Preserving the World Heritage: Post-Earthquake Monitoring Based on Structural Break Testing with Deep Temporal Convolutional Features

Francesco Dente[1,2] (✉), Andy Combey[1], Alix Lhéritier[3], Rodrigo Acuna-Agost[3], and E. Diego Mercerat[1]

[1] Université Cote d'Azur, CEREMA, IRD, CNRS, Observatoire de la Cote d'Azur, Géoazur, France diego.mercerat@cerema.fr, andy.combey@geoazur.unice.fr
[2] EURECOM, France {francesco.dente}@eurecom.fr
[3] Amadeus, France {alix.lheritier, rodrigo.acunaagost}@amadeus.com

Abstract. Built heritage faces nowadays increasing vulnerability due to the combined impact of climatic, seismic, and anthropogenic forcings. In this context, vibration-based monitoring has become a key non-invasive method for assessing the integrity of historical buildings. However, little attention has been given to the development of automatic tools, which are crucial for rapid and effective decision-making. This study examines San Cristobal Church, a 17th-century building located in the UNESCO World Heritage site of Cusco, Peru. The church has been continuously monitored during 17 months using a seismic sensor located on one of its walls. First, we develop machine learning models to predict the church's natural frequencies based only on weather data. Then, we analyze deviations from the expected frequency variations to detect anomalies that may indicate structural changes in the building, especially following strong transient events such as earthquake-induced motions. We evaluate multiple machine learning approaches, including Ridge Regression, Feedforward Neural Networks, and Temporal Convolutional Networks, with the latter outperforming other models in capturing nonlinear temporal dependencies. To estimate the post-seismic recovery time of the natural frequencies following a Mw 4.2 earthquake occurred in August 13th, 2024, we employ the Bai-Perron test for structural break detection on the learned deep temporal convolutional features. As this recovery time is influenced by the damage state, changes in its duration can reflect alterations in masonry mechanical properties. By accurately assessing the post-seismic recovery time, our methodology offers a promising approach for developing early warning systems to identify damage in historical buildings.

Keywords: Structural Health Monitoring · Historical Masonry · Vibration Analysis · Temporal Convolutional Networks · Structural Break Analysis · Bai-Perron Test

1 Introduction

In the 21st century, preserving and protecting the cultural heritage is among the principal missions of UNESCO organization. This heritage faces increasing

vulnerability due to the combined impact of climatic, seismic, and anthropogenic forcings. In the field of Structural Health Monitoring (SHM), the characterization of the dynamic response of structures — to both continuous environmental and transient seismic loadings — is widely recognized as an effective tool for identifying the modal properties of existing structures [17, 15, 12].

In particular, the use of ambient vibration instead of external artificial excitation is highly valuable, as it is quick and easy to implement for structural health assessment and is often integrated into continuous monitoring systems [28]. This approach enables a data-driven evaluation of structural conditions, which is particularly beneficial for heritage structures where invasive testing methods should be avoided. As highlighted by numerous studies, variations in the natural frequencies of buildings have been observed both during and after earthquakes [7, 27], even in the absence of structural damage [9]. If no permanent structural damage occurs, the frequency shifts gradually diminish over time, normally returning to pre-seismic values. The duration of this recovery period depends on the mechanical properties of the materials and the damage state of the structure [19, 16, 1]. However, the effects of environmental factors can mask and complicate an accurate estimation of this recovery period. A proper understanding of the environmental response is essential to develop a reliable model that distinguishes between transient environmental effects and actual structural changes.

Unlike reinforced concrete structures, historical masonry buildings exhibit complex responses that are less studied; the lack of experimental data, and consequently robust models, limits preservation strategies for these heritage structures. No research has focused on the nonlinear elastic response (recovery period) of historical buildings, mainly due to the complexity of their behavior. This highlights the need to better understand the factors influencing their response using machine learning algorithms. While some studies are emerging in Western Europe, particularly in Italy [13, 25, 21, 2], the field remains largely unexplored elsewhere. Studying and better characterizing the seismic response of historical buildings is promising for improving the assessment of their structural health and damage state.

This research focuses on Peru, which has a unique heritage of traditional masonry structures made of stone and earthen materials, distinct from the typical materials used in European historical masonry. The goal is to develop innovative, efficient and lightweight computational tools for a more accurate evaluation of their structural health. This study focuses on the San Cristobal Church, a 17th-century colonial building in the UNESCO World Heritage site of Cusco, Peru. The church has been continuously monitored for 17 months using a tri-axial Raspberry Shake (3C) velocimeter to record its response to ambient and transient vibrations. In addition, environmental data, including temperature, humidity, wind speed, atmospheric pressure, and rainfall, have been collected from a weather station in Cusco located 3.2 km from the site. The unique value of our dataset is that it includes the recording of a $M_w 4.2$ earthquake that occurred on August 13, 2024, which temporarily altered the church's dynamic properties. A key contribution of this study is to focus on a historical masonry structure

Fig. 1: a) San Cristobal church from Cusco city centre, view from the South. The small green triangle indicates the position of the Raspberry Shake sensor. Modified from Martin St-Amant - Wikipedia - CC-BY-SA-3.0. b) Interior of the church with the Raspberry Shake sensor: small white box on the window sill.

built using stone and earthen materials (known as "adobe") Such materials are underrepresented in the literature and, to the best of our knowledge, no prior machine learning study has attempted to model their environmental response. Although previous work [29, 30] examined the response of earthen masonry to weather conditions, they do not develop predictive models.

Our approach is to develop machine learning models capable of predicting the church's natural frequencies under normal environmental conditions, establishing a baseline response of the undamaged structure. By doing so, we aim to predict the natural frequency variations induced by weather parameters, allowing us to identify non-predicted deviations as a proxy for detecting structural anomalies.

The key focus of this work is to estimate the post-seismic recovery period of the structure's dynamic properties. This estimation is crucial: when an earthquake affects a structure, its natural frequency should not return to its previous value if significant damage has occurred. However, with only a single sensor, localized damage elsewhere may not always be detectable from the data. As a result, the natural frequency could appear to recover even if there is some damage in parts of the church that are not monitored. Laboratory experiments and data from reinforced concrete buildings suggest that the recovery period is directly related to the degree of heterogeneity of the materials and therefore with their mechanical properties [26, 16]. Consequently, a change in recovery period may indicate evolving masonry properties, even in the absence of visible damage. This leads us to our primary research hypothesis: machine learning can assess, model, and predict these recovery periods, potentially identifying structural changes before they result in visible, significant damage.

In summary, the contributions of this paper are twofold: i) we develop and compare different machine learning models to predict the natural frequency of the San Cristobal Church using exclusively environmental parameters before the

earthquake, assumed as the undamaged state; ii) we use a methodology based on structural break testing with Deep Temporal Convolutional Features to refine the estimate of post-seismic recovery time.

2 Methodology

Our methodology has three main components: (i) data preprocessing and feature engineering, (ii) predictive modeling of the natural frequency under normal conditions using machine learning, and (iii) post-seismic recovery period estimation based on structural break testing with deep temporal convolutional features.

2.1 Data Collection and Preprocessing

The dataset consists of two primary data sources: i) Seismic data recorded by a triaxial seismometer installed in the south wall of the church's nave, at a height of around 6 meters, with a sampling rate of 100 Hz. The natural frequencies are extracted from this data using the Random Decrement Technique [10, 11]; ii) Environmental data obtained from a Davis Vantage Pro 2 weather station located approximately 3.2 km from the church, providing 10 minutes resolution measurements of temperature, humidity, wind speed, atmospheric pressure, and rainfall. To align the environmental data with the seismic data, all time series are downsampled to an hourly resolution. Weather parameters are averaged, except for rainfall, which is summed. Additionally, missing values are estimated using linear interpolation, as the gaps are short, with a maximum of one week of missing data. In July 2023, we carried out the Operational Modal Analysis [8] of the church from which several structural modes were identified. In the present paper, we focus specifically on the wandering of the natural frequency corresponding to the first torsional mode of the bell tower (around 6.7 Hz), as it is the most excited mode at the sensor location and provides a clean time series with minimal uncertainty. The complete time series is shown in Figure 2.

2.2 Predictive Modeling

To predict the natural frequency from environmental data, we evaluate multiple machine learning models:

- **Baseline Linear Model**: Ridge Regression [18] is used as a baseline to benchmark more complex models. Linear models are commonly applied in SHM, as noted in [24, 23].
- **Feedforward Neural Network**: a feedforward neural network (FNN) is implemented to model nonlinear dependencies between environmental parameters and the natural frequency. The model is trained using the Adam optimizer with mean squared error as the loss function, and early stopping is applied to prevent overfitting.

– **Temporal Convolutional Networks**: To better capture temporal dependencies, we consider Temporal Convolutional Networks (TCN) [20]. A TCN is a generic 1D-convolutional architecture for sequence prediction with additional key characteristics: (i) It employs causal convolutions, ensuring that the output at time t depends only on the current and past inputs. (ii) A simple causal convolution can only look back at a history (receptive field) that grows linearly with the network depth. To efficiently capture long-range dependencies, dilated convolutions are used, expanding the receptive field exponentially with respect to the number of hidden layers. The transformation of an input sequence $\mathbf{x}_0^t$ through a hidden dilated causal convolutional layer is defined as

$$h^{(l)}(s) = \sum_{i=0}^{k-1} f^{(l)}(i) x_{s-di} \ , \ \forall s \in \{0, \ldots, n-k\} \ , \tag{1}$$

where $h^{(l)}(s)$ is the output of the convolutional layer at position s in layer l, $f^{(l)}(i)$ is the learnable convolutional filter at layer l with i indexing the filter coefficients, k is the filter size and d is the dilation factor, which controls the spacing between kernel elements. (iii) Residual connections facilitate stable training, allowing the TCN to model long-term dependencies by stacking multiple layers. Finally, differently from a standard sequence prediction TCN, we focus only on the final time step: $\hat{y}_t = f(x_0, x_1, ..., x_t)$ where $\mathbf{x}_0^t \equiv x_0, x_1, \ldots, x_t$ is the input sequence, and $\hat{y}_t$ is the predicted output corresponding to time t. The model is trained using the Adam optimizer with mean squared error as the loss function, and early stopping is applied to prevent overfitting. The final prediction is obtained through a linear layer, which maps the extracted deep convolutional features to the output.

2.3 Structural Break Testing

As the building's natural frequency responds to weather variations, we cannot assess its recovery by considering the variation of the frequency alone. Instead, our approach is to identify changes in the frequency response to the weather variables. In Economics and Statistics, Structural Break Analysis aims at identifying changes over time in the parameters of regression models, called *structural breaks*. The Bai-Perron (BP) sequential test [4, Sec. 5.2.2][5, Sec. 5.3] identifies an unknown number of structural breaks in linear regression models and assesses their statistical significance, by estimating them one at a time as in [3]. After $l \geq 0$ breaks $\hat{T}_1, \ldots, \hat{T}_l$ have been found (when $l = 0$, the sequence is empty), the test considers an additional break at each possible location within each subsample defined by the l breaks (or within the whole sample, if $l = 0$). Let $\hat{T}_0 = 0$ and $\hat{T}_{l+1} = T$, with T being the sample size, and $S_T(\hat{T}_1, \ldots, \hat{T}_m)$ denote the sum of squared residuals obtained by fitting a linear model on each of the subsamples defined the break points $\hat{T}_1, \ldots, \hat{T}_m$. Then, the following statistic is considered to test the null hypothesis of l breaks $\hat{T}_1, \ldots, \hat{T}_l$ against the alternative of $l + 1$

breaks:

$$F_T(l+1 \mid l) = \left\{ S_T\left(\hat{T}_1, \ldots, \hat{T}_l\right) - \min_{1 \le i \le l+1} \inf_{\tau \in \Lambda_{i,\eta}} S_T\left(\hat{T}_1, \ldots, \hat{T}_{i-1}, \tau, \hat{T}_i, \ldots, \hat{T}_l\right) \right\} / \hat{\sigma}^2 \tag{2}$$

where

$$\Lambda_{i,\eta} = \left\{ \tau; \hat{T}_{i-1} + \left(\hat{T}_i - \hat{T}_{i-1}\right)\eta \le \tau \le \hat{T}_i - \left(\hat{T}_i - \hat{T}_{i-1}\right)\eta \right\}$$

with η being a parameter defining the minimal length between two adjacent breaks and $\hat{\sigma}^2$ a consistent estimate of the variance of the disturbance σ^2 under the null hypothesis. Critical values for this statistic are provided in [4, Table II]. When $F_T(l+1 \mid l)$ is above the critical value, the null is rejected and the value of τ minimizing the second term of the right-hand side Eq. 2 is retained as the $l+1$-th break and appended to the previous list of breaks. The test starts with $l = 0$ breaks and keeps increasing l by 1 as long as the null hypothesis is rejected and a maximum number of breaks is not reached. The linear assumption allows the optimization problem to be efficiently solved, resulting in $O(T)$ complexity for the entire procedure.

To extend the structural break analysis to a non-linear framework, our approach consists in processing the raw weather features using the learned TCN and taking the output of the last hidden layer as input features for the structural break analysis—thus replacing the last linear layer of the TCN by the linear models fitted by the BP test. As shown in Sec. 3.5, we use this test to estimate the length of the recovery process and finally assess if it recovered at all by concatenating the time series before the earthquake with that after the hypothesized recovery and testing for breaks—no rejection supporting the full recovery hypothesis, up to a type II error.

3 Application Case: San Cristobal Church

We now present the results obtained with our methodology on the frequency response of the San Cristobal Church to the weather conditions.

The code and the dataset necessary to reproduce the results can be found at: https://github.com/francdente/SHM-post_earthquake_monitoring and https://zenodo.org/records/15641078.

3.1 Experimental Setup

Temporal Constraints and Limited Data Availability A key challenge is maintaining temporal relationships in the data, ensuring that validation and test sets occur strictly after the training set to prevent the model from predicting the past using future data. This issue is particularly critical for highly seasonal datasets such as ours, which has distinct wet and dry seasons. If the training set

consists of dry season data while the validation set is from the rainy season, the model may struggle to generalize, making validation results less meaningful. Even with time series cross-validation [6], which preserves temporal order, validation and test splits may still fail to represent the full distribution, particularly in cases of strong seasonality. The dataset consists of approximately one year and four months of data, whereas at least two years would be ideal to better capture seasonal variations in the validation and test sets. This limitation is complicated by the occurrence of an earthquake on August 13, 2024, which disrupts normal structural behavior, leaving only about a year of usable pre-earthquake data. To mitigate this constraint, only 10% of the pre-earthquake data is allocated for validation when training neural networks. This allows the network to be trained on nearly a full year of pre-earthquake data, reducing the risk of overfitting while still maintaining a sufficient validation set to evaluate performance.

Challenges with Neural Networks and Cross-Validation Time series cross-validation poses challenges for neural networks, primarily due to training size bias, where increasing training data in each split artificially improves performance, making comparisons difficult. Additionally, small training splits in early folds can be problematic as neural networks require large datasets in relation to the number of parameters they have to perform well. Smaller training splits in early cross-validation iterations result in underperforming models, and these results are not representative of the network's true potential.

Dataset Partitioning Following the previous considerations, the dataset was partitioned according to the earthquake of August 13. The pre-earthquake period is used for training, including validation data, while the post-earthquake period is designated as the test set. The post-earthquake test set may not fully represent the generalization error of the model due to the sudden frequency disturbance caused by the earthquake and the gradual recovery of elastic properties. The recovery duration depends on material properties and structural damage [19][16]. Since no visible damage is observed and the frequency value return to pre-earthquake levels, there is no evidence of significant structural damage. Thus, the period during which the frequency had not yet stabilized is classified as abnormal behavior. To determine a first naive estimate of the recovery duration, we applied a conservative moving average with a 12-day window to identify when frequency values returned to pre-earthquake levels ($f = 6.76$ Hz). This resulted in around 2.5 months of estimated recovery period. The natural frequency values calculated after this period are assumed to reflect normal behavior and are used to estimate the generalization error of the predictive model. This setup is visually shown in Figure 2. It allows both the evaluation of the model generalization under normal conditions and the investigation of the lingering effects of the earthquake on the structure. By applying our methodology, we aim to refine the estimate of the recovery period, moving beyond the naive assumptions made during the preliminary analysis.

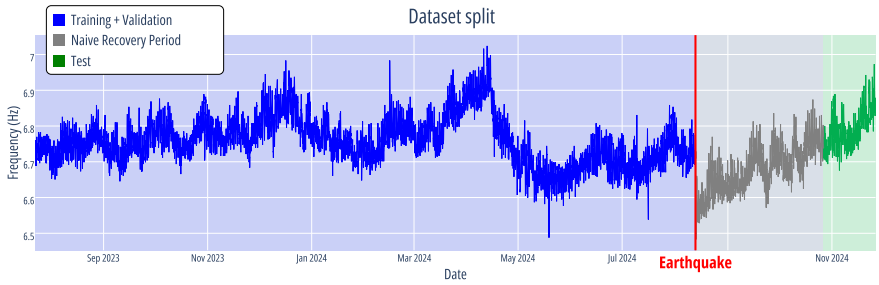


Fig. 2: Visual representation of dataset splits.

Evaluation Metric: Root Mean Squared Error (RMSE) We use the Root Mean Squared Error (RMSE) as the primary evaluation metric, as it is widely adopted in regression tasks and retains the same unit as the target variable (Hz), providing an intuitive measure of error magnitude.

3.2 Feature Engineering

Feature engineering is essential for enhancing the performance of linear models. This section focuses on modeling the impact of past weather conditions and examining the correlation between weather variables and natural frequencies over various time shifts. To effectively capture these relationships, we use both shifted weather values and rolling window averages as newly crafted features.

Rolling windows features Using only current weather values (t) ignores past influences, while adding multiple lagged values increases model complexity and risks overfitting. To address this, rolling window averages are used to capture past trends and cumulative effects. Window sizes are treated as hyper-parameter.

Correlation Analysis To better understand the relationships between input weather features and natural frequency, we compute the correlation coefficients for various weather parameters with different time shifts. Figure 3 shows the correlations across these shifts. Shifting weather data significantly improves correlation for humidity and temperature, with temperature changes linked to thermal expansion and humidity showing a 13-hour lag due to its anti-correlation with temperature. A longer analysis indicates that rainfall has the highest correlation with natural frequency after 4.4 days (106 hours), likely due to moisture absorption in adobe walls[4] and soil saturation effects on the foundation. Based on these findings, the original values of temperature, humidity, and rainfall at time t are replaced with their respective lagged counterparts: temperature ($t-2$ hours), humidity ($t-13$ hours), and rainfall ($t-106$ hours).

[4] Adobe walls, composed of earth, clay, and straw, offer good insulation but are highly susceptible to moisture from prolonged rainfall.

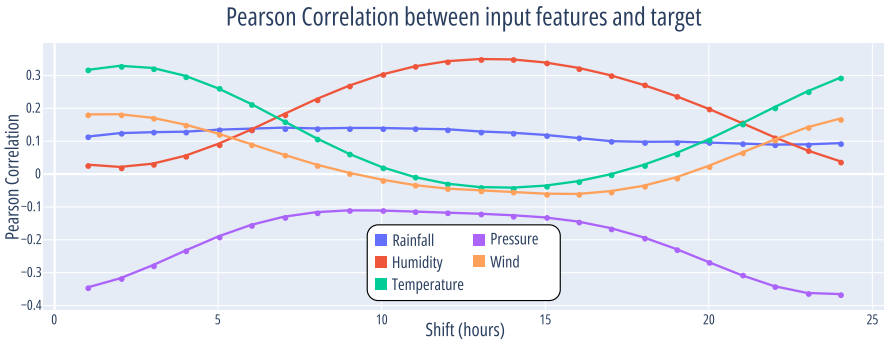


Fig. 3: Pearson correlation between weather parameters and natural frequency at different shifts. The x-axis represents the time shift (up to 24 hours in the past), while the y-axis shows the correlation coefficient for each weather variable.

3.3 Ridge Regression

Training Strategy and Feature Engineering We optimize key hyperparameters using time series cross-validation [6] with 3 splits, as implemented in `TimeSeriesSplit` from `scikit-learn`. For the regularization weight, values of 0.1, 1, 10, 100, and 1000 are tested. Rolling window sizes are also explored, with different ranges for each variable. For pressure and temperature, window sizes from 0 to 300 past hours are evaluated in increments of 50, as the correlation analysis highlights the importance of recent values. For humidity and rainfall, window sizes from 0 to 600 past hours are tested in increments of 50, as past values in the correlation analysis appear to have a significant influence.

Results and Observations We evaluate the Ridge Regression model against a dummy model that predicts the mean of the training set. Cross-validation optimization selects the best hyperparameters: a regularization weight of 0.1, a rolling window of 450 hours for rainfall, and 150 hours for temperature, with no rolling window for pressure, humidity, or wind speed. Ridge Regression outperforms the dummy model but still struggles to capture natural frequency peaks accurately. Table 1 summarizes the RMSE values for both models. Poorly modeled peaks in natural frequency often coincide with periods of heavy rainfall in Cusco. Comparing models with and without a past window for rainfall (Figure 4) reveals that aggregating past rainfall data significantly improves peak prediction. Simply including rainfall as a feature without applying a past window does not enhance performance, highlighting the importance of capturing cumulative effects (see Figure 5). This is evident in the test set, where early rainfall in 2024 caused a peak that the model without past aggregation failed to capture. These findings suggest that prolonged rainfall impacts the structure, likely through moisture absorption in adobe walls or changes in the foundation.

Table 1: Comparison of RMSE values for Ridge Regression with (w) and without (w/o) past rainfall window and Dummy Model.

Model	Train RMSE	Validation RMSE	Test RMSE
Dummy Model	0.0659	0.0788	0.0917
Ridge Regression (w/o)	0.0545	0.0599	0.0459
Ridge Regression (w)	**0.0469**	**0.0521**	**0.0414**

Table 2: Comparison of RMSE values (mean ± standard error of the mean) for TCNNs trained with different sequence lengths across 10 independent runs.

Sequence Length	Train RMSE	Validation RMSE	Test RMSE
400	0.0229 ± 0.0003	0.0306 ± 0.0001	$\mathbf{0.0349 \pm 0.0001}$
500	$\mathbf{0.0229 \pm 0.0002}$	$\mathbf{0.0298 \pm 0.0000}$	0.0375 ± 0.0001
600	0.0244 ± 0.0004	0.0320 ± 0.0001	0.0364 ± 0.0003

3.4 Temporal Convolutional Networks

Training Strategy and Feature Engineering Our implementation of the TCN is based on the PyTorch-TCN library.[5] As in Section 3.1, we adopt a fixed train-validation-test partitioning strategy. The training set consists of 90% of the pre-earthquake data, while the validation set, comprising 10%, is used to monitor overfitting and select the best model based on validation loss. The test set includes post-earthquake data after a naive recovery period and is used solely to evaluate the model's generalization performance. Since TCNs process sequential data, raw instantaneous input features are used without rolling window transformations, and the sequence length determines how far back the model can analyze past data. To deal with the randomness of the training process, instead of relying on a single training run, 10 independent TCN models are trained for 25 epochs at each sequence length, and the mean and standard deviation of their RMSEs are reported. Based on findings from previous models, where rolling rainfall windows (450 hours for Ridge Regression) improved performance, we evaluate sequence lengths ranging from 400 to 600 time steps. To ensure that the network's receptive field spans the entire sequence length, we tune the TCN architecture by varying the kernel size and the number of hidden layers.

Results and Observations Table 2 shows the mean and standard deviation of the RMSE values across the 10 independent runs for different sequence lengths. While the 500 sequence length model achieves the best mean validation RMSE, we select the 400 sequence length model. The decision to select the latter model is based on the real-world nature of this case study. Unlike a specific benchmark dataset, real-world data often presents challenges, as discussed in Sec. 3.1, including issues with the representativeness of the validation set, therefore basing

[5] available at https://github.com/paul-krug/pytorch-tcn.

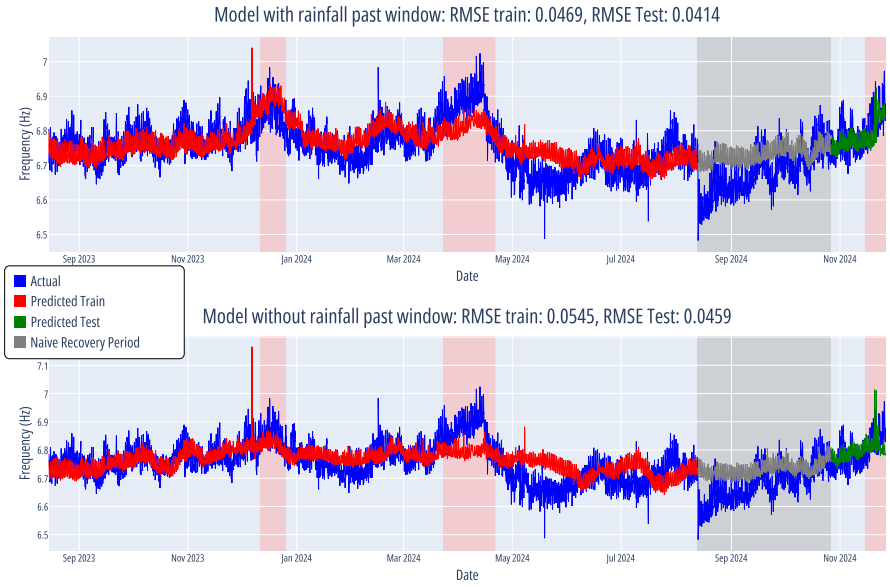


Fig. 4: (Top) Ridge Regression results **with** the rainfall past window as a feature, peaks are better modeled. (Bottom) Ridge Regression results **without** the rainfall past window as a feature, the model struggles to capture peaks and completely fails in capturing the peak in the test set. Poorly modeled peaks are highlighted in red.

model selection solely on validation performance may not be optimal. Instead, we choose the 400 sequence length model because it exhibits a more balanced behavior across training, validation, and test RMSE values. This suggests that it is likely to generalize better and be less prone to overfitting. The selected model is a TCN with seven layers (six with 16 channels and one with 8 channels), a kernel size of 6, dropout of 0.3, and residual connections. The output from the final TCN layer is processed through two fully connected layers ($8 \rightarrow 4 \rightarrow 1$) to generate the final prediction. We also show results of a Feedforward Neural Network (FFNN) with three fully connected layers (128, 64, and 1 neurons, respectively), batch normalization, LeakyReLU activation, and dropout, in which we performed feature engineering by adding a rolling window mean of past values for the rainfall feature. This choice was motivated by our observations from Ridge Regression, where incorporating such transformations improved the ability to model peak values. The results in Fig. 6 show that while the FFNN performs better than Ridge Regression in capturing peak values, it still struggles to accurately model the peak occurring between March and May. As a result, in certain intervals, the residuals during the normal period are almost as high as those observed during the anomalous period. In contrast, the TCN model provides a more stable and well-behaved solution. As seen in Fig. 6, the residuals are better

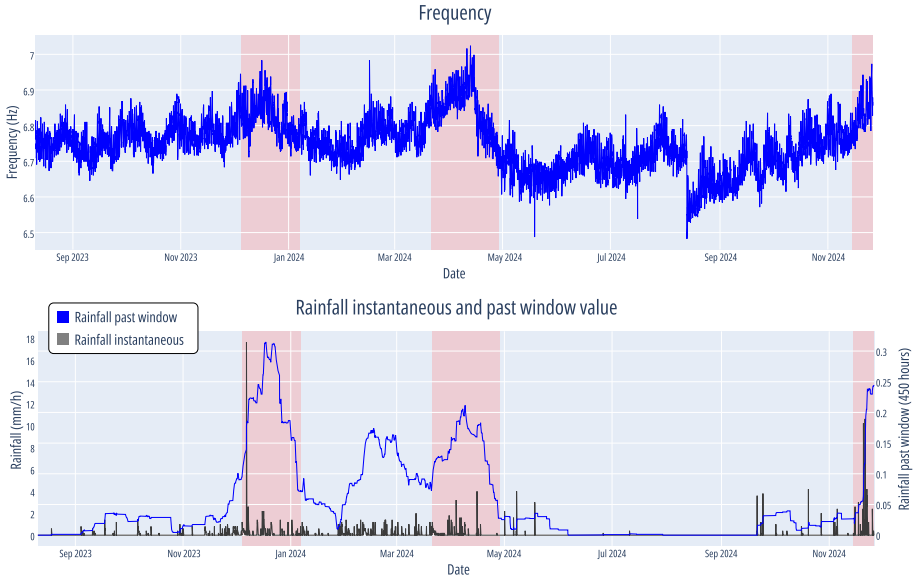


Fig. 5: (Top) Natural frequency time series with the three poorly modeled peaks highlighted. They correspond to periods where the model struggles to accurately predict the natural frequency. (Bottom) Instantaneous rainfall values (measured in millimeters over a one-hour period) and the past rainfall rolling window (450-hour window). The past window feature aligns well with these peaks, suggesting its effectiveness in capturing the cumulative effects of rainfall.

distributed, and the RMSE values further confirm its better performance. The other key advantage of TCNs is their ability to learn temporal dependencies directly from raw data, eliminating the need for manually engineered features like rolling window averages. Table 3 summarizes the comparison across different architectures, where the TCN achieves the best performance.

3.5 Structural Break Testing

Our primary goal is to automatically identify the end of the post-earthquake recovery period and refine the naive estimate of 2.5 months. To achieve this, we implement the BP sequential test presented in Section 2.3, by using the `dosequa` function provided in the R package `mbreaks` [22]. We perform the test using a 1% significance level, which means that only breakpoints with strong statistical evidence for a structural change are retained. The package requires a minimum segment length for reliable breakpoint detection. We set this value to 5% of the total dataset length, which is the smallest allowed by the package. We set a maximum of 10 breaks, which is the largest value allowed by the package.

Table 3: Comparison of RMSE values for the best-performing models across different architectures. For TCN, RMSE values of the best training run are reported. TCN achieves the best performance.

Model	**Train RMSE**	**Test RMSE**	**Overall RMSE**
Ridge Regression	0.0469	0.0414	0.0465
Feedforward Neural Network	0.0366	0.0394	0.0371
Temporal Convolutional Network	**0.0243**	**0.0340**	**0.0308**

Results and Observations Fig. 7a shows the result of our approach that combines the BP test with Deep Temporal Convolutional features, on the whole time series. The $F_T(l+1 \mid l)$ statistic (denoted supF) representing the gain in SSR obtained by splitting at a given break (see Sec. 2.3) is shown on the top. The test correctly identifies a structural break corresponding to the earthquake (Date 4) and two breaks after suggesting two important steps of the recovery process. Before the earthquake, three breaks are detected, however, the breaks at Date 1 and Date 2 have a much lower supF statistic compared to the others. The detection of Date 3 can be attributed to the fact that it marks the beginning of the validation set, where the model exhibits a higher RMSE compared to the training set. Finally, as explained in Sec. 2.3, we concatenate the portion of the time series between Date 3 and 4 of Fig. 7a (i.e., just before the earthquake)[6] with the portion after Date 6 (i.e., after the supposed recovery) and run the test on it, resulting in a non rejection of the null hypothesis (i.e., the recovery). The linear approach (see Fig. 7b), while correctly identifying the recovery period, detects the maximum allowed 10 breaks, many of which may not reflect actual structural changes. In contrast, our method avoids spurious breakpoints, providing a more reliable interpretation and reducing the risk of false positives, especially in less evident cases. These results demonstrate that our approach provides a more precise estimate of the recovery period, improving upon both the initial naive estimate and the results obtained with the linear approach.

4 Conclusions

This study combined machine learning and statistical methods to analyze the relationship between environmental factors and the natural frequency of a historical adobe church. By developing predictive models based on environmental parameters, we established a baseline response that enabled the detection of anomalies potentially linked to changes in the building's structure.

Our findings demonstrate that incorporating past weather data, particularly a rolling window for rainfall, significantly improves predictions. Ridge Regression served as a useful baseline, but it struggled to capture complex nonlinear

[6] We do not use previous segments since they belong to the training set and would induce a break due to the difference between training and test set as for Date 3.

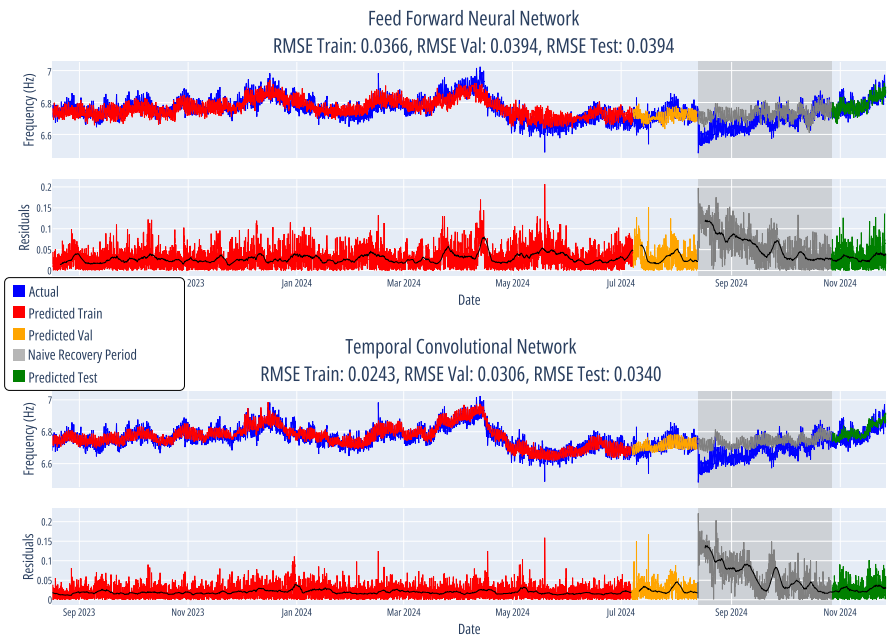
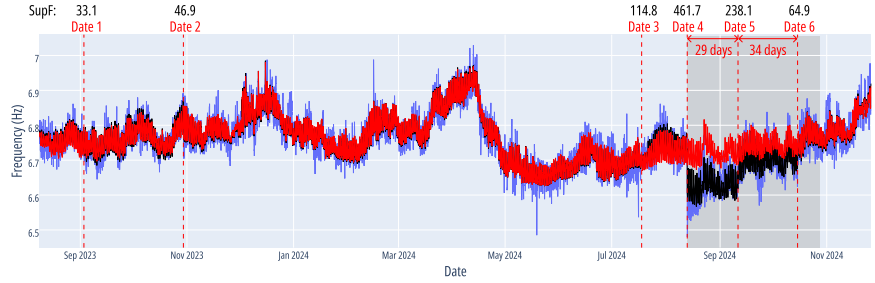


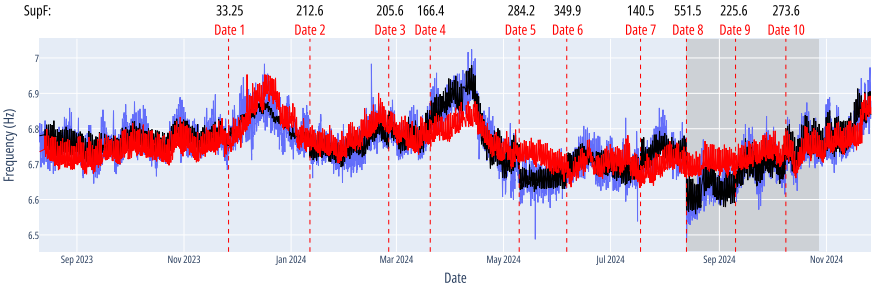
Fig. 6: (Top) Feedforward Neural Network results with rolling window feature engineering (window size 600). While peak modeling improves compared to Ridge Regression, the residuals remain problematic, suggesting further limitations. (Bottom) Temporal Convolutional Neural Network results of the best model across the 10 training runs with sequence length 400.

dependencies. In contrast, Temporal Convolutional Networks effectively modeled temporal dependencies, outperforming other approaches in predictive accuracy, as they could naturally take into account the time-delayed dynamic response of the church to changes in weather parameters. The achievement of a reliable baseline response model using exclusively weather data demonstrates that a large portion of the observed frequency variations can be explained by environmental conditions. Furthermore, our novel approach combining deep temporal convolutional features with the Bai-Perron test allowed us to refine the estimation of the post-seismic recovery period and to conclude that San Cristobal Church has completely recovered its dynamic behavior on October 15th, 63 days after the Mw 4.2 earthquake of August 13th, 2024.

By accurately assessing the post-seismic recovery time, our methodology provides a promising approach for developing early warning systems to detect damage in historical buildings. This capability is particularly valuable for heritage conservation, where non-invasive monitoring solutions are essential for preserving structural integrity. Heritage management is generally reactive, with interventions occurring only after damage has already taken place [14]. This approach is costly and, in some cases, too late to preserve the building's authenticity. The



(a) Result of the BP test using Deep Temporal Convolutional Features as input.



(b) Result of the BP test using Ridge Regression features as input (see Section 3.3).

Fig. 7: Structural Breaks: The black curve represents the predictions with one linear model per segment, the red one corresponds to predictions using a single linear model, and the blue one represents the actual values. In a), Date 4 corresponds to the earthquake, while Date 6 indicates the end of the recovery period. The grayed segment corresponds to the naive estimate.

ability to detect subtle changes, prior to damage, in the mechanical properties of masonry could improve risk assessment strategies and support more effective maintenance interventions.

References

1. Astorga, A., Guéguen, P.: Structural health building response induced by earthquakes: Material softening and recovery. Engineering Reports **2**(9), e12228 (2020)
2. Azzara, R.M., De Roeck, G., Girardi, M., Padovani, C., Pellegrini, D., Reynders, E.: The influence of environmental parameters on the dynamic behaviour of the san frediano bell tower in lucca. Engineering Structures **156**, 175–187 (2018)
3. Bai, J.: Estimating multiple breaks one at a time. Econometric theory **13**(3), 315–352 (1997)
4. Bai, J., Perron, P.: Estimating and testing linear models with multiple structural changes. Econometrica pp. 47–78 (1998)
5. Bai, J., Perron, P.: Computation and analysis of multiple structural change models. Journal of applied econometrics **18**(1), 1–22 (2003)

6. Bergmeir, C., Benítez, J.M.: On the use of cross-validation for time series predictor evaluation. Information Sciences **191**, 192–213 (2012)
7. Bodin, P., Vidale, J., Walsh, T., Çakir, R., Çelebi, M.: Transient and long-term changes in seismic response of the natural resources building, olympia, washington, due to earthquake shaking. Journal of Earthquake Engineering **16**(5), 607–622 (2012)
8. Brincker, R., Zhang, L., Andersen, P.: Modal identification of output-only systems using frequency domaindecomposition. Smart materials and structures **10**(3), 441 (2001)
9. Çelebi, M.: On the variation of fundamental frequency (period) of an undamaged building—a continuing discussion. In: International Conference on Experimental Vibration Analysis for Civil Engineering Structures. Porto (2007)
10. Cole Jr, H.A.: On-line failure detection and damping measurement of aerospace structures by random decrement signatures. Tech. rep., NASA (1973)
11. Combey, A., Mercerat, D.E., Gueguen, P., Langlais, M., Audin, L.: Postseismic survey of a historic masonry tower and monitoring of its dynamic behavior in the aftermath of le teil earthquake (ardèche, france). Bulletin of the Seismological Society of America **112**(2), 1101–1119 (2022)
12. Ewins, D.: Exciting vibrations: the role of testing in an era of supercomputers and uncertainties. Meccanica **51**(12), 3241–3258 (2016)
13. Falchi, F., Girardi, M., Gurioli, G., Messina, N., Padovani, C., Pellegrini, D.: Deep learning and structural health monitoring: Temporal fusion transformers for anomaly detection in masonry towers. Mechanical Systems and Signal Processing **215**, 111382 (2024)
14. Feilden, B.M.: Between two earthquakes: cultural property in seismic zones. Getty Publications (1987)
15. Gattulli, V., Lepidi, M., Potenza, F.: Dynamic testing and health monitoring of historic and modern civil structures in italy. Structural Monitoring and Maintenance **3**(1), 71–90 (2016)
16. Guéguen, P., Brossault, M.A., Roux, P., Singaucho, J.C.: Slow dynamics process observed in civil engineering structures to detect structural heterogeneities. Engineering Structures **202**, 109833 (2020)
17. Gueguen, P., Gallipoli, M.R., Navarro, M., Masi, A., Michel, C., Guillier, B., Karakostas, C., Lekidis, V., Mucciarelli, M., Ponzo, F., et al.: Testing buildings using ambient vibrations for earthquake engineering: a european review. In: Proceedings of the 2nd European conference on earthquake engineering and seismology (2ECEES), Istanbul (2014)
18. Hoerl, A.E., Kennard, R.W.: Ridge regression: Biased estimation for nonorthogonal problems. Technometrics **12**(1), 55–67 (1970)
19. Johnson, P., Sutin, A.: Slow dynamics and anomalous nonlinear fast dynamics in diverse solids. The Journal of the Acoustical Society of America **117**(1), 124–130 (2005)
20. Lea, C., Flynn, M.D., Vidal, R., Reiter, A., Hager, G.D.: Temporal convolutional networks for action segmentation and detection. In: proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 156–165 (2017)
21. Lorenzoni, F., Caldon, M., da Porto, F., Modena, C., Aoki, T.: Post-earthquake controls and damage detection through structural health monitoring: applications in l'aquila. Journal of Civil Structural Health Monitoring **8**, 217–236 (2018)
22. Nguyen, L., Yamamoto, Y., Perron, P.: mbreaks: Estimation and inference for structural breaks in linear regression models [computer software manual]. R package version **1**(0) (2023)

23. Peeters, B., De Roeck, G.: One-year monitoring of the z24-bridge: environmental effects versus damage events. Earthquake engineering & structural dynamics **30**(2), 149–171 (2001)
24. Peeters, B., Maeck, J., De Roeck, G.: Vibration-based damage detection in civil engineering: excitation sources and temperature effects. Smart materials and Structures **10**(3), 518 (2001)
25. Pellegrini, D., Padovani, C., Messina, N., Girardi, M., Carrara, F., Falchi, F.: Deep learning for structural health monitoring: An application to heritage structures. Materials Research Proceedings **26** (2022)
26. Van Den Abeele, K.A., Carmeliet, J., Johnson, P., Zinszner, B.: Influence of water saturation on the nonlinear elastic mesoscopic response in earth materials and the implications to the mechanism of nonlinearity. Journal of Geophysical Research: Solid Earth **107**(B6), ECV–4 (2002)
27. Vidal, F., Navarro, M., Aranda, C., Enomoto, T.: Changes in dynamic characteristics of lorca rc buildings from pre-and post-earthquake ambient vibration data. Bulletin of Earthquake Engineering **12**, 2095–2110 (2014)
28. Williams, E.F., Heaton, T.H., Zhan, Z., Lambert, V.R.: Variability in the natural frequencies of a nine-story concrete building from seconds to decades. The Seismic Record **2**(4), 237–247 (2022)
29. Zonno, G., Aguilar, R., Boroschek, R., Lourenço, P.B.: 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**, 109299 (2019)
30. Zonno, G., Aguilar, R., Boroschek, R., Lourenço, P.B.: Experimental analysis of the thermohygrometric effects on the dynamic behavior of adobe systems. Construction and Building Materials **208**, 158–174 (2019)