obtained by SMLR. Additionally, the statistical moments of the residuals by PCA indicate more consistency with Gaussianity. Note that the kurtosis (κ) is 3.05 and 3.02 for the SMLR and PCA models, respectively, with $\kappa = 3$ being the theoretical value for a perfect Gaussian distribution. Similarly, the skewness (γ) takes values of 3.5E-1 and -2.0E-2 for the SMLR and PCA models, respectively, being $\gamma = 0$ the theoretical value for an ideal Gaussian distribution. The Kolmogorov–Smirnov (K-S) normality test results of the residuals of the considered DSFs obtained by the SMLR model are reported in Table 3. It is noted that not all the residuals can be considered perfectly normally distributed. In particular, considering a standard significance level of 0.05, the residuals of the resonant frequencies of Modes 4 and 5 and crack amplitudes D1 would fail to pass the normality test. This circumstance, which is common in practice, inevitably represents a certain bias in the determination of the T^2 values (Eq. (2)) and the UCL (Eq. (3)). In this regard, no substantial differences were found when implementing PCA, achieving the same number of residuals strictly passing the K-S normality test.

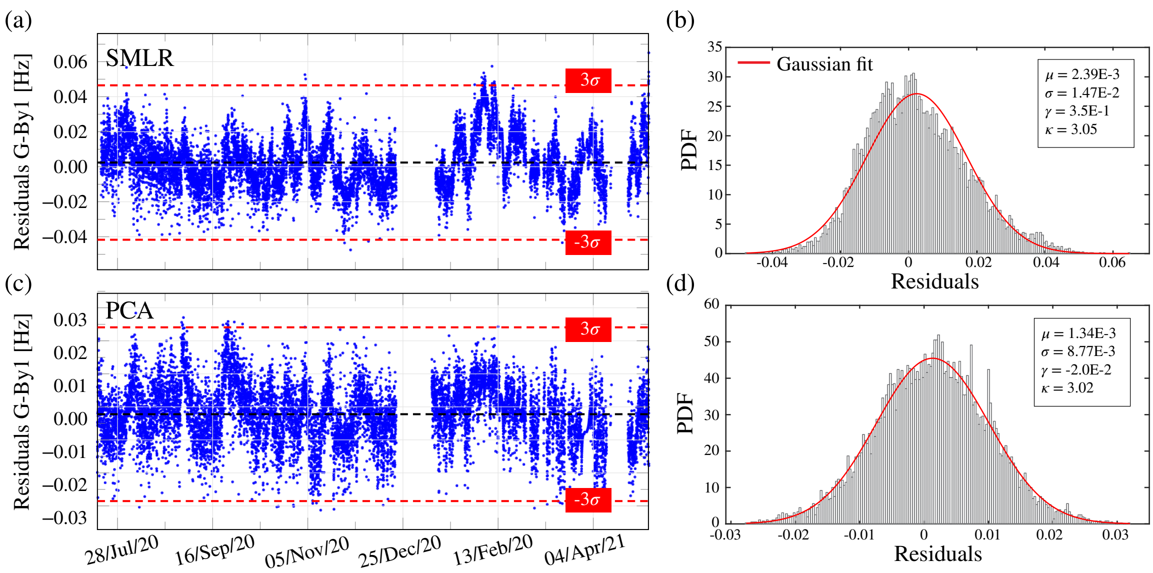

Figure 15: Analysis of residuals in the predictions of the fundamental frequency of the Consoli Palace using SMLR (a,b) and PCA (3 PCs) (c,d) obtained through the training period from July 14th 2020 until May 3rd 2021 (12878 data samples).

The previous closer fittings by PCA are expectable given its fundamental definition. Consider that PCA is a dimensionality reduction approach based upon the eigenvalue decomposition of the covariance of the observation matrix. In this light, EOC-induced variability is assumed to be contained within the eigenvectors associated with the largest eigenvalues (PCs), i.e. the PCs contributing the most to the variance. Indeed, perfect reconstruction is achieved when considering a number of PCs equal to the number of estimators. Nevertheless, since PCA does not rely on any EOC independent from the structural damage, certain defects may go unnoticed in the damage assessment period. This would be the case, for instance, of a consistent damage-induced shift in the resonant frequencies with minimal effect upon their correlations. This limitation is evidenced in Figs. 16 (a) and (b) where the experimentally identified resonant frequencies of the Consoli Palace throughout all the monitoring period along with the reconstructions obtained by SMLR and PCA are presented, respectively. After the seismic sequence initiated on May 15th 2021 and analysed in more details in the next section, some drops in most of the resonant frequencies are evident in the zoom views in Fig. 16. As further analysed below, the earthquake-induced damage did not significantly affect the correlations between the resonant frequencies of the palace. This fact makes the predictions by PCA replicate the drops found in the experimental resonant frequencies, limiting its effectiveness for damage detection. Conversely, the SMLR model do not predict any drop in the resonant frequencies, which will facilitate the identification of damage-induced anomalies in the residuals. This is particularly evident in mode L-By1, where the experimental data follow a shifted parallel tendency to the statistical predictions.

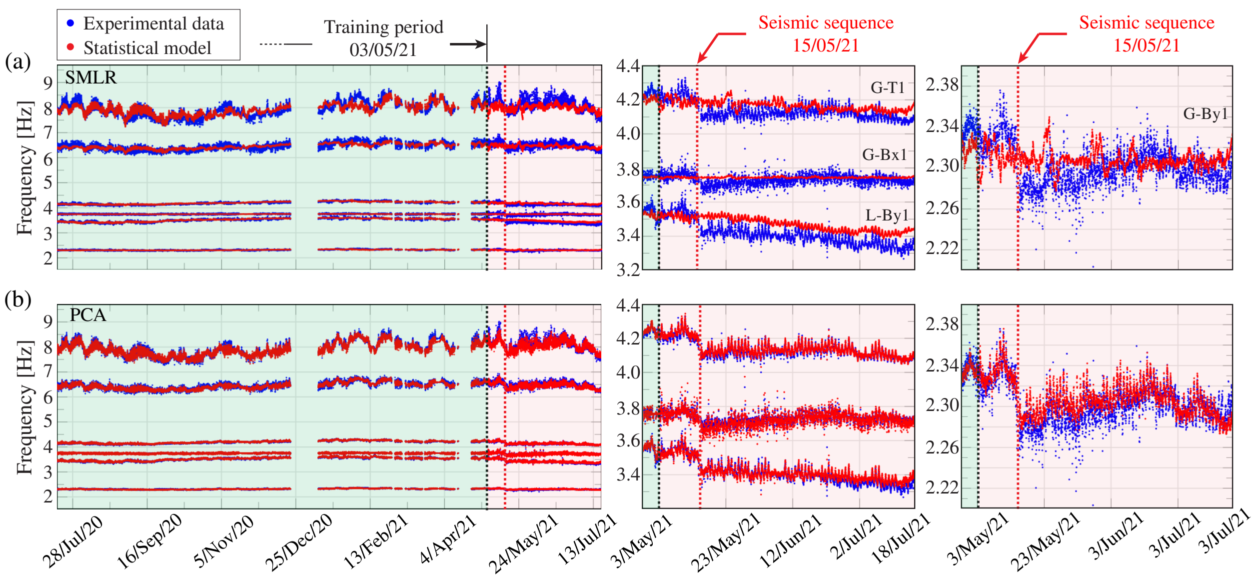

Figure 16: Prediction of the resonant frequencies of the Consoli Palace using SMLR (a) and PCA (3 PCs) (b) using a training period from July 14th 2020 until May 3rd 2021 (12878 data samples) followed by a damage assessment period until July 18th 2021 (3631 data samples).

3.4. *May 15th 2021 seismic sequence - Damage identification results*

As anticipated above, a relatively important seismic sequence initiated on May 15th 2021 with epicentre in Gubbio. The seismic sequence comprised six earthquakes of moderate intensity, with the strongest shock of magnitude Mw 4.0 at 07:56 UTC. Ground motion records for these earthquakes have been obtained from the data provided by the Italian Strong Motion Network (RAN) of the Department of Civil Protection (DPC) and the Italian Seismic Network (RSN) of the National Institute of Geophysics and Vulcanology (INGV). Specifically, seismic records have been taken from the Gubbio Parcheggio Santa Lucia station, which is located only 600 m far from the Consoli Palace. Figure 17 (a) shows the geographical location of the epicenter of the main shock and the location of the palace, and the waveforms recorded at the Gubbio Parcheggio Santa Lucia station are presented in Fig. 17 (b). Table 4 reports the registered seismic events, including their PGA, depth and distance from the seismic station. Note that the Consoli Palace is located at a distance of less than 3 km from the epicenters, thereby this case study represents a unique example of a monumental building subjected to impulsive near-field earthquakes. The location of the epicenter is almost identical to the seismic sequences started on December 18th 2013 with a major shock of similar intensity Mw 3.9, followed by seven aftershocks with intensities between Mw 2.9 and 3.6. Therefore, it is conceivable that this new sequence may have been originated by the same activation mechanism of the Gubbio fault.

To illustrate the transient response of the palace, Figs. 18 (a) and (b) show the acceleration time-histories and time-frequency analysis of the accelerations recorded by sensor A5 under the seismic events on May 15th 2021 at 08:07 UTC and 10:19 UTC, respectively. The time-frequency analysis is performed using the Wigner-Ville distribution evaluated in the frequency broadband from 0 to 10 Hz. In Fig. 18 (a), it can be observed that the fundamental frequency experiences a decrease down to 2.16 Hz, however during the coda it recovers to ≈2.20 Hz after 13 s. The recovery is not complete and a mild shift of frequency exists compared to the pre-event frequency of 2.26 Hz (indicated by a dashed line in Fig. 18 (a)). The largest decays in the resonant frequencies are though expected to have appeared during the main Mw 4.0 shock. Unfortunately, acceleration records during this event are not available because of an electrical interruption which affected the SHM system until 08:00 UTC. Nonetheless, herein we focus on the analysis of the pre- and post-earthquake behaviour of the palace, and the analysis of its transient response under base strong motions falls out the scope of this work.

Table 4: Seismic events registered in May 2021 at the Gubbio Parcheggio Santa Lucia station (Latitude: 43.3558, Longitude: 12.5717, Elevation: 515 m). Source: Italian Strong Motion Network (RAN).

Event	Date	Mw	PGA [cm/s^2]	Depth [km]	Dist. epic. [km]
E1	15/05/21 07:56:01 UTC	4.0	102.4	9.9	1.4
E2	15/05/21 08:07:20 UTC	3.1	35.3	9.6	1.0
E3	15/05/21 10:19:17 UTC	3.0	18.07	10.5	2.4
E4	15/05/21 21:27:25 UTC	2.8	18.07	9.4	2.2
E5	23/05/21 20:51:22 UTC	3.0	17.38	8.1	1.8
E6	27/06/21 13:27:16 UTC	3.0	7.65	6.8	2.2

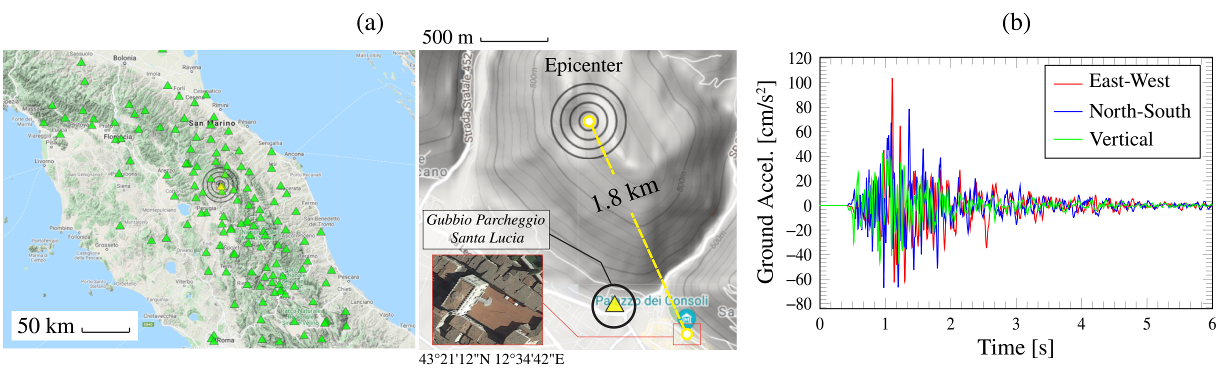

Figure 17: Geographical map highlighting the epicenter of the earthquake from May 15th 2021 at 07:56:01 UTC (a), and E-W, N-S and vertical components of the near-field accelerations recorded by the Gubbio Parcheggio Santa Lucia station (200 Hz sampling frequency) (b).

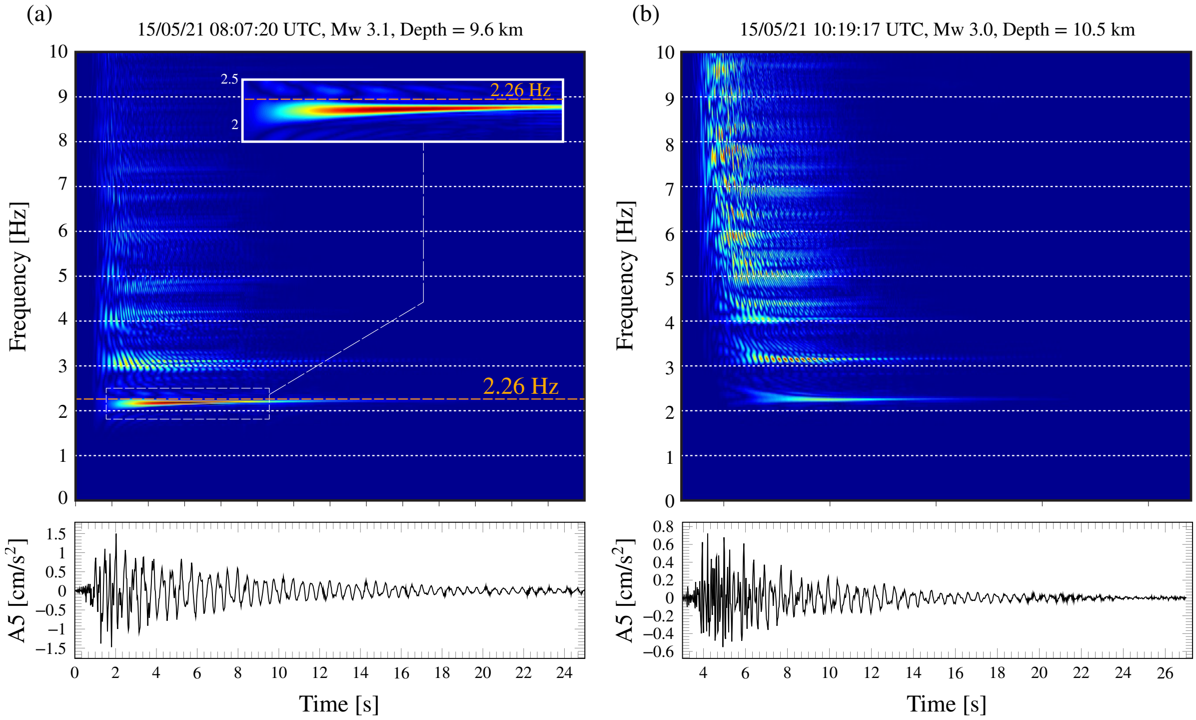

Figure 18: Time series and time-frequency analysis (Wigner-Ville distribution) of the acceleration records by channel A5 during the seismic events on May 15$^{\text{th}}$ 2021 at 08:07 UTC (a) and 10:19 UTC (b).

With the aim of assessing the potential appearance of damage in the Consoli Palace after the seismic sequence, novelty analyses have been conducted on the basis of the theoretical framework previously overviewed in Section 3.3. The Hotelling's T^2 control charts in terms of resonant frequencies and crack amplitudes are furnished Figs. 19 and 21, respectively. Let us firstly focus on the novelty analysis of the resonant frequencies of the palace in Fig. 19. For comparison purposes, the control chart obtained using SMLR is benchmarked again against the one resulting from using PCA (3 PCs). In this figure, a marked anomaly is clearly observable after May 15$^{\text{th}}$, either using SMLR or PCA. In both cases, the T^2 distances experience a shift right after the onset of the seismic sequence, although some slightly better results are visible when using SMLR. Specifically, in the case of SMLR, the number of out-of-control processes overpassing the 95% UCL amounts to 8.55% until May 15$^{\text{th}}$ 2021, and increases up to 71.95% in the remaining damage assessment period. Instead, in the case of PCA, the number of outliers amount to 7.85% and 68.8% before and after the seismic sequence for the same confidence level. The quality of the damage classifications by SMLR and PCA is also appraised in Fig. 19 (b) through the assessment of the confusion matrices, including receiver operating characteristic (ROC) and Precision/Recall (PR) curves. For their calculation, a dense range of UCL values is swept and the frequency of outliers is computed and stored independently before and after the seismic sequence. Then, outliers before May 15$^{\text{th}}$ are assumed as false positives, while those arising after May 15$^{\text{th}}$ are considered true positives. Note that, in this particular case study, precision-recall curves may be more informative than ROC curves since the size of the dataset after the seismic sequence (3631 data samples) is considerably smaller than in-control set (12878 data samples), being the damage/undamaged classes considerably imbalanced. The analyses are also performed considering PCA and SMLR without outliers elimination to demonstrate the importance of cleansing the training dataset. Finally, with the aim of providing further insight into the effectiveness of the classifications, the areas under the ROC (AUC) and Precision-Recall (PAUC) have been also computed as comprehensive quality metrics. All things considered, it is evident in Fig. 19 (b) that the pattern recognition including data cleansing (denoted with solid lines) proves far better performance using both SMLR or PCA. Indeed, the classification conducted without outliers elimination achieves limited areas of AUC=0.8698 and PAUC=0.5463 in the case of SMLR, and AUC=0.8811 and PAUC=0.5002 in the case of PCA. Instead, when the cleansing of the training population is included, the classifications using both SMLR (AUC=0.9749, PAUC=0.8766) or PCA (AUC=0.9739, PAUC=0.8834) approach the perfect classifier (AUC=PAUC=1.0). With regard to the comparison between SMLR and PCA, very limited differences are found in terms of ROC/PR curves when all the selected resonant frequencies are included in the classification. Nevertheless, larger differences arise when inspecting the classifications obtained accounting for an increasing number of features as shown in Fig. 20. It is noted in Fig. 20 (a) that almost equally effective classifications are obtained using SMLR for every subset of features, while the classification using PCA in Fig. 20 (b) cannot achieve a comparable effectiveness until including all the modes in the classification. Indeed, a comparable classification is only found when including high-order Mode 9 (HO2). This is also confirmed by examining the AUCs of the previous ROC curves in Fig. 20 (c), which demonstrates that the damage classification using SMLR considerably outperforms the one obtained by PCA. Concerning crack amplitudes, almost perfect detection can be observed in Fig. 21. In the control chart shown in Fig. 21 (a), the number of outliers considering a 95% UCL increases from 4.99% to 84.79% before and after the seismic sequence. Furthermore, the ROC and PR curves reported in Fig. 21 (b) evidence almost perfect classification, with global metrics of AUC=0.9943 and PAUC=0.9760.

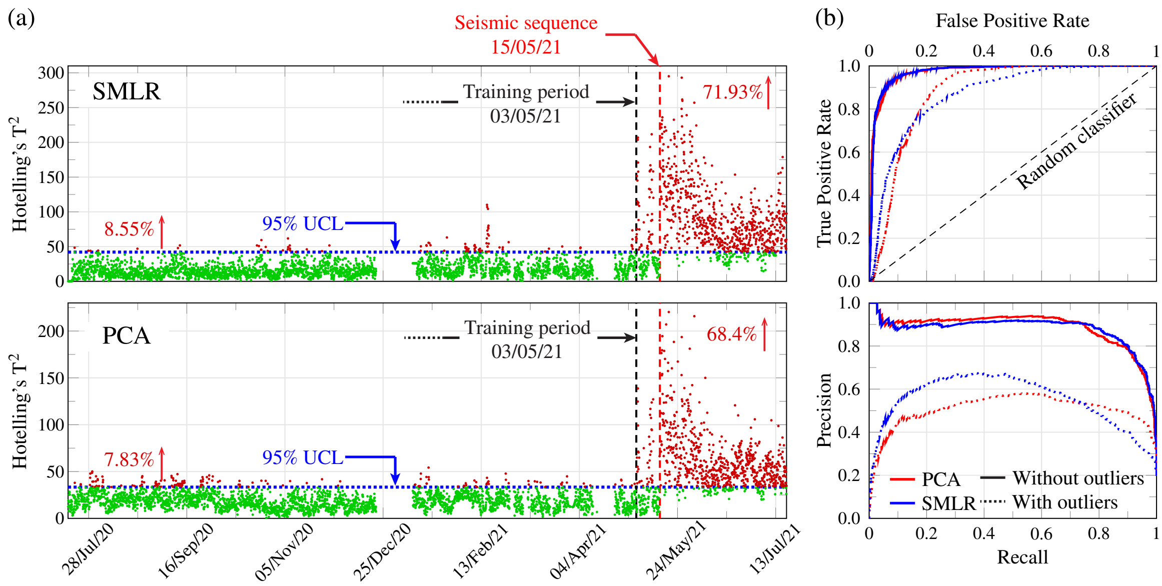

Figure 19: Hotelling's T^2 control charts of the residuals of the resonant frequencies of the Consoli Palace considering SMLR and PCA (3 PCs) (a), and quality assessment in terms of ROC/PR curves (b).

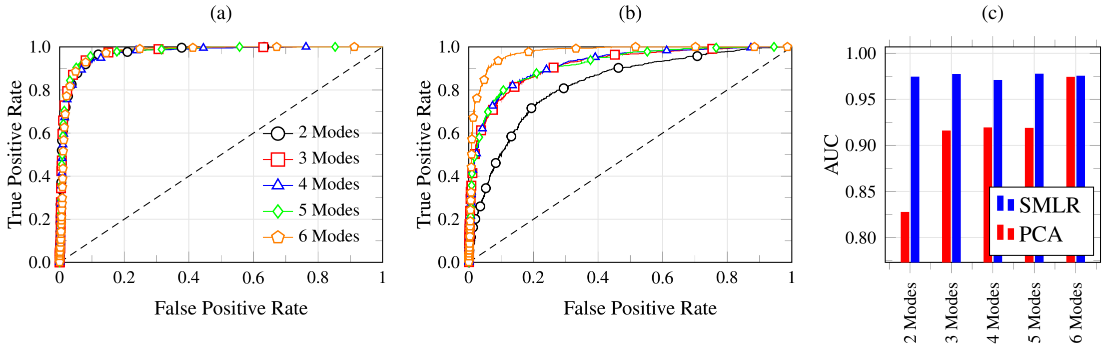

Figure 20: ROC curves considering an increasing number of resonant frequencies using SMLR (a) and PCA (b) (3 PCs), and comparison in terms of AUCs (c). Modes are included in the classification in increasing order of frequency.

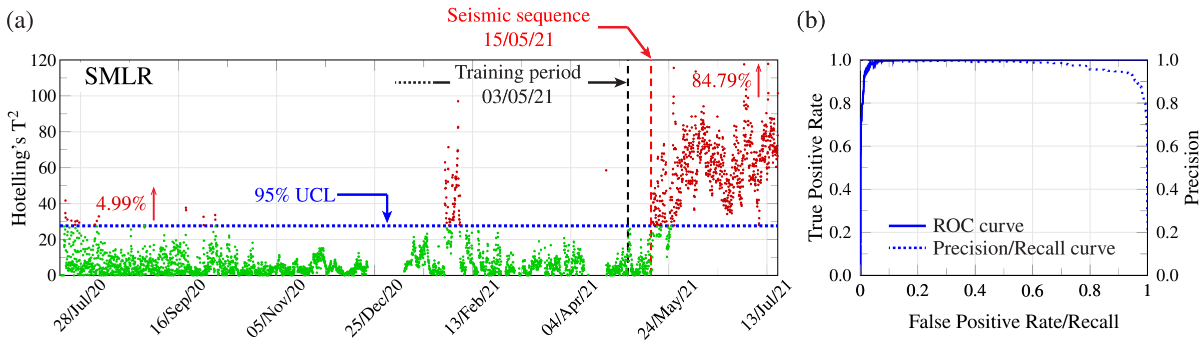

Figure 21: Hotelling's T^2 control charts of the residuals of the crack amplitudes D1 and D2 of the Consoli Palace considering SMLR (a), and quality assessment through a ROC/PR curve (b).

The characterization of the earthquake-induced decays in the resonant frequencies of the Consoli Palace is reported in Fig. 22. The analysis is performed by the individualized study of the residuals of the resonant frequencies obtained using both SMLR and PCA and presented in Figs. 22 (a) and (b), respectively. To facilitate the identification of shifts and minimize the effects of residual EOC variability, moving averages of order 192 (4 days) are included with solid black lines. Additionally, the convergence of the average of the residuals before and after the seismic sequence of May 15th are also included and denoted with dashed yellow lines. Although it was concluded from the previous analysis in Fig. 19 that PCA may show a comparable classification performance to SMLR when including all the considered resonant frequencies, it is evident in Fig. 22 (b) that this approach fails to provide a clear interpretation of the decays in the resonant frequencies. In fact, only a clear shift starting on May 15th is noticeable in Mode 4, while just mild deviations are recognized for Modes 1 and 3. Moreover, since PCA only exploits correlations between the estimators, the sign of the shifts may be hardly interpreted and/or related to physical phenomena. On the contrary, clear decays starting right after the onset of the seismic sequence are observable when implementing SMLR in Fig. 19 (a). Note that the signs of the shifts observed in this figure are all negative, indicating the appearance of earthquake-induced stiffness losses in the palace. It is important to remark that, although the number of data samples in the evaluation period is still limited, the mean convergence curves in Fig. 19 (a) suggest certain stabilization which allows to state the appearance of persistent damage in the palace. These results are further investigated in Fig. 23 through the analysis of the squared Mahalanobis distances (D^2)

of the residuals with respect to the training population. Data samples before and after the seismic sequence are denoted in this figure with Regions I and II, respectively. The probability distributions of the distances in Fig. 23 (a) (plotted in logarithmic scale and normalized to have unit maximum probability) exhibit clear shifts in terms of mode and mean values after the seismic sequence, which further supports the claim of the appearance of persistent structural damage. The earthquake-induced variations in the correlations between the resonant frequencies and the environmental temperature are investigated in Fig. 22 (b). In these analyses, the variation range of environmental temperature (channel T1) has been divided into 50 equally spaced disjoint intervals, and the statistical distribution of resonant frequencies has been described interval-wise through a frequentistic analysis. In is noted in this figure that an almost constant decay is found in Mode 3 throughout all the temperature range. Conversely, decays in Modes 1 and 4 concentrate in the temperature range up to 20°, while almost no variation is observed at higher temperatures. Nonetheless, future analyses should appraise a longer damage assessment period to fully characterize the permanent damage-induced variations in the environmental effects, covering the temperature range below 15° which remains unexplored in the present work.

Following a similar approach, Figs. 24 (a) and (b) report the obtained residuals of crack amplitudes D1 and D2 using SMLR and the analysis of their correlations with the environmental temperature (channel T1), respectively. In this case, it may be clearly concluded from Fig. 24 (a) that crack amplitudes D1 (in the south façade of the palace) experienced almost no variation after the seismic sequence, while a steep and stable shift is found in D2 (in the north façade of the palace). This fact is also confirmed when analysing the correlations with environmental temperature in Fig. 24 (b), where consistent decreases (closing) are found in the whole temperature range.

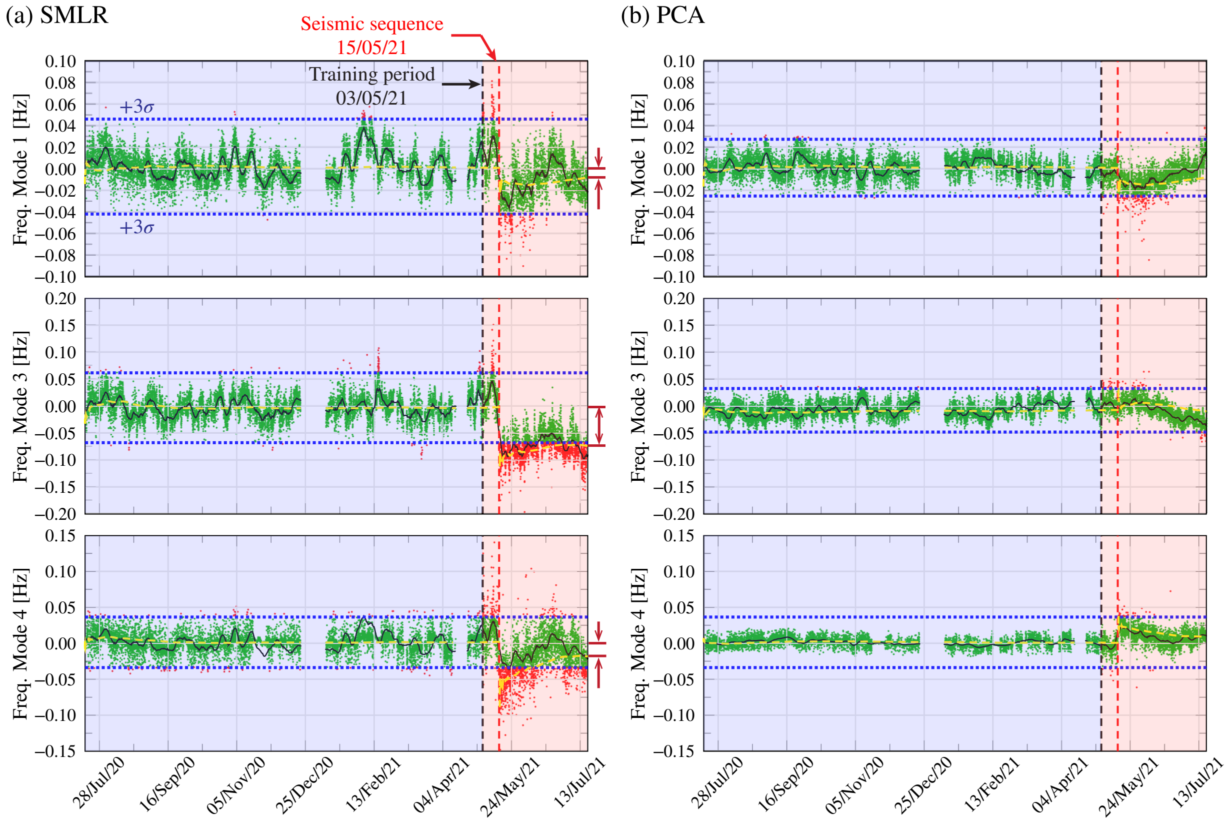

Figure 22: Residuals of the first three considered resonant frequencies of the Consoli Palace using SMLR (a) and PCA (b). Black solid lines represent the moving averages of order 192 (4 days) of the residuals, while yellow dashed lines represent the convergence of the mean values of the residuals before and after the seismic sequence occurred on May 15th 2021.

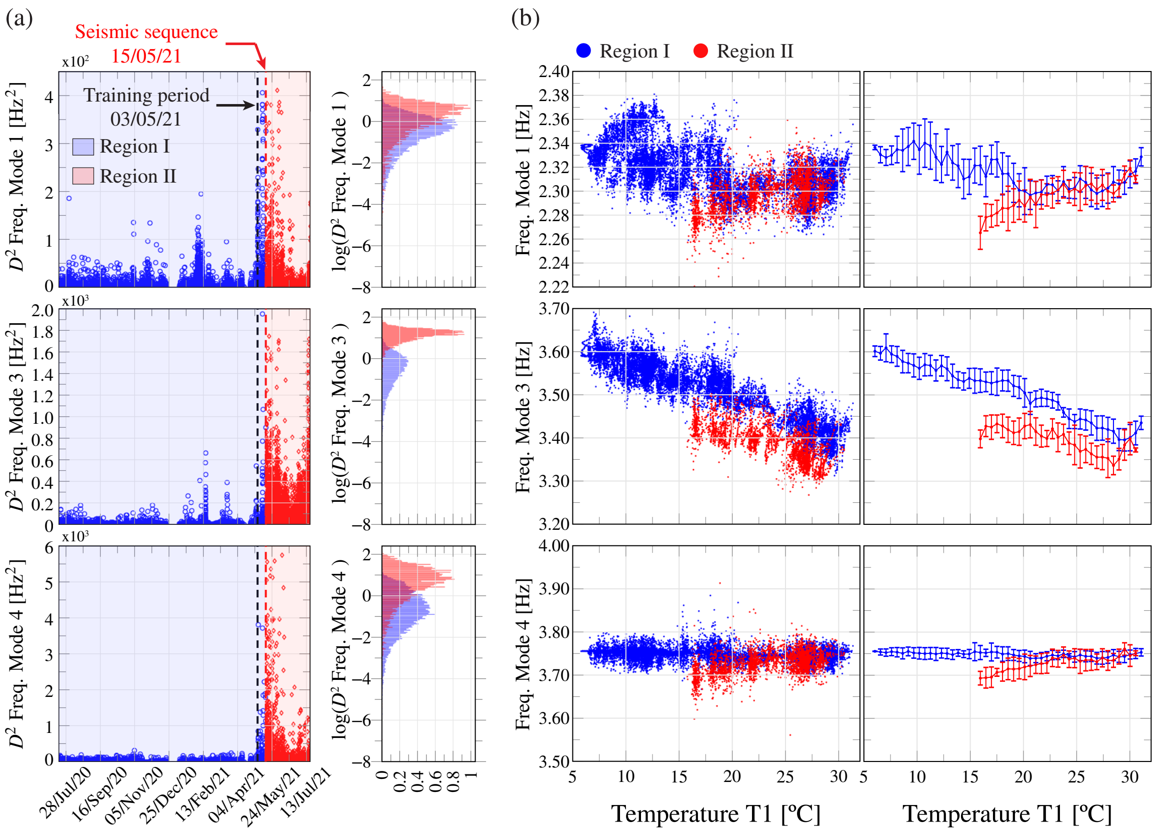

Figure 23: Identification of earthquake-induced decays in the resonant frequencies of Modes 1, 3 and 4 of the Consoli Palace. Statistical analysis of the distribution of Mahalanobis distances of the residuals against the training population (a), and characterization of the earthquake-induced damage in the frequency/temperature correlations (b). Error bars in (b) indicate the standard deviation.

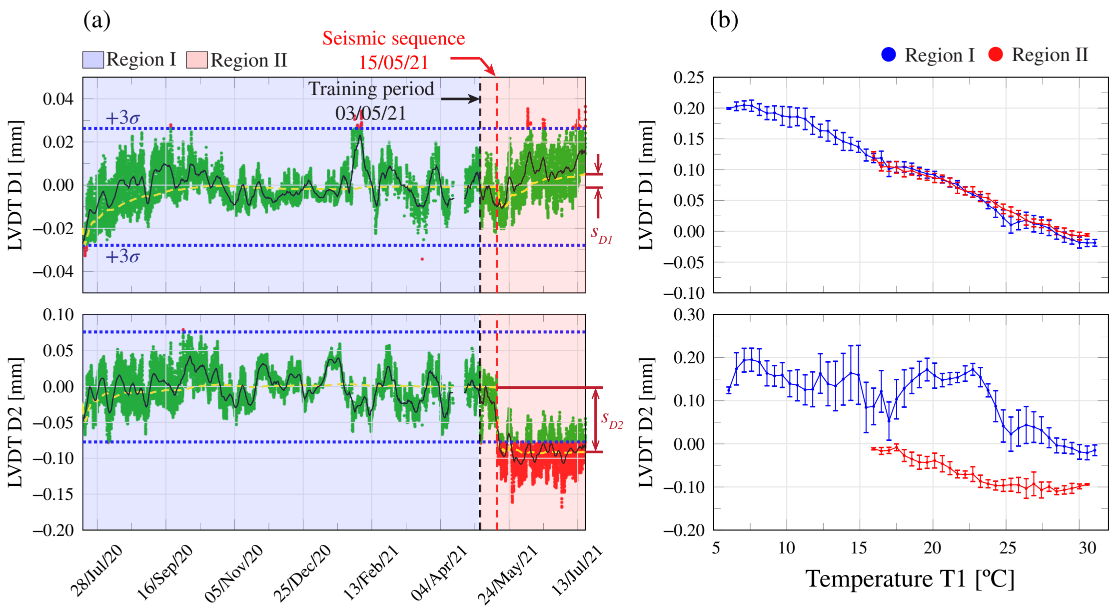

Figure 24: Residuals of crack amplitudes D1 and D2 of the Consoli Palace using SMLR (a), and identification of earthquake-induced variations in the crack/temperature correlations (b). Black solid lines in (a) represent the moving averages of order 192 (4 days) of the residuals, while yellow dashed lines represent the convergence of the mean values of the residuals before and after the seismic sequence occurred on May 15th 2021. Error bars in (b) indicate the standard deviation.

To conclude the previous analysis, Table 5 summarizes the identified earthquake-induced damage in the Consoli Palace. To compute the earthquake-induced shifts in the considered estimators, the statistical moments of the residuals presented in Figs. 22 and 24 are estimated by non-parametric bootstrap with 800 repetitions. As a measure of the uncertainty in the identification, the quantities in parentheses in Table 5 indicate the standard deviations of the empirical mean values computed in the bootstrap repetitions. In terms of resonant frequencies, decays concentrate in Modes 3 (L-T1), 5 (G-T1), 1 (G-By1) and 4 (G-Bx1) in decreasing order. Note that, given the relative orientation of the palace with respect to the epicenter of the main shock (see Fig. 17 (a)), damage is expected to concentrate along the y-direction of the building, primarily affecting the first bending mode along that direction (G-By1) as well as the torsional modes (L-T1 and G-T1). This claim is supported by the results reported in Table 5, where the decays in the first order bending modes along the y- and x-directions amount to 0.62 and 0.54%, respectively. Besides, the largest decays are found for torsional Modes 3 (local) and 5 (global) with values of 2.04 and 0.93%, respectively, indicating that the seismic events had largest influence upon the torsional stiffness of the palace. Interestingly, an increase of 0.62% is found for local Mode 9, which may indicate some earthquake-induced rearrangement of the tower/building interaction. However, given the limited performance of the data normalization of this mode shown in Fig. 16, as well as the largest uncertainty found in its estimation as reported in Table 5, the analysis of a longer monitoring period would be required to confirm whether this is a persistent variation or not. It is important to remark that no significant variations are found in the MAC values between the mode shapes after and before the seismic sequence, with values very close to 1 as reported in Table 5. No clear persistent variations were observed either in the time series of MAC values all throughout the monitoring period, thereby their analysis has been omitted herein. This circumstance may indicate the registered damage is very moderate and sustains no severe structural risk to the palace. In fact, preliminary in-situ inspections have not revealed any new pathology in the palace, which suggests that the developed damage remains at a degree not observable by visual examinations. With regard to the crack amplitudes, the results in Table 5 confirm that crack D1 was not affected by the seismic sequence, while a clear closure of crack D2 of about 8.9E-2 mm is found. Note that LVDT D2 is located bridging a major crack located in the north façade of the building, which was presumably originated as a result of an incipient overturning failure mechanism of the western façade. Instead, LVDT D1 is monitoring a crack relating the local overturning of the loggia in the south façade. Following the previous discussion on the incidence direction of the seismic shocks, it is reasonable to state that these will mainly affect the north façade of the palace. Finally, it is important to emphasize that, whilst the detected anomaly in D2 is significant in relatively terms, the closure of crack D2 is very limited and supports the previous statement on the mild severity of the earthquake-induced damage.

Table 5: Characterization of the shifts in the resonant frequencies and crack amplitudes of the Consoli Palace after the May 15th 2021 seismic sequence. Subscripts I and II relate the statistical moments calculated before and after the seismic sequence, respectively.

	July 17th 2020 to May 15th 2021		May 15th 2021 to July 18th 2021		Comparison	
Estimator $\mathbf{y}$	Mean $\bar{\mathbf{y}}_I$	Std. dev. $\sigma_{\mathbf{y},I}$	Mean $\bar{\mathbf{y}}_{II}$	Std. dev. $\sigma_{\mathbf{y},II}$	$100 \cdot (\bar{\mathbf{y}}_{II} - \bar{\mathbf{y}}_I)/\bar{\mathbf{y}}_I$	MAC
Mode 1 (G-By1) [Hz]	2.319 (±1.33E-04)	0.020	2.304 (±3.37E-04)	0.016	-0.62	0.977
Mode 3 (L-By1) [Hz]	3.514 (±1.99E-04)	0.068	3.442 (±4.26E-04)	0.041	-2.04	0.943
Mode 4 (G-Bx1) [Hz]	3.750 (±1.20E-04)	0.014	3.730 (±4.69E-04)	0.027	-0.54	0.903
Mode 5 (G-T1) [Hz]	4.186 (±2.38E-04)	0.057	4.147 (±6.39E-04)	0.031	-0.93	0.974
Mode 7 (L-T1) [Hz]	6.477 (±8.03E-04)	0.136	6.479 (±1.79E-03)	0.110	0.04	0.962
Mode 9 (HO2) [Hz]	7.971 (±1.61E-03)	0.292	8.020 (±3.57E-03)	0.248	0.62	0.977
Crack D1 [mm]	0.115 (±7.20E-05)	0.074	0.122 (±1.93E-04)	0.037	-	-
Crack D2 [mm]	0.120 (±2.21E-04)	0.072	0.031 (±4.26E-04)	0.036	-	-

4. Conclusions

This paper has presented the development of a novel methodology for statistical pattern recognition and identification of earthquake-induced structural damage. The proposed methodology comprises a first data cleansing stage using the MCD method to mitigate the adverse effects related to the existence of outliers in the training population. Afterwards, a sparse multivariate linear regression model is trained using LAR to eliminate the influence of EOC upon the dataset of damage-sensitive features. The proposed SMLR model allows to automatically identify the optimal set of EOC predictors (both static and dynamic), with the subsequent enhancement in the data normalization and minimal need for expert judgement. The effectiveness of the proposed approach is demonstrated with an application case study of a monumental masonry palace, the Consoli Palace in Gubbio. The Consoli Palace has been instrumented since July 14th 2020 with an aggregated static/dynamic/environmental SHM system. A seismic sequence of moderate intensity hit the palace on May 15th to May 27th 2021, including a main earthquake of magnitude Mw 4.0 followed by five 2.9 <Mw< 3.6 aftershocks with epicenters located only 2-3 km far from the palace. The reported results have demonstrated the effectiveness of the proposed approach to detect and quantify the earthquake-induced effects upon the resonant frequencies and the amplitudes of two major cracks of the Consoli Palace. The key findings and contributions of this work can be summarized as follows:

- A new damage identification methodology is proposed combining data cleansing and sparse multivariate regression. The reported results have demonstrated that the proposed approach can handle large sets of potential predictors in a fully automated way, including both static and dynamic (i.e. time delayed) EOC time series with minimal assistance by expert judgement.
- The reported results have highlighted the critical influence of data normalisation for achieving effective damage identification. Additionally, it has been evidenced that output-only data normalisation models such as PCA may fail at quantifying structural damage, providing anomalies in the residuals hardly attributable to physical phenomena.
- Maximum earthquake-induced decays of 2% are found for the fundamental bending and torsional modes of the Consoli Palace. Slightly stronger effects are found in the bending mode along the east-west direction

of the palace, which is conceivably explained by the relative orientation of the palace with respect to the incidence direction of the seismic shock. This observation is also justified by the analysis of the static data, where a persistent crack closure in the north façade of the palace has been clearly identified.

- No significant earthquake-induced effects are observed upon the time series of mode shapes of the palace. This circumstance, along with the impossibility to find new damage patterns in the palace by preliminary in-situ inspections, may indicate that the newly acquired damage condition remains at an early stage of development not visually observable and with no critical risk to the structural integrity of the building.

Acknowledgements

This work was supported by the Italian Ministry of Education, University and Research (MIUR) through the funded project of national interest "DETECT-AGING - Degradation Effects on sTructural safEty of Cultural heriTAGe constructions through simulation and health monitorING" (Protocol No. 201747Y73L).

References

[1] C. Gustafsson, B. Mellár, Research for CULT Committee-Best Practices in Sustainable Management and Safeguarding of Cultural Heritage in the EU., European Parliament, 2018.

[2] European tourism – Trends & Prospects – Quarterly Report Q1/2021., Tech. rep., ETC-European Travel Commission (2021).

[3] T. M. Ferreira, R. Vicente, J. A. R. M. Da Silva, H. Varum, A. Costa, Seismic vulnerability assessment of historical urban centres: case study of the old city centre in Seixal, Portugal, Bulletin of Earthquake Engineering 11 (5) (2013) 1753–1773.

[4] M. Villar-Vega, V. Silva, H. Crowley, C. Yepes, N. Tarque, A. B. Acevedo, M. A. Hube, C. D. Gustavo, H. S. María, Development of a fragility model for the residential building stock in South America, Earthquake Spectra 33 (2) (2017) 581–604.

[5] A. Rossi, A. Tertulliani, R. Azzaro, L. Graziani, A. Rovida, A. Maramai, V. Pessina, S. Hailemikael, G. Buffarini, F. Bernardini, The 2016-2017 earthquake sequence in Central Italy: macroseismic survey and damage scenario through the EMS-98 intensity assessment, Bulletin of Earthquake Engineering 17 (5) (2019) 2407–2431.

[6] K. Van Breugel, Societal burden and engineering challenges of ageing infrastructure, Procedia Engineering 171 (2017) 53–63.

[7] L. Binda, G. Gatti, G. Mangano, C. Poggi, G. S. Landriani, Collapse of the civic tower of Pavia: a survey of the materials and structure, Masonry International 6 (1) (1992) 11–20.

[8] J. S. Mitchell, From vibration measurements to condition-based maintenance, Sound and vibration 41 (1) (2007) 62–79.

[9] K. Gkoumas, F. L. Marques Dos Santos, M. van Balen, A. Tsakalidis, A. Ortega Hortelano, M. Grosso, G. Haq, F. Pekár, Research and Innovation in Bridge Maintenance, Inspection and Monitoring, Tech. rep. (2019).

[10] P. Cawley, Structural health monitoring: Closing the gap between research and industrial deployment, Structural Health Monitoring 17 (5) (2018) 1225–1244.

[11] C. R. Farrar, S. W. Doebling, D. A. Nix, Vibration-based structural damage identification, Philosophical Transactions of the Royal Society of London. Series A: Mathematical, Physical and Engineering Sciences 359 (1778) (2001) 131–149.

[12] L. M. Gandham, J. R. Kota, P. Kalapatapu, V. D. K. Pasupuleti, A survey on current heritage structural health monitoring practices around the globe., in: Digital Heritage. Progress in Cultural Heritage: Documentation, Preservation, and Protection, Springer International Publishing, 2020, pp. 565–576.

[13] G. Zini, M. Betti, G. Bartoli, A quality-based automated procedure for operational modal analysis, Mechanical Systems and Signal Processing 164 (2022) 108173.

[14] M. L. Pecorelli, R. Ceravolo, R. Epicoco, An automatic modal identification procedure for the permanent dynamic monitoring of the sanctuary of Vicoforte, International Journal of Architectural Heritage (2018) 1–15.

[15] L. F. Ramos, L. Marques, P. B. Lourenço, G. De Roeck, A. Campos-Costa, J. Roque, Monitoring historical masonry structures with operational modal analysis: two case studies, Mechanical Systems and Signal Processing 24 (5) (2010) 1291–1305.

[16] F. Ottoni, C. Blasi, Results of a 60-year monitoring system for santa maria del fiore dome in florence, International Journal of Architectural Heritage 9 (1) (2015) 7–24.

[17] M. Mishra, Machine learning techniques for structural health monitoring of heritage buildings: A state-of-the-art review and case studies, Journal of Cultural Heritage 47 (2020) 227–245.

[18] S. Sony, K. Dunphy, A. Sadhu, M. Capretz, A systematic review of convolutional neural network-based structural condition assessment techniques, Engineering Structures 226 (2021) 111347.

[19] C. Scuro, F. Lamonaca, S. Porzio, G. Milani, R. S. Olivito, ´Internet of Things (IoT) for masonry structural health monitoring (SHM): Overview and examples of innovative systems, Construction and Building Materials 290 (2021) 123092.

[20] G. D. Zhou, T. H. Yi, A summary review of correlations between temperatures and vibration properties of long-span bridges, Mathematical Problems in Engineering 2014.

[21] G. Zonno, R. Aguilar, R. Boroschek, P. B. Lourenço, Analysis of the long and short-term effects of temperature and humidity on the structural properties of adobe buildings using continuous monitoring, Engineering Structures 196 (2019) 109299.

[22] R. Ceravolo, G. Coletta, G. Miraglia, F. Palma, Statistical correlation between environmental time series and data from long-term monitoring of buildings, Mechanical Systems and Signal Processing 152 (2021) 107460.

[23] F. Ubertini, G. Comanducci, N. Cavalagli, A. L. Pisello, A. L. Materazzi, F. Cotana, Environmental effects on natural frequencies of the San Pietro bell tower in Perugia, Italy, and their removal for structural performance assessment, Mechanical Systems and Signal Processing 82 (2017) 307–322.

[24] R. M. Azzara, G. De Roeck, M. Girardi, C. Padovani, D. Pellegrini, E. Reynders, The influence of environmental parameters on the dynamic behaviour of the San Frediano bell tower in Lucca, Engineering Structures 156 (2018) 175–187.

[25] C. Gentile, A. Ruccolo, F. Canali, Long-term monitoring for the condition-based structural maintenance of the Milan Cathedral, Construction and Building Materials 228 (2019) 117101.

[26] E. García-Macías, F. Ubertini, Automated operational modal analysis and ambient noise deconvolution interferometry for the full structural identification of historic towers: A case study of the Sciri Tower in Perugia, Italy, Engineering Structures 215 (2020) 110615.

[27] E. M. Tronci, M. De Angelis, R. Betti, V. Altomare, Vibration-based structural health monitoring of a RC-masonry tower equipped with non-conventional TMD, Engineering Structures 224 (2020) 111212.

[28] A. Cabboi, C. Gentile, A. Saisi, From continuous vibration monitoring to FEM-based damage assessment: application on a stone-masonry tower, Construction and Building Materials 156 (2017) 252–265.

[29] A. Deraemaeker, K. Worden, A comparison of linear approaches to filter out environmental effects in structural health monitoring, Mechanical Systems and Signal Processing 105 (2018) 1–15.

[30] H. F. Zhou, Y. Q. Ni, J. M. Ko, Eliminating temperature effect in vibration-based structural damage detection, Journal of Engineering Mechanics 137 (12) (2011) 785–796.

[31] A. Entezami, H. Shariatmadar, S. Mariani, Early damage assessment in large-scale structures by innovative statistical pattern recognition methods based on time series modeling and novelty detection, Advances in Engineering Software 150 (2020) 102923.

[32] G. Coletta, G. Miraglia, M. Pecorelli, R. Ceravolo, E. Cross, C. Surace, K. Worden, Use of the cointegration strategies to remove environmental effects from data acquired on historical buildings, Engineering Structures 183 (2019) 1014–1026.

[33] F. Lorenzoni, F. Casarin, M. Caldon, K. Islami, C. Modena, Uncertainty quantification in structural health monitoring: Applications on cultural heritage buildings, Mechanical Systems and Signal Processing 66 (2016) 268–281.

[34] I. Flood, N. Kartam, Neural networks in civil engineering. II: Systems and application, Journal of computing in civil engineering 8 (2) (1994) 149–162.

[35] R. Kromanis, P. Kripakaran, Support vector regression for anomaly detection from measurement histories, Advanced Engineering Informatics 27 (4) (2013) 486–495.

[36] J. Maeck, B. Peeters, G. De Roeck, Damage identification on the Z24 bridge using vibration monitoring, Smart materials and structures 10 (3) (2001) 512.

[37] B. Peeters, G. De Roeck, One-year monitoring of the Z24-Bridge: environmental effects versus damage events, Earthquake Engineering & Structural Dynamics 30 (2) (2001) 149–171.

[38] R. Langone, E. Reynders, S. Mehrkanoon, J. A. K. Suykens, Automated structural health monitoring based on adaptive kernel spectral clustering, Mechanical Systems and Signal Processing 90 (2017) 64–78.

[39] P. Singh, M. Keyvanlou, A. Sadhu, An improved time-varying empirical mode decomposition for structural condition assessment using limited sensors, Engineering Structures 232 (2021) 111882.

[40] M. Mousavi, A. H. Gandomi, Prediction error of Johansen cointegration residuals for structural health monitoring, Mechanical Systems and Signal Processing 160 (2021) 107847.

[41] A. Pierdicca, F. Clementi, P. Mezzapelle, A. Fortunati, S. Lenci, One-year monitoring of a reinforced concrete school building: Evolution of dynamic behavior during retrofitting works, Procedia engineering 199 (2017) 2238–2243.

[42] S. K. Kourkoulis, Recent advances in structural health monitoring of restored elements of marble monuments, Procedia Structural Integrity 10 (2018) 3–10.

[43] M. G. Masciotta, L. F. Ramos, P. B. Lourenço, The importance of structural monitoring as a diagnosis and control tool in the restoration process of heritage structures: a case study in Portugal, Journal of Cultural Heritage 27 (2017) 36–47.

[44] E. Mesquita, A. Arêde, N. Pinto, P. Antunes, H. Varum, Long-term monitoring of a damaged historic structure using a wireless sensor network, Engineering Structures 161 (2018) 108–117.

[45] A. Saisi, C. Gentile, M. Guidobaldi, Post-earthquake continuous dynamic monitoring of the Gabbia Tower in Mantua, Italy, Construction and Building Materials 81 (2015) 101–112.

[46] F. Ubertini, N. Cavalagli, A. Kita, G. Comanducci, Assessment of a monumental masonry bell-tower after 2016 Central Italy seismic sequence by long-term SHM, Bulletin of Earthquake Engineering 16 (2) (2018) 775–801.

[47] H. Hotteling, Multivariate quality control, illustrated by the air testing of sample bombsights, Techniques of statistical analysis (1947) 111–184.

[48] P. J. Rousseeuw, Least median of squares regression, Journal of the American statistical association 79 (388) (1984) 871–880.

[49] M. Hubert, M. Debruyne, P. J. Rousseeuw, Minimum covariance determinant and extensions, Wiley Interdisciplinary Reviews: Computational Statistics 10 (3) (2018) e1421.

[50] P. J. Rousseeuw, K. V. Driessen, A fast algorithm for the minimum covariance determinant estimator, Technometrics 41 (3) (1999) 212–223.

[51] P. J. Rousseeuw, A. M. Leroy, Robust regression and outlier detection, Vol. 589, John Wiley & Sons, 2005.

[52] M. Hubert, P. J. Rousseeuw, S. Van Aelst, Multivariate outlier detection and robustness, Handbook of Statistics 24 (2005) 263–302.

[53] B. Efron, T. Hastie, I. Johnstone, R. Tibshirani, Least angle regression, Annals of statistics 32 (2) (2004) 407–499.

[54] T. Hastie, R. Tibshirani, M. Wainwright, Statistical learning with sparsity: the lasso and generalizations, Chapman and Hall/CRC, 2019.

[55] G. Schwarz, Estimating the dimension of a model, The annals of statistics (1978) 461–464.

[56] H. Akaike, On entropy maximization principle, Applications of Statistics, in: Proceedings of the Symposium held at Wright State University, North-Holland Publishing Company, 1977, pp. 27–41.

[57] N. Cavalagli, A. Kita, S. Falco, F. Trillo, M. Costantini, F. Ubertini, Satellite radar interferometry and in-situ measurements for static monitoring of historical monuments: The case of Gubbio, Italy, Remote Sensing of Environment 235 (2019) 111453.

[58] N. Cavalagli, A. Kita, V. L. Castaldo, A. L. Pisello, F. Ubertini, Hierarchical environmental risk mapping of material degradation in historic masonry buildings: An integrated approach considering climate change and structural damage, Construction and Building Materials 215 (2019) 998–1014.

[59] 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 (2019) 180–202.

[60] 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) 106830.

[61] F. Mirabella, M. G. Ciaccio, M. R. Barchi, S. Merlini, The Gubbio normal fault (Central Italy): geometry, displacement distribution and tectonic evolution, Journal of Structural Geology 26 (12) (2004) 2233–2249.

[62] P. Balocchi, G. Riga, Considerations on the Seismotectonics and Seismogenesis of Tiberina-Gubbio Valley Extensional System (Central Italy), Atti della Societa dei Naturalisti e Matematici di Modena 148 (2017) 65–82.

[63] L. Valoroso, L. Chiaraluce, R. Di Stefano, G. Monachesi, Mixed-mode slip behavior of the Altotiberina low-angle normal fault system (Northern Apennines, Italy) through high-resolution earthquake locations and repeating events, Journal of Geophysical Research: Solid Earth 122 (12) (2017) 10–220.

[64] F. Ubertini, C. Gentile, A. L. Materazzi, Automated modal identification in operational conditions and its application to bridges, Engineering Structures 46 (2013) 264–278.

[65] S. M. Kay, Modern spectral estimation: theory and application, Pearson Education India, 1988.