

INTEGRATION OF STAY LOAD CELL MEASUREMENTS AND DATA NORMALIZATION FOR STRUCTURAL HEALTH MONITORING OF CABLE-STAYED BRIDGES

S. Quarchioni¹, L. Tentella¹, V. Nicoletti¹, S. Carbonari¹ and F. Gara¹

¹ Università Politecnica delle Marche, Department of Construction, Civil Engineering and Architecture (DICEA)
Via Brece Bianche 12, Ancona, Italy
e-mail: s.quarchioni@pm.univpm.it

Keywords: Cable-stayed bridges, Data normalization, Modal tracking, Structural health monitoring, Load cells on stays, Sensor optimization.

Abstract. *This study presents an integrated and data driven monitoring framework for a cable-stayed bridge that integrates traditional accelerometric measurements with load cell data to achieve a comprehensive structural assessment. Accelerometers, load cells, a weather station, and thermocouples are integrated into the monitoring system to capture both structural and environmental parameters affecting the dynamic behaviour of the bridge. The work investigates how different data normalization approaches and environmental conditions affect the accuracy of damage detection. After a procedure of modal tracking, advanced methods for data normalization are applied, including neural network-based normalization procedures aimed at reducing the effects of temperature, wind, and traffic fluctuations. The normalized outputs serve as inputs to Hotelling control charts that are analysed to assess the sensitivity, stability, and reliability of each normalization approach. A comparative study is also conducted between control charts derived from load cell based modal frequencies and those obtained from accelerometric data, assessing their relative diagnostic capability under varying environmental conditions. Results show that the different normalization strategies significantly affect control chart robustness and their ability to discern environmental effects from potential structural anomalies. The proposed methodology presents a reliable basis for long term monitoring of cable-stayed bridges by using a limited but efficient sensor network.*

1 INTRODUCTION

Structural Health Monitoring (SHM) is a key tool for ensuring the safety and reliability of civil infrastructure. Cable-stayed bridges, in particular, are highly flexible systems whose dynamic response is strongly affected by environmental and operational conditions such as temperature, wind, and traffic. These influences can hide changes in modal parameters caused by structural damage, making detection difficult. Long-term vibration-based monitoring provides large datasets of modal properties (natural frequencies, damping ratios, and mode shapes). However, these parameters often vary due to environmental and operational conditions rather than damage [1]. To address this, normalization techniques are used to reduce environmental effects and extract damage-sensitive features. In addition to conventional acceleration measurements, this framework incorporates dynamic responses derived from load cells installed on stay cables. These signals are dynamically analysed using the same procedures of accelerometric data, allowing the monitoring strategy to extend beyond traditional vibration-based approaches with accelerometers. Details on the procedures for extracting modal parameters from load cells can be found in [2]. Various normalization methods have been proposed, ranging from linear statistical techniques to machine learning models. Multivariate linear regression and principal component analysis are widely used for their simplicity and interpretability, while neural networks (e.g., feedforward and autoassociative models) can capture more complex non-linear relationships. Despite these advances, there is no universally optimal method, and performance depends on the specific application. To support decision-making, multivariate statistical process control techniques, particularly the Hotelling T^2 control chart [3,4], are used as damage indicators. These methods enable the simultaneous monitoring of multiple correlated variables through a single index, improving sensitivity to structural behaviour changes. In this context, control charts are applied to the residuals between the frequencies predicted by the statistical model and the frequencies identified using both accelerometers and load cells, thereby providing a more comprehensive assessment of the structural behaviour and investigating the possibility of using load cells for dynamic monitoring together with the static one.

The present study proposes an integrated SHM framework that combines four normalization approaches (input-output and output-only, linear and non-linear) with a control chart strategy based on Hotelling T^2 , further applied to a real-case cable-stayed bridge. The main objective is to assess how the complexity of normalization models affects the entire monitoring process and the reliability of structural assessment within a statistical approach, using both dynamic modal parameters based on acceleration measurements and those based on load cells.

2 DESCRIPTION OF THE BRIDGE AND MONITORING SYSTEM

The ‘Filomena Delli Castelli’ cable-stayed bridge, located in central Italy, is 188.6 metres long and includes three spans; its steel-concrete composite deck accommodates two traffic lanes and, on the downstream side, a cycle path and a pavement. The deck, supported by 40 inclined stay cables, consists of a 25 cm thick concrete slab on two main steel I-shaped girders, connected by shear studs and transverse beams with lateral cantilevers. To enhance the seismic performance of the deck, Lead Rubber Bearings (LRBs) are installed at the supports of the two transverse beams located near the pylons. The pylons have a hollow circular steel section (1.9 m diameter), partially filled with concrete at the base, with total heights of about 33-36 m. The abutments are reinforced concrete walls with inclined wing walls and a retaining wall, while the deep foundations consist of 1.2 m diameter drilled piles. A picture of the bridge is presented in Figure 1a, while Figure 1b illustrates its side and plan views, including the main geometric dimensions.

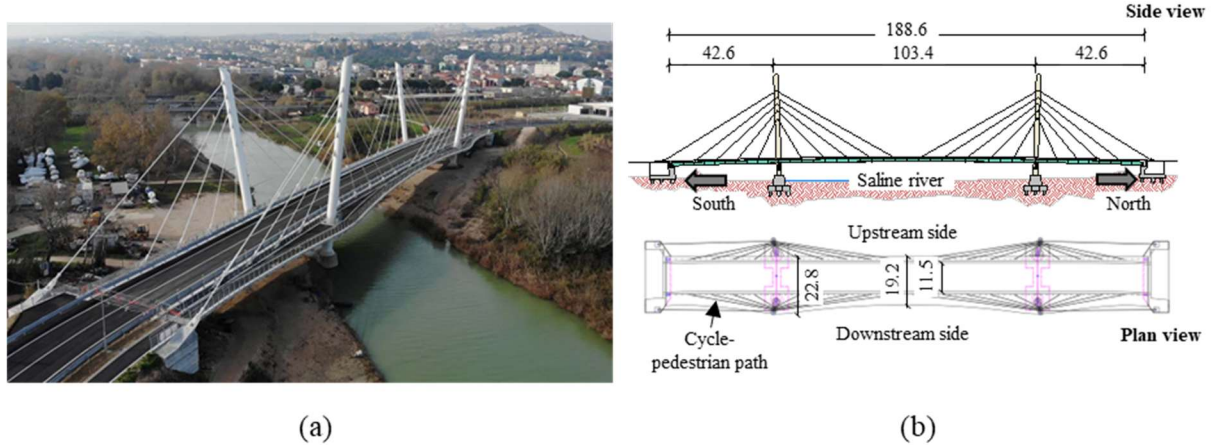


Figure 1: (a): aerial view of the bridge; (b) schematic view with main dimensions in elevation and plan.

The monitoring system includes sensors for static, dynamic, and environmental measurements. Static instrumentation includes load cells on eight stay cables and displacement transducers on each seismic isolator. Dynamic behaviour is identified using four deck-mounted accelerometers (one triaxial, one biaxial, two uniaxial), installed on only one half of the bridge due to structural symmetry. Additional sensors include a triaxial velocimeter, a triaxial MEMS accelerometer at the base of a pylon and another triaxial MEMS device on the ground to record seismic input. Environmental conditions are monitored through a weather station (temperature, wind, humidity) and five thermocouples placed on structural components and inside the acquisition unit. All structural sensors are sampled at 100 Hz, enabling both quasi-static and dynamic measurements, including seismic response. Environmental data are recorded once per minute. Data are continuously collected in 30-minute files, with trigger-based acquisition for seismic events, and automatically transmitted via FTP to a remote server, where real-time modal identification and dynamic analysis are performed. A schematic layout of the SHM system is presented in Figure 2.

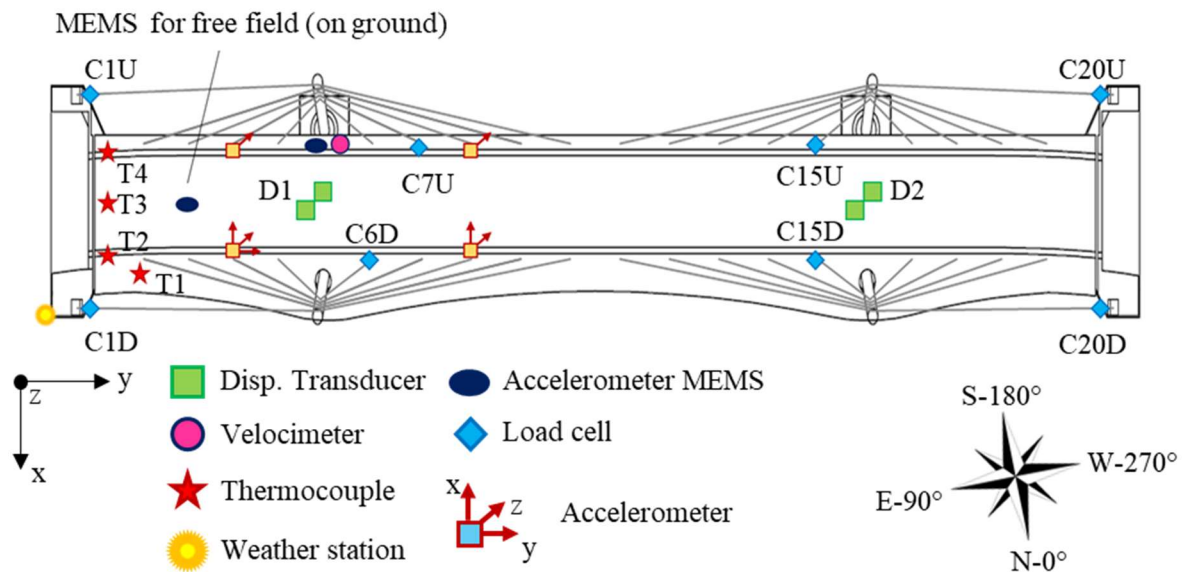


Figure 2: schematic layout of the SHM system.

3 DATA NORMALIZATION

3.1 Overview of data normalization techniques

Data normalization is a key step in dynamic SHM, as it reduces the influence of environmental and operational conditions on modal parameters. This allows changes due to structural damage to be distinguished from normal variability, improving detection reliability. The literature includes a wide range of approaches, from linear statistical methods to advanced artificial intelligence techniques. However, no general guidelines exist for selecting the most appropriate method, and their effectiveness largely depends on the specific application. For this reason, this study adopts a set of representative normalization methods to cover different scenarios related to environmental effects on modal parameters. Both input-output and output-only methods are selected, including linear and nonlinear approaches, as summarized in Table 1.

	Linear	Non-linear
Input-output	Multivariate Linear Regression (MLR)	Feedforward Neural Network (FNN)
Output-only	Principal Component Analysis (PCA)	Autoassociative Neural Network (ANN)

Table 1: Normalization methods for modal parameters identified during dynamic monitoring.

Multivariate Linear Regression (MLR) (Figure 3a) is a statistical technique used to model the relationship between one or more independent variables and multiple dependent variables. In this case, the dependent variables are the system frequencies f_i identified over time. The MLR model can be expressed as:

$$f_i(t) = \beta_0 + \sum_{j=1}^p \beta_j x_j(t) + e_i \quad (1)$$

for $i \in \{1, \dots, n\}$, where p is the number of independent variables, n is the number of dependent variables, β_0 is the intercept, β_j is the j -th regression coefficient, x_j is the j -th independent variable and e_i is a Gaussian error term.

Another linear technique for data normalization, which, however, takes into account only the outputs and thus the identified frequencies, is Principal Component Analysis (PCA). It is a linear transformation technique that converts an original set of correlated variables into a reduced set of uncorrelated variables, while preserving most of the information contained in the dataset [5]. In this study, PCA is applied to identify the principal components mainly influenced by environmental and operational parameters, which are then removed to obtain a dataset filtered from such effects. To perform PCA, all dependent variables are arranged in a matrix $\mathbf{F} \in \mathbb{R}^{m \times n}$, where m is the number of time observations and n is the number of modes. The mean value of each variable is computed and subtracted to obtain the centered matrix $\hat{\mathbf{F}}$. The covariance matrix is then evaluated as:

$$\mathbf{C} = \hat{\mathbf{F}}\hat{\mathbf{F}}^T \quad (2)$$

The principal component space is obtained through spectral decomposition:

$$\mathbf{C} = \mathbf{A}\mathbf{\Lambda}\mathbf{A}^T \quad (3)$$

where the columns of $\mathbf{A}$ are the eigenvectors and $\mathbf{\Lambda}$ is the diagonal matrix of eigenvalues. Each eigenvector represents a principal component, ordered according to its contribution to the total variance. The original variables can then be projected onto the principal component space:

$$\mathbf{Z} = \mathbf{F}\mathbf{A} \quad (4)$$

where $\mathbf{Z}$ is the matrix of transformed variables. By selecting the first k components that explain the majority of the variance (typically more than 75%), a reduced matrix $\mathbf{A}_{\text{rid}}$ is obtained so, the filtered dependent variables (frequencies) can then be reconstructed as:

$$\mathbf{F}_r = \mathbf{Z}\mathbf{A}_{\text{rid}}^T \quad (5)$$

Figure 3c illustrates a schematic representation of principal components in a three-dimensional space.

However, in recent years more advanced methods based on machine learning have been emerging. In this context artificial neural networks represent a major class of computational models in artificial intelligence, inspired by the structure of the human brain. Among them, the Feedforward Neural Network (FNN) is the simplest architecture and serves as the basis for more complex models. It is widely used in applications such as regression, classification, and pattern recognition.

A FNN consists of three main types of layers: an input layer, one or more hidden layers, and an output layer (Figure 3b). The input layer receives raw data, where each node corresponds to an independent variable [6]. Hidden layers process the inputs through weighted connections and linear or nonlinear activation functions, passing the results forward through the network. The learning process involves optimizing the weights to minimize the error between predicted and measured outputs. The final layer provides the model output, which in this case corresponds to the estimated modal frequencies. The network is trained using the Levenberg-Marquardt algorithm, a widely adopted method for solving nonlinear least-squares problems, particularly effective for regression tasks [7].

Another widely used output-only neural network is the Autoassociative Neural Networks (ANNs) that constitute a specific class of neural networks designed to reproduce their input at the output (Figure 3d). In this framework, the input vector coincides with the desired output. This structure enables the network to learn a compact representation of the data, which can be exploited for feature extraction and regression tasks. The network is trained to minimize the reconstruction error between input and output, effectively learning a compressed representation that retains the essential information required to reproduce the original data [8]. As in the FNN case, the learning parameters consist of the weights associated with nonlinear activation functions, which are optimized during the training phase.

To provide a more formal mathematical description, the FNN can be expressed as a sequence of nonlinear transformations across layers. For a network with L layers, the forward propagation can be written as:

$$\mathbf{x}^{(l)} = \sigma^{(l)} \left(\mathbf{W}^{(l)} \mathbf{y}^{(l-1)} + \mathbf{b}^{(l)} \right), l = 1, \dots, L \quad (6)$$

where $\mathbf{x}$ is the input vector (independent variables), $\mathbf{W}^{(l)}$ and $\mathbf{b}^{(l)}$ are the weight matrix and bias vector of the l -th layer, $\sigma^{(l)}(\cdot)$ is the activation function (e.g., sigmoid, ReLU, tanh), $\mathbf{y}^{(l)}$ is the output of the l -th layer. In the present application, $\mathbf{y}$ represents the estimated modal

frequencies. The training process aims to minimize a loss function J , typically the Mean Squared Error (MSE):

$$J = \frac{1}{N} \sum_{k=1}^N \| \mathbf{y}_k - \mathbf{t}_k \|^2 \quad (7)$$

where $\mathbf{t}_k$ are the target (measured) frequencies and N is the number of training samples.

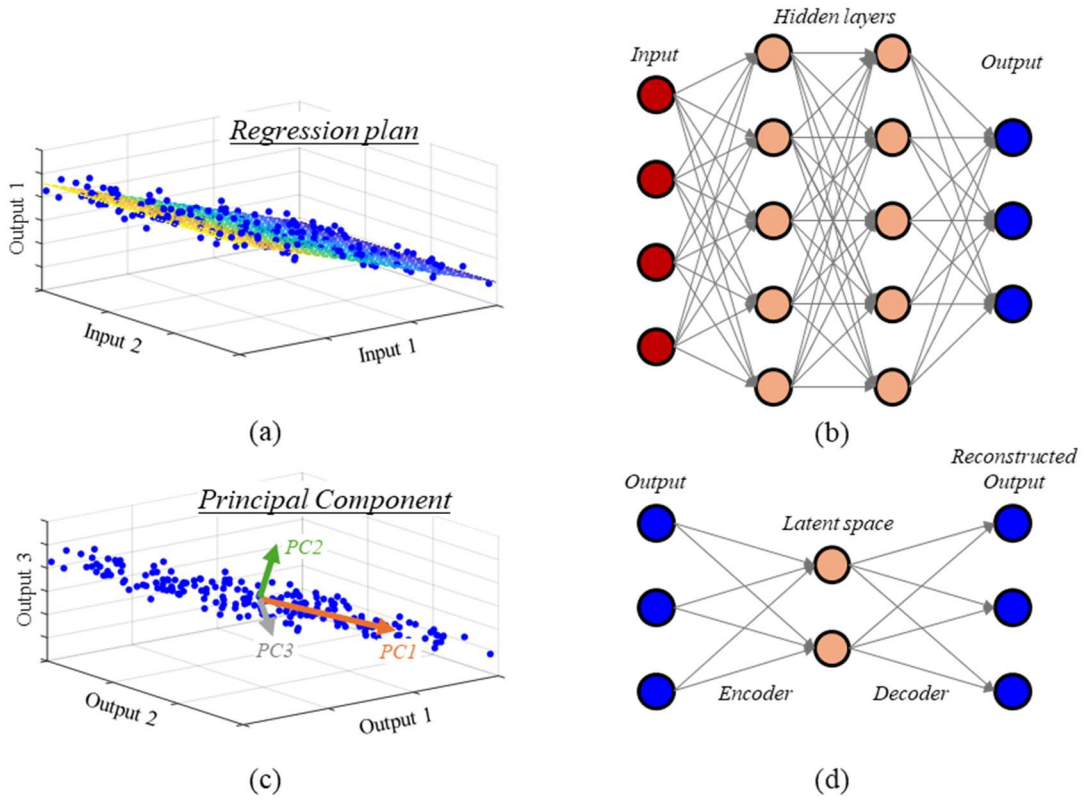


Figure 3: graphic representation of the four proposed normalization methods: (a) multiple linear regression; (b) feedforward neural network; (c) principal component analysis; (d) autoassociative neural networks.

For ANN, the mathematical formulation is similar, but with the constraint that the output must reproduce the input. The network can be decomposed into two parts: an encoder $\mathbf{z}$ and a decoder $\hat{\mathbf{x}}$:

$$\mathbf{z} = \sigma_e(\mathbf{W}_e \mathbf{x} + \mathbf{b}_e) \quad (8)$$

$$\hat{\mathbf{x}} = \sigma_d(\mathbf{W}_d \mathbf{z} + \mathbf{b}_d) \quad (9)$$

where $\mathbf{x}$ is the input vector (modal frequencies), $\mathbf{z}$ is the latent (compressed) representation, $\hat{\mathbf{x}}$ is the reconstructed input, $\mathbf{W}_e$, $\mathbf{W}_d$ and $\mathbf{b}_e$, $\mathbf{b}_d$ are weights and biases of encoder and decoder and σ_e , σ_d are activation functions. Also, in this case the objective function to be minimized is the reconstruction error J .

The effectiveness of each method depends on the correct selection of input variables and initial parameters. The selection of variables capable of adequately capturing environmental and operational effects is essential for accurately describing their influence on modal parameters; however, the choice of algorithm parameters is also crucial. For example, selecting a number of principal components in PCA that account for more than 75% of the variance, as well as

adjusting the epochs and learning rates in neural networks to ensure stable training and avoid overfitting. All normalization methods require the definition of a training period for the development of the statistical or machine learning model, during which the structure is assumed to be in an undamaged state. The training period is a selected interval of historical data used to calibrate the model parameters. During this phase, the model learns patterns from the data. In input-output approaches, both environmental/operational variables and identified modal frequencies are used, whereas in output-only approaches only the identified modal frequencies are considered. The model parameters are then optimized to minimize the error between predicted and observed values over the training interval. The training period must be sufficiently long to capture the full variability of modal frequencies induced by changing environmental and operational conditions. In practice, approximately one year of continuous monitoring data is commonly used, as this duration allows both short-term (daily and weekly) fluctuations and long-term (seasonal) effects to be properly represented.

3.2 Application to the case study

The four normalization methods are applied to the monitored bridge to assess the capability of both linear and nonlinear approaches to reproduce its dynamic behaviour under varying environmental and operational conditions. In the adopted input-output methods, air temperature, the root mean square (RMS) of accelerations (introduced as a representation for traffic-induced excitation), and wind speed and direction are selected as independent variables, as they represent the main sources of variability affecting the structural response. The first year of monitoring data (September 2022 – August 2023) is used as the training period, ensuring that the models are calibrated over a complete seasonal cycle and therefore exposed to the full range of environmental conditions. Figure 4 reports the probability density functions (PDFs) of the natural frequencies associated with the identified vibration modes during this period. Specifically, 12 modes are clearly identified from accelerometric measurements (Figure 4a) and subsequently adopted for dynamic monitoring, while 10 modes are extracted from load cell data (Figure 4b). Further details on the selected modes, along with a comparison between the results identified by the load cells and those identified by the accelerometers, are available in [2]. In Figure 4, the experimentally identified modal frequencies are represented by solid lines and are directly compared with those predicted by the four statistical models. The strong overlap between the corresponding PDFs indicates that the models are able to accurately capture both the mean and the standard deviation of the frequencies for each mode. This agreement demonstrates that the influence of environmental and operational variables has been effectively modeled, and that both linear and nonlinear approaches provide a reliable representation of the bridge's dynamic behaviour. On this basis, the calibrated models can be confidently employed to define reference conditions and to construct control charts for structural health monitoring purposes. Such charts enable the detection of statistically significant deviations from the baseline, which may be associated with potential structural changes or damage.

4 STRUCTURAL MONITORING USING STATISTICAL PATTERN RECOGNITION APPROACH

Once the statistical model has been defined, it serves as a predictive tool capable of estimating the modal frequencies of the monitored structure filtered from environmental and operational effects. The subsequent step consists in evaluating the residual R , defined as the difference between the model-predicted frequencies F_{mod} and the experimentally identified ones F_{id} :

$$R = F_{\text{mod}} - F_{\text{id}} \quad (10)$$

Residuals close to zero indicate that the model effectively reproduces the structural dynamics and the variability due to environmental and operational influences. In contrast, large residuals suggest a deviation from the baseline behaviour identified during the training period. Such deviations may be associated with either planned actions (e.g., maintenance or retrofiting) or unplanned changes (e.g., damage or deterioration). In the literature on SHM, multivariate control charts are widely used to analyse such residuals, with the Hotelling T^2 chart being one of the most common approaches. This method enables the simultaneous control of multiple correlated variables, such as residuals between predicted and experimentally identified modal frequencies, by summarising them into a single statistical indicator for each normalization method.

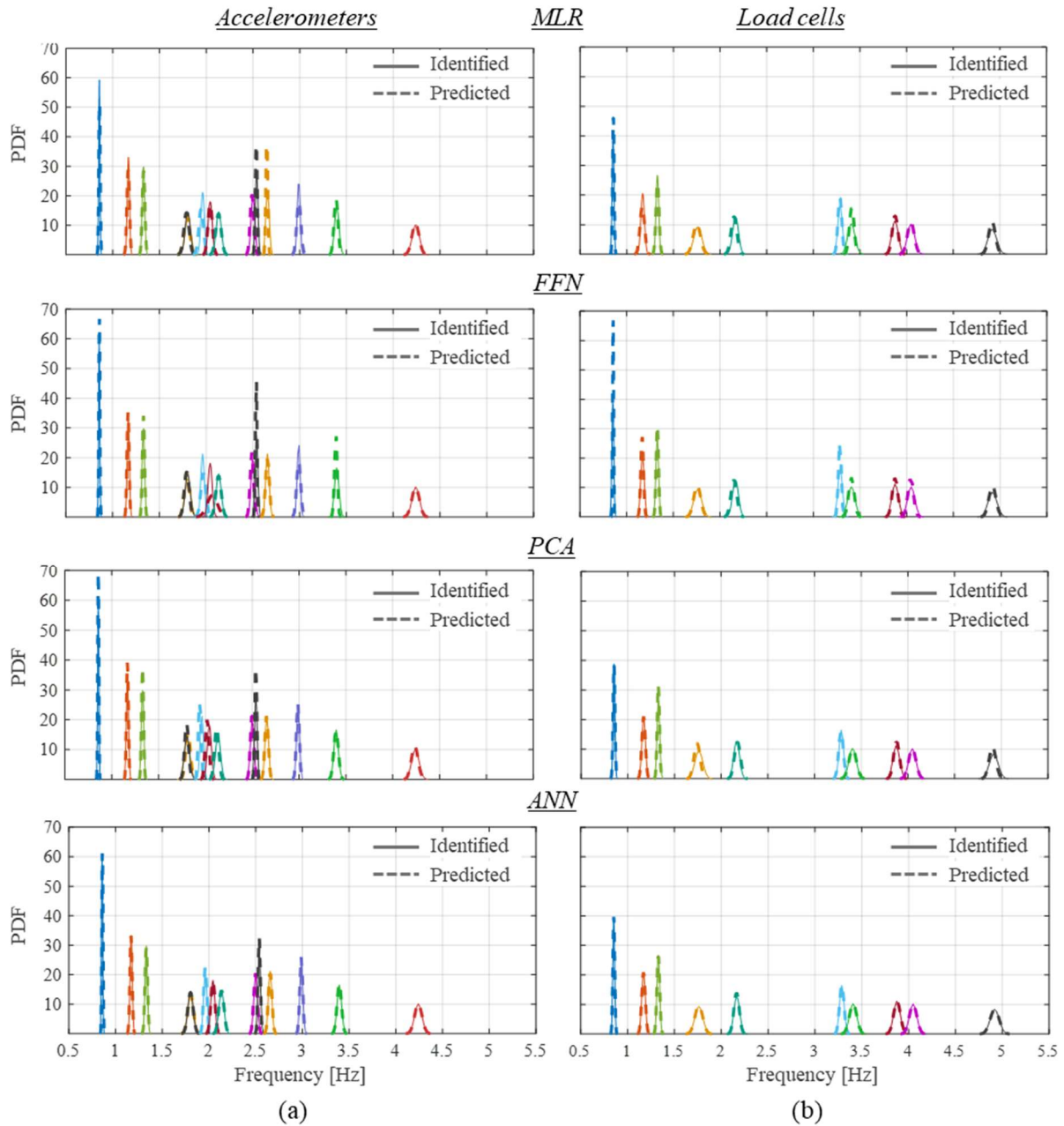


Figure 4: probability density functions of the frequencies of the monitored modes superimposed on those predicted by statistical models: (a) derived from accelerometers; (b) derived from load cells.

For each observation (or set of observations, e.g., dynamic identifications over a given time window), the statistical indicator is computed as follows:

$$T_i^2 = n(\bar{\mathbf{r}}_i - \bar{\mathbf{r}})\mathbf{S}^{-1}(\bar{\mathbf{r}}_i - \bar{\mathbf{r}})^T \quad (11)$$

where $\bar{\mathbf{r}}$ is the mean residual vector over the training period, $\bar{\mathbf{r}}_i$ is the mean residual vector for the i -th subgroup (time window), and $\mathbf{S}$ is the covariance matrix estimated from the training data. The choice of subgroup size n is critical, as it directly influences the reliability of the estimated statistical parameters. Each subgroup should contain a sufficient number of observations achieved under comparable environmental and operational conditions to ensure a stable estimation of the covariance matrix. A common empirical guideline is $n = k \cdot p$, where p is the number of variables and k typically ranges between 5 and 10; however, this rule should be regarded as indicative and verified against data quality and physical consistency of results.

Finally, a Control Limit (CL) is defined as a threshold corresponding to a selected percentile of the T^2 values computed during the training phase, commonly in the range of 95–98%; in this study, a 98% threshold is adopted. Exceedances of the CL indicate potential anomalies. However, isolated or marginal exceedances comparable to those observed during the training period may reflect natural variability, whereas sustained or significantly higher values are more likely associated with meaningful changes in structural behaviour.

4.1 Control charts based on frequencies identified with accelerometer measurements

The Hotelling T^2 control chart is computed for the bridge using all four normalization methods, resulting in different sets of residuals. Including all methods allows the assessment of how a suboptimal normalization strategy may influence the robustness, accuracy, and reliability of the resulting control chart.

A daily aggregation is adopted, with each subgroup containing all identifications performed within a 24-hour period. This choice provides one consistent SHM update per day, yielding a stable and interpretable control point for the monitoring of the bridge.

An analysis of the control charts represented in Figure 5 for all four methods reveals, most clearly in the input-output methods, that the dynamic behavior of the bridge between December 2023 and January 2026 has deviated from the statistical model constructed during the training period. This could be due to a change in the influence of environmental parameters on dynamic behavior, perhaps associated with variations on the boundary conditions, with no other signs of damage observed, either on the load cells or on the displacement transducers at the supports. It is clear that, in methods based exclusively on output (PCA and ANN), the T^2 values are slightly above the threshold following the change in dynamic behavior, so this analysis leads to the conclusion that output-only methods are less effective at identifying small variations in the structure's properties, even though they provide better results in terms of frequency normalization. This is probably because the variation in the structure's properties from December 2023 onwards is so small that the output-only statistical model regards it as a fluctuation already experienced during the training period and to which, however, it cannot associate any input parameter (for example, temperature).

4.2 Control charts based on frequencies identified with load cell measurements

In addition to the accelerometer-based monitoring, Hotelling T^2 control charts are also developed using dynamic responses identified by load cells installed on the stay cables.

As in the accelerometer-based case, all four normalization methods are applied to the load cell-based results to generate corresponding residual datasets. This approach enables a

consistent comparison of how different normalization strategies influence the statistical behaviour of the control charts when applied to heterogeneous dynamic measurements.

A daily aggregation is still adopted for constructing the control charts, providing a single consistent SHM update per day.

This formulation allows the assessment of whether load cell measurements can provide consistent indications of structural condition changes, and how their statistical behaviour compares with that obtained from accelerometer-based frequencies residuals under different normalization methods. The control charts are shown in Figure 6. A change is also observed in the load cells, and also in this case, the PCA and ANN models mask this phenomenon to a greater extent. This demonstrates how it is possible to assess changes in the bridge's dynamic behaviour using load cells that measure data in real time, thereby enabling both static (load on the stay cables) and dynamic monitoring using load cells only.

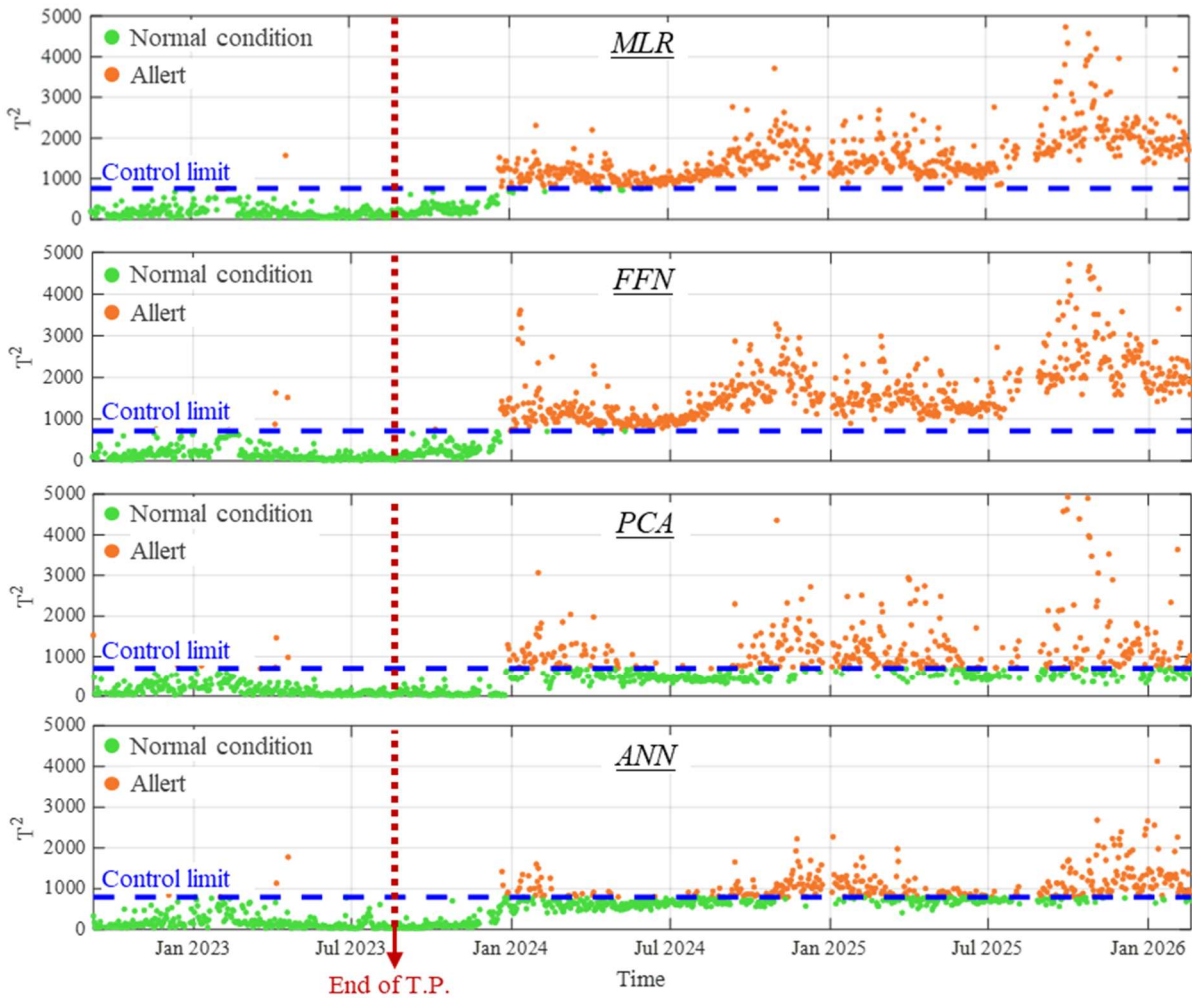


Figure 5: Hotelling T^2 control charts based on all four normalization methods for dynamic monitoring using accelerometers.

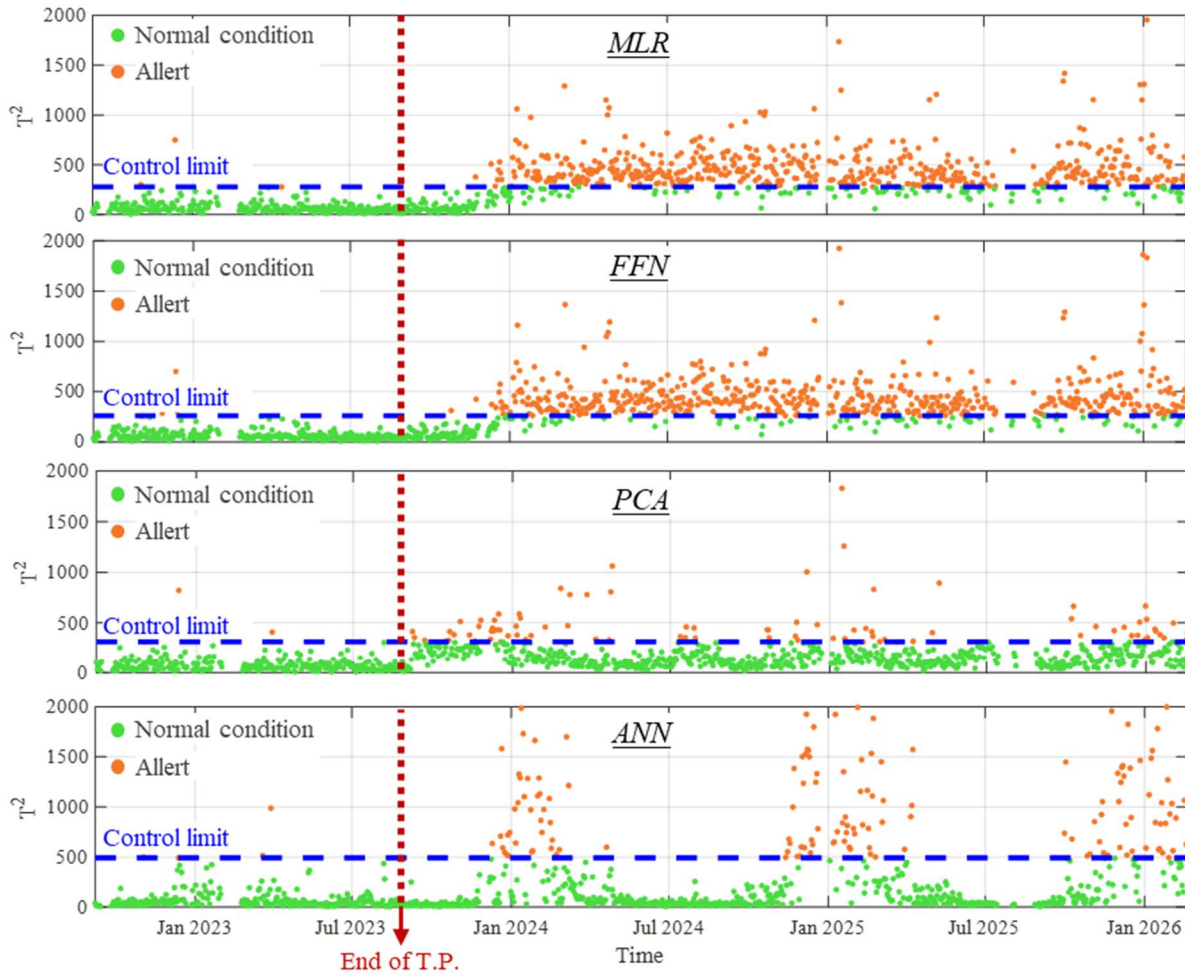


Figure 6: Hotelling T^2 control charts based on all four normalization methods for dynamic monitoring using load cells.

5 CONCLUSIONS

This study presented a SHM framework for a cable-stayed bridge based on the long-term monitoring of modal parameters and four frequencies normalization approaches, including linear statistical methods and non-linear machine learning methods, within input-output and output-only formulations. Multivariate regression, PCA filtering and neural network models (feedforward and autoassociative) are used to reduce environmental and operational effects with varying degrees of effectiveness. The residuals between the predicted and identified frequencies are integrated into a Hotelling T^2 control chart for continuous monitoring. The study demonstrated how the normalization strategy influences the quality of the residuals used for the extraction of damage-sensitive features. Furthermore, it shown how the well-established statistical pattern recognition approach can capture even small variations in dynamic behaviour of the bridge. In addition, this approach is also applied to the results of dynamic identification obtained from load cells, demonstrating how load cells can also be utilised for dynamic monitoring, thereby leveraging dual information (load on stay cables and modal parameters of the bridge) increasing the efficiency of the monitoring system.

REFERENCES

- [1] García-Macías E., Ubertini F.. “MOVA/MOSS: Two integrated software solutions for comprehensive Structural Health Monitoring of structures”. In: *Mechanical Systems and Signal Processing*, Vol. 143, (2020). DOI: 10.1016/j.ymssp.2020.106830.
- [2] Quarchioni, S., Nicoletti, V., Carbonari, S., Gara, F.. “Mode Shape Reconstruction Using Operational Modal Analysis of Stay Cables Load Cells Data”. In: *Cunha, Á., Caetano, E. (eds) Experimental Vibration Analysis for Civil Engineering Structures. EVACES 2025. Lecture Notes in Civil Engineering*, vol. 675. Springer, Cham. DOI: 10.1007/978-3-031-96106-9_23.
- [3] Tiriyaki S., Aydin A. “Multivariate Hotelling T2 Control Chart for Monitoring Some Quality Characteristics in Medium Density Fiberboard Manufacturing”. In: *Drvna industrija*, 73, 1, (2022). DOI: 10.5552/drvid.2022.2046.
- [4] Tracy N., Young J., Mason R. “Multivariate Control Charts for Individual Observations”. In: *Journal of Quality Technology*, 24, 2, (1992), 88-95. DOI: 10.1080/00224065.1992.12015232.
- [5] Tharwat, A. “Principal component analysis - a tutorial”. In: *International Journal of Applied Pattern Recognition*, Vol. 3, No. 3, (2016), pp. 1-41. DOI: 10.1504/IJAPR.2016.079733. DOI: 10.1504/IJAPR.2016.079733.
- [6] Sazlı, M. H. “A brief review of feed-forward neural networks”. In: *Communications Faculty of Sciences University of Ankara Series A2-A3 Physical Sciences and Engineering*, 50(01). (2006). DOI: 10.1501/commua1-2_0000000026.
- [7] Lim, J., Sung, K. “FNN (Feedforward Neural Network) Training Method Based on Robust Recursive Least Square Method”. In: *Lecture Notes in Computer Science*, vol. 4492. Springer, Berlin, Heidelberg (2007). DOI: 10.1007/978-3-540-72393-6_48.
- [8] Kramer, M.A. “Autoassociative neural networks”. In: *Computers & Chemical Engineering*, 16, 4, (1992), pp. 313-328. DOI: 10.1016/0098-1354(92)80051-A.