13th International Conference on
Structural Health Monitoring of Intelligent Infrastructure
DOI: 10.3217/978-3-99161-057-1-143

Setting an optimal threshold for novelty detection in data-driven Structural Health Monitoring

Alessio De Corso[1, 2], Carlo Rainieri[2, 0000-0003-4854-0850]
[1]Polytechnic University of Bari, via E. Orabona 4, 70125 Bari, Italy
[2]Construction Technologies Institute, National Research Council of Italy, corso N. Protopisani, 80146 Naples, Italy
email: a.decorso@phd.poliba.it, rainieri@itc.cnr.it

ABSTRACT: The data-driven approach to vibration-based Structural Health Monitoring aims to detect anomalies in the monitored modal properties. A key step in this framework is compensating for the normal variability in the data, which is due to the strong influence of environmental and operational variables on the structure's dynamic behavior. The decision-making process is then formulated as a binary classification problem, supported by an appropriate alarm threshold to distinguish between normal and anomalous structural conditions. The threshold is typically set on the statistical distribution of the novelty index computed during the training phase, often assuming a Gaussian distribution of the data. However, anomaly detection requires a more refined modeling of the distribution tails. The present paper investigates the use of Extreme Value Theory for threshold setting, focusing on the Block Maxima sampling technique and the Generalized Extreme Value distribution. A comparison with conventional approaches demonstrates the significant accuracy achievable through the extreme value theory. The natural frequency time histories of the KW51 bridge are used as benchmark data to highlight the method's effectiveness in improving the reliability of early damage detection.

KEY WORDS: Data-driven Structural Health Monitoring; Novelty detection; Threshold setting; Environmental and operational variables; Extreme Value Theory.

1 INTRODUCTION

The remote and automated evaluation of the structural conditions through Structural Health Monitoring (SHM) is crucial in the modern management of civil engineering assets. SHM enables a transition from the traditional scheduled maintenance approaches to proactive strategies that exploit the early damage identification, thereby enhancing safety and reducing long-term maintenance costs [1]. In the context of SHM, damage detection is the first step of damage identification, and it is commonly approached through data-driven methodologies. In this framework, the damage detection problem is cast as a novelty detection one [2], [3]. This strategy involves extracting damage-sensitive features (DSFs) from sensor data through automatic Operational Modal Analysis (OMA) techniques, which are further analyzed to detect deviations from a baseline condition. As such, damage detection is framed as a binary classification problem, aiming to distinguish the anomalous structural behavior, caused by either progressive degradation or sudden events, from the normal operating state [4].

In vibration-based SHM the modal properties or other related parameters are often selected as DSFs. Several applications reported in the literature consider the natural frequencies as DSFs because they can be easily obtained from measurements of the ambient vibration response of structures by a few, appropriately located sensors. Even if natural frequencies are relatively easy to monitor and informative for the first level damage detection, they are also very sensitive to the influence of environmental and operational variables (EOVs), such as temperature changes over time. An accurate damage detection therefore requires the application of appropriate compensation techniques to isolate the changes in the structural behavior due to damage or degradation phenomena from environmental and operational effects on the selected DSFs [5], and, as a consequence, enhance the reliability of the SHM outcomes. Such a compensation relies on setting data normalization models developed with reference to data collected in a training phase.

After the data normalization stage, the DSFs are transformed into novelty indexes (NIs), which are scalar indicators used to quantify how much a given observation deviates from the expected behavior. In order to assess whether the observed structural response should be considered anomalous, appropriate threshold values must be set, so that if the NIs overcome the threshold a warning can be issued. This is, therefore, another key step in the implementation of reliable modal based SHM strategies, in addition to the previously mentioned compensation of environmental and operational influence on DSFs (Figure 1).

A critical aspect in threshold setting is related to the need of defining it in an unsupervised context, that is to say, by using only data from the reference (nominally, healthy) condition of the structure. In the common practice, a Gaussian distribution for NIs is often assumed for the sake of threshold setting. However, this assumption is frequently inadequate for the novelty detection tasks [1]. As a result, setting the threshold based on a predefined data distribution can be misleading. Moreover, this approach does not take into account that detecting rare, extreme deviations is the focus of any SHM strategy, and, as such, an appropriate data-driven threshold setting approach should rely on the careful analysis of the tails of NI distribution.

DOI: 10.3217/978-3-99161-057-1-143

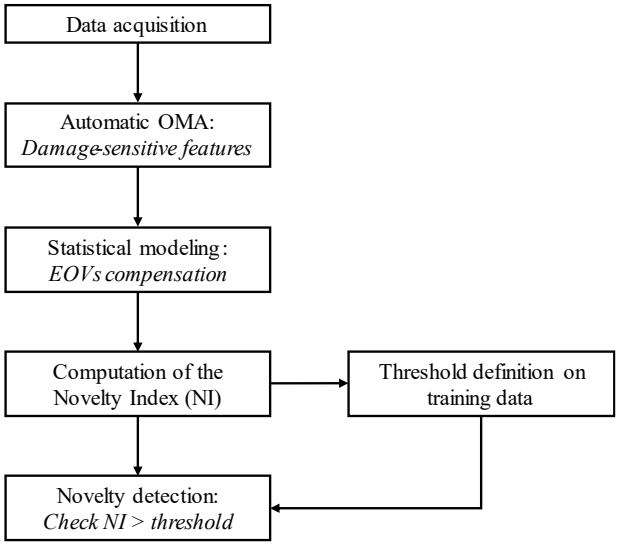


Figure 1. Flowchart of a typical data-driven SHM framework.

The Extreme Value Theory (EVT) represents a suitable alternative to the approaches based on the assumption of a Gaussian data distribution, as it focuses on the tail of the distribution – where anomalies are most likely to occur – thereby enabling a more precise threshold estimation [6].
This paper discusses the problem of the appropriate alarm threshold setting in the context of modal-based damage detection. A comparative assessment of different threshold setting strategies is presented by processing a benchmark dataset available in the literature. The analysis starts from the computation of NIs from natural frequency time histories through the combination of Gaussian Mixture Model (GMM) and Mahalanobis Squared Distance (MSD) to mitigate the influence of EOVs [7]. Afterwards, an EVT-based approach is applied for threshold setting. It resorts on the Block Maxima (BM) method to identify extreme observations and to model them according to the Generalized Extreme Value (GEV) distribution [8]. The effectiveness of the proposed approach in enhancing the robustness of novelty detection in modal-based SHM systems is demonstrated through quantitative comparisons with standard threshold setting methods.

The paper is structured as follows: after the introduction, Section 2 describes the methodological framework, detailing both the compensation strategy for data normalization with respect to the EOV influence and the computation of the NI time series. Moreover, Section 2 also outlines the EVT-based procedure for threshold setting. Section 3 presents the applicative case study and the characteristics of the benchmark dataset, followed by the analysis and discussion of results. The key findings of the study are finally summarized in the conclusions.

2 METHODOLOGY

2.1 Multivariate modeling of damage sensitive features under EOV influences

The present section describes the approach adopted for the compensation of EOV effects on DSFs and the computation of the NI. In this context, the matrix $\mathbf{X} \in \mathbb{R}^{n \times m}$ represents the training dataset, holding n observations of the natural frequencies of m vibration modes. These are experimentally collected under varying environmental and operational conditions at the beginning of the monitoring period or, more generally, in a reference monitoring period.

GMM is herein applied to represent $\mathbf{X}$ as a finite mixture of multivariate Gaussian distributions. The objective of this data processing stage is the effective modeling of the dominant feature clusters associated with the reference states of the monitored structure. The mixture density function is formally defined as:

$$f_{mix}(\boldsymbol{x}) = \sum_{q=1}^{Q} \eta_q f_q\left(\boldsymbol{x}|\boldsymbol{\mu}_q, \boldsymbol{\Sigma}_q\right) \quad (1)$$

Here, $f_q\left(\boldsymbol{x}|\boldsymbol{\mu}_q, \boldsymbol{\Sigma}_q\right)$ denotes the multivariate Gaussian probability density function of the q-th component, fully characterized by the mean vector $\boldsymbol{\mu}_q$, the covariance matrix $\boldsymbol{\Sigma}_q$, and the mixture weight η_q. The model parameters are obtained by the Maximum Likelihood Estimation (MLE) method, where the maximization of the likelihood function is achieved by the Expectation-Maximization (EM) algorithm [9].

The optimal number of components Q in Equation (1) is determined by minimizing the Bayesian Information Criterion (BIC), a standard model selection metric that penalizes model complexity to prevent overfitting [7]. This probabilistic framework supports the implementation of a robust anomaly detection methodology by leveraging the different components of the mixture model while inherently accounting for the influence of EOVs.

In order to obtain the NI time series for anomaly detection, the MSD is adopted as a multivariate metric to measure the distance between the observed DSFs and the GMM components. It incorporates both variable scales and correlations [10], and, given the generic test observation $\mathbf{z}$, its MSD relative to each GMM component can be computed as follows:

$$MSD_q(\boldsymbol{z}) = \left(\boldsymbol{z} - \boldsymbol{\mu}_q\right)\boldsymbol{\Sigma}_q^{-1}\left(\boldsymbol{z} - \boldsymbol{\mu}_q\right)^T \quad (2)$$

where $\boldsymbol{\mu}_q$ and $\boldsymbol{\Sigma}_q$ denote the mean vector and covariance matrix of the q-th GMM component, respectively. The NI corresponding to the generic test observation $\mathbf{z}$ is then given by the minimum distance across all components:

$$NI(z) = \min\left\{MSD_q(\boldsymbol{z})\right\} \quad (3)$$

Following the above-described approach, if a new observation is consistent with the reference structural condition, it will be close to one of the GMM components computed in the reference training period, and it will yield a low NI value. Conversely, if the structure has transitioned to a damaged state, the new observations will significantly diverge from all the GMM components in the training stage, resulting in larger NI values with respect to the undamaged condition.

2.2 Threshold setting methods for Novelty Detection

In the context of novelty detection, the EVT-based approach provides a robust statistical framework for threshold determination. This method is grounded in the theorem stating that the distribution of extreme values can converge only to one of three canonical forms: Gumbel, Weibull, or Fréchet distributions. To simplify the process, the GEV distribution is

DOI: 10.3217/978-3-99161-057-1-143

employed, as it unifies all three types within a single parametric family. The GEV distribution is expressed as follows [8]:

$$G(Y) = exp\left\{-\left[1+\xi\left(\frac{Y-\lambda}{\sigma}\right)\right]^{-\frac{1}{\xi}}\right\} \quad (4)$$

where ξ, σ and λ are the shape, scale, and location parameters, respectively. It is defined on the set of maxima Y, satisfying the condition $1+\xi(Y-\lambda)/\sigma > 0$, with λ and ξ real-valued parameters, and $\sigma > 0$. The unknown parameters are estimated by using the MLE method.

In order to define the population of extreme values to be fitted by the GEV distribution, the BM method is employed. Thus, the NI time series in the training period is divided into non-overlapping blocks of equal length, and the maximum value is selected in each block. Assuming that the structure is initially undamaged (null hypothesis), and selecting a significance level α, the threshold can be defined as the corresponding quantile of the fitted distribution [8]:

$$t = \begin{cases} \lambda - \frac{\sigma}{\xi}\left[1-\{-log(1-\alpha)\}^{-\xi}\right], & \xi \neq 0 \\ \lambda - \sigma log\{-log(1-\alpha)\}, & \xi = 0 \end{cases} \quad (5)$$

In addition to the previously described approach based on EVT, in this study also a more conventional method for threshold setting is considered for the purpose of comparative assessment. It consists in setting a predefined False Alarm Rate (FAR), interpreted as the tolerable proportion of false alarms in the training data. The threshold t is then calculated as the cut-off value that satisfies [11]:

$$\frac{number\ of\ false\ alarms\ (t)}{n} = FAR \quad (6)$$

Another widely adopted strategy for threshold setting is based on the assumption of Gaussian distribution of the DSFs and the use of MSD as the novelty index. Under these assumptions, the NI follows a χ^2 distribution, and the threshold can be directly obtained as the quantile of the distribution at the significance level α [7].

For the sake of the comparative performance assessment of the considered threshold setting approaches, the following parameters are computed: the number of false positives (FPs), the number of false negatives (FNs), and the Youden index. The latter, derived from the Receiver Operating Characteristic (ROC) curve, provides a measure of the balance between sensitivity (the true positive rate) and specificity (the true negative rate) and results in a single value that reflects the overall diagnostic performance of the considered approach [12].

3 ANALYSIS OF THE BENCHMARK DATASET

The data collected from the SHM system installed on the KW51 bridge in Belgium are processed for the objectives of the present study. The KW51 bridge is a steel bowstring railway bridge located in Leuven, Belgium (Figure 2). A detailed 15-month monitoring program was conducted between late 2018 and early 2020 to capture the dynamic behavior of the bridge under operational conditions [13]. Acceleration data collected during this period were processed using Operational Modal Analysis techniques [14], allowing the identification of the first 14 natural frequencies of the bridge.

Figure 2. KW51 bridge in Leuven, Belgium [13].

During the monitoring period, the bridge was retrofitted to correct a construction defect identified during inspection. The intervention involved strengthening the connection between the diagonals, arches, and bridge deck by welding a steel box around each original bolted joint. Specifically, the bridge was monitored before the retrofit intervention between October 2nd, 2018, and May 15th, 2019, and after that in the period between September 27th, 2019, and January 15th, 2020.

In the application of the data processing and threshold setting approaches described in Section 2, the natural frequency time series corresponding to modes 1, 2, 7, 8, 10, 12, and 14 were excluded from the analysis due to significant data gaps that prevented successful monitoring of these modes. This was made in agreement with similar considerations reported in [13]. Thus, only the natural frequency time series of modes 3, 4, 5, 6, 9, 11, and 13 were considered for the present analysis. Minor data gaps in these time series were filled by linear interpolation.

In addition to mode selection, since the retrofit intervention introduced a significant shift in the considered natural frequency time histories, the difference in the average value of corresponding natural frequencies before and after the intervention was intentionally reduced to make more challenging the appropriate alarm threshold setting as a result of a reduced change in the observed structural behavior.

Specifically, the frequency scatter was scaled down to 25% of its original value to avoid a straightforward or even trivial novelty detection (Table 1).

Table 1. Reduced scatter between average frequency before and after the retrofitting.

Mode	Frequency scatter	
	Original	Reduced
3	0.53%	0.13%
4	1.17%	0.29%
5	0.33%	0.08%
6	2.07%	0.52%
9	0.76%	0.19%
11	2.02%	0.51%
13	1.44%	0.36%

DOI: 10.3217/978-3-99161-057-1-143

Indeed, without this correction, the frequency shift due to retrofit results in an overly obvious differentiation between pre- and post-intervention states, thereby undermining the relevance of the novelty detection process and threshold setting procedure.

A detailed inspection of the collected natural frequency time histories also reveals some sharp increases in frequency values in the first monitoring period (Figure 3). A detailed investigation about the occurrence of these particular patterns is reported in [13], where the correlation between the natural frequencies of the bridge and the measured temperature has been evaluated. That study showed that those singular patterns occur when the temperature falls below 0 °C. Indeed, before the retrofit intervention, the observed structural behavior is characterized by a bilinear trend in the frequency-temperature relationship, with a knee-point around 0 °C. The interpretation of this phenomenon has been guided by insights gained from similar previously analyzed case study where a similar relationship was observed and attributed to the freezing of the asphalt layer [15]. Further investigations specifically focused on the KW51 case study, also supported by finite element model updating, confirmed that the observed singularities in the natural frequency patterns were associated with the freezing of the porphyry ballast layer beneath the railway tracks.

Excluding the period during which the intervention took place, the dataset employed in this study comprises 6287 observations of the seven selected natural frequencies of the bridge, 3977 of which were collected before the retrofit intervention, while the remaining 2310 were gathered after the completion of the works.

In the context of the present study, the first 3579 samples collected before the retrofit – approximately corresponding to 90% of the available observations in the same period – have been used to train the GMM and to define threshold values for the subsequent comparative analyses. The remaining 10% of the dataset collected before the structural intervention has been used as a validation set, in order to check that no structural changes are detected before the onset of the retrofit. The whole natural frequency time series collected after the retrofit intervention are instead employed as the test data (Figure 3) to assess the accuracy of the different novelty detection strategies.

The trained GMM has been specifically designed to model the variability induced by EOVs, including the effects of freezing conditions observed between late January and early February 2019. The optimal number of GMM components has been selected as discussed in Section 2.1, resulting in a five-component GMM, which has been identified as the appropriate representation of the training dataset. Once the model of the operational variability of natural frequencies has been established, the NI time series has been computed according to Equations (2) and (3). The resulting NI values, shown in Figure 4, demonstrate the model's capability to effectively account for the influence of EOVs. Specifically, the NIs computed over the training data exhibit a consistent and stable behavior, indicating that the model successfully captures the normal variability of data, even under freezing conditions. Moreover, the NIs

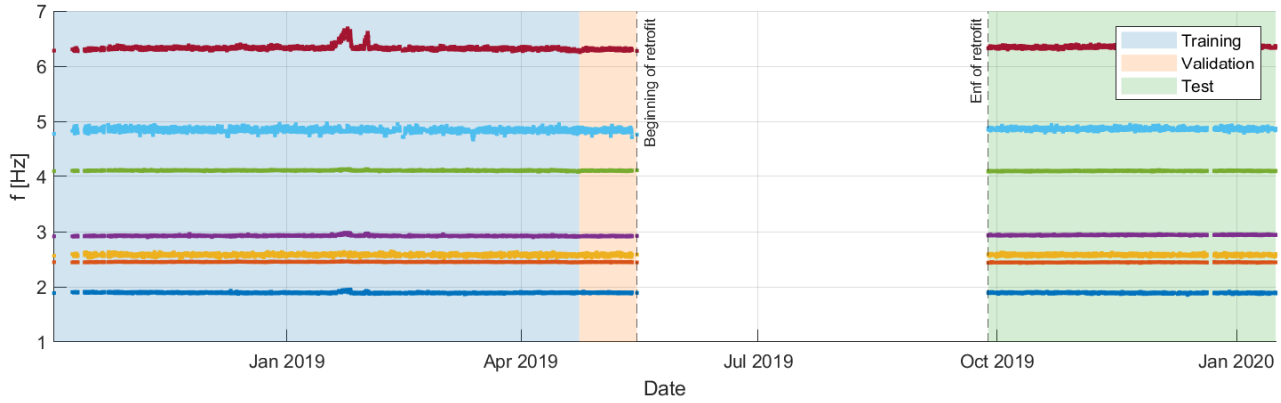


Figure 3. Time histories of the seven selected natural frequencies and partitioning of the dataset into training, validation and testing sets.

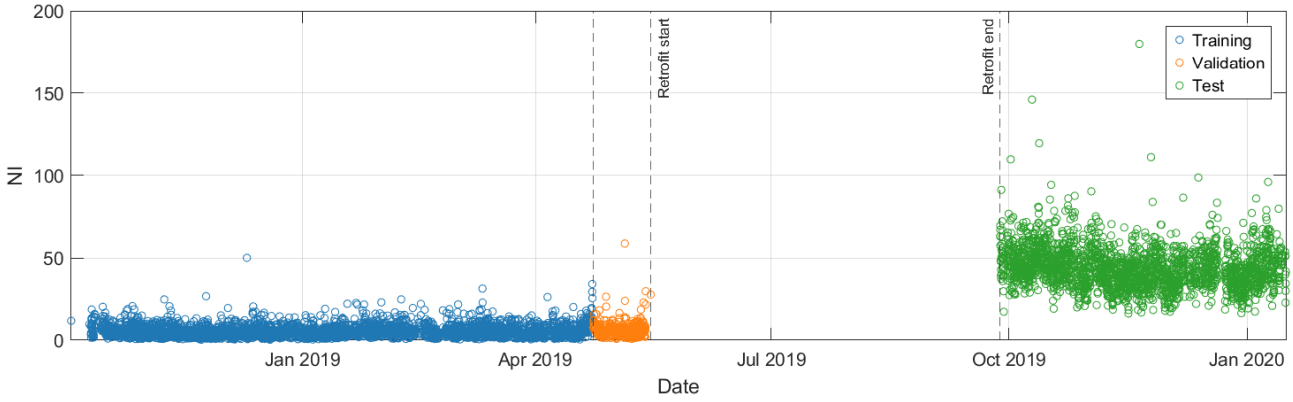


Figure 4. NIs time series during the entire monitoring period.

DOI: 10.3217/978-3-99161-057-1-143

calculated on the validation set remain aligned with those obtained during the training period, thus confirming the generalization capability of the considered model.

In contrast with these results, a marked upward shift in the NI trend can be observed when data collected after the retrofit intervention are considered, clearly suggesting the transition to a different structural state. This result confirms the effectiveness of the proposed approach in distinguishing between the normal environmental and operational variability of the selected DSFs and that associated to the occurrence of changes in the structure as a result of damage or, as in the present case, of a retrofit intervention. In the context of the novelty detection framework, an alarm threshold has been established based on the NIs computed from the training data. To this aim, a combination of the BM method and the GEV distribution fitting has been applied, as further illustrated hereafter.

The training NI time series has been divided into consecutive, non-overlapping blocks, and the maximum value from each block has been extracted to collect a set of extreme values. The choice of the number of blocks plays a critical role in the process [8]. In this work, the number of blocks has set equal to 300.

Figure 5 shows the comparison between the empirical cumulative distribution function (CDF) of the extracted maxima and the CDF of the fitted GEV distribution. The close agreement between the two curves indicates that the GEV distribution effectively describes the statistical variability of the observed maxima.

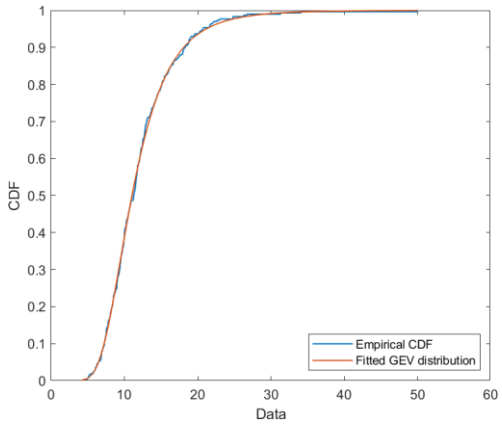


Figure 5. GEV distribution modeling: comparison between empirical CDF and fitted GEV distribution.

The fitted GEV model has been exploited to determine the alarm threshold corresponding to a significance level α of 0.05, as per Equation (5). Figure 7a illustrates the application of the determined threshold for novelty detection analysis of the full dataset, highlighting the occurrences of misclassification. During the training and validation periods, corresponding to the structural condition before the retrofit intervention, only a few isolated points exceeded the threshold, indicating a low false positives rate and, therefore, a high specificity of the proposed approach in characterizing the structural behavior in this state. Notably, after the retrofit intervention, only a very limited number of observations remained under the threshold. This demonstrates also the high sensitivity of the method in detecting the transition to a new structural condition, as it successfully identifies nearly all test data points, referring to the structural response after the retrofit intervention, as anomalous.

For the comparative assessment of the effectiveness of the method for threshold setting based on the GEV distribution, threshold values are also determined by means of the previously mentioned alternative procedures. The related results are presented in Figures 7b and 7c.

Table 2 summarizes the resulting threshold values, along with the corresponding number of misclassifications. Setting a fixed cut-off threshold is the most straightforward approach for the present task. For the considered application, a fixed cut-off threshold has been defined by setting the FAR to 0.05, meaning that up to 5% of the training data points are tolerated as FPs, in agreement with Equation (6).

As a second alternative approach, a threshold has been set based on the assumption of normal distribution of the natural frequencies in the training stage. As a result, the NIs derived through the MSD are expected to follow a χ^2_m distribution, with $m = 7$ degrees of freedom (with m corresponding to the number of modes considered). From this distribution, a threshold corresponding to a given significance level α has been determined.

The analysis of the results reported in Table 2 indicates that, although the threshold values obtained as the 95% cut-off value or through the χ^2 CDF achieve a zero false negative rate – meaning that all observations in the testing stage are correctly classified as anomalous –, they still suffer from a relatively high number of FPs, which can jeopardize the reliability and practicality of the monitoring system (Figure 7b and 7c). On the other hand, the EVT-based method for threshold setting yields a small number of FPs as well as a small number of FNs. While this method might appear less conservative, it establishes a threshold that better approximates the optimal balance between FPs and FNs. This can be demonstrated by looking at the coordinates associated with the various thresholding strategies when they are plotted on the ROC curve (Figure 6).

Table 2. Number of misclassifications for the different threshold setting procedures.

Approach	Threshold	FPs	FNs
BM-GEV	21.2	24	29
Cut-off	12.4	199	0
χ^2 CDF	14.1	120	0

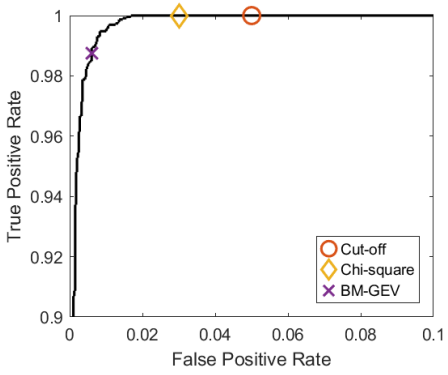
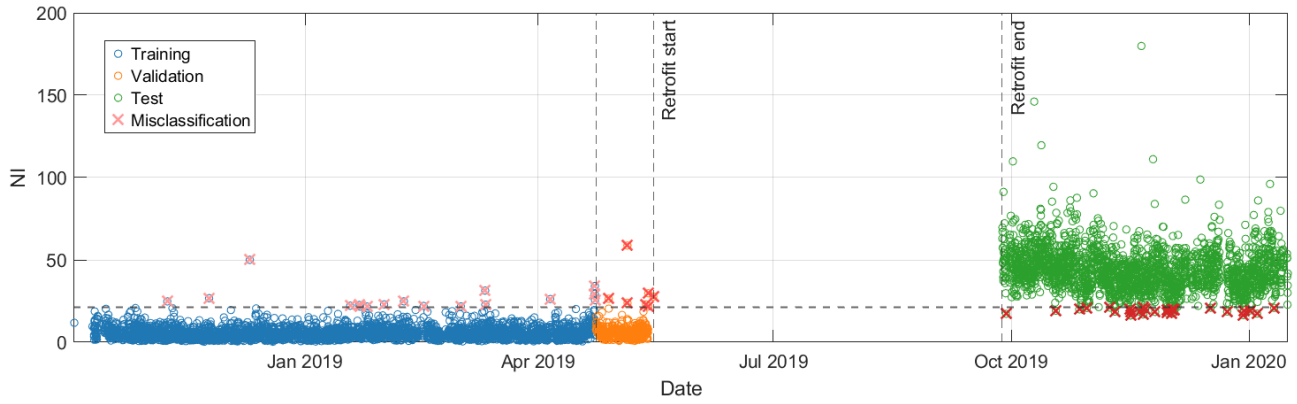
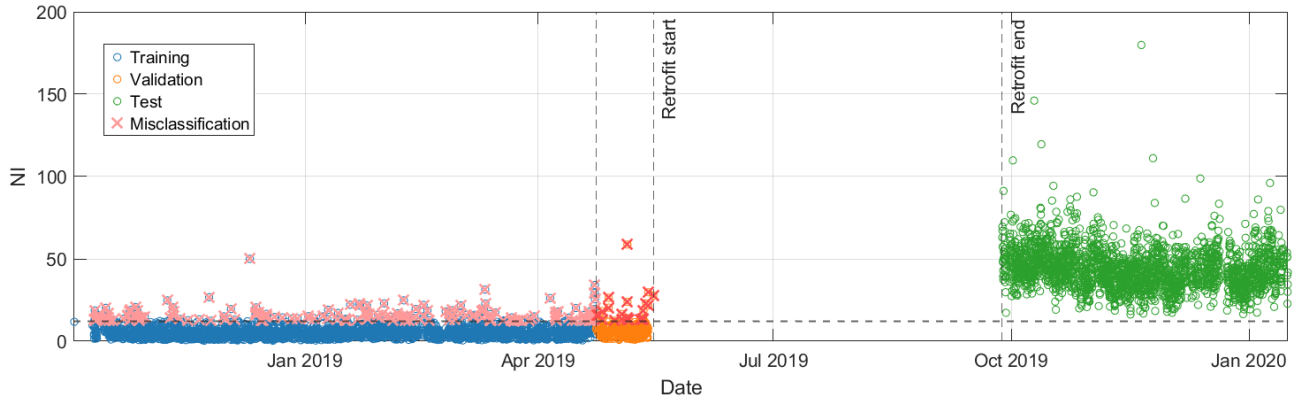


Figure 6. ROC curve and points corresponding to different threshold setting approaches.

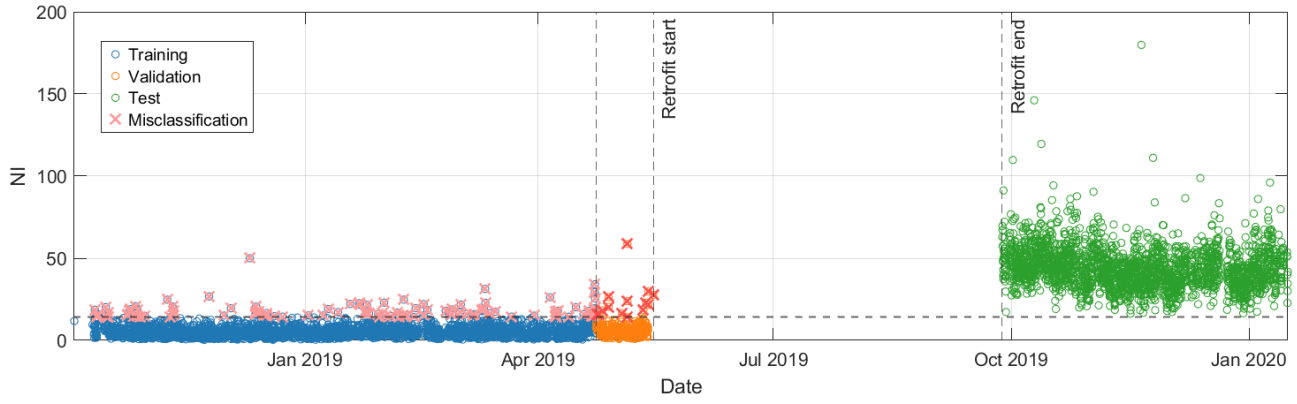
DOI: 10.3217/978-3-99161-057-1-143



(a)



(b)



(c)

Figure 7. Novelty detection according to: EVT-based threshold (a), 95% cut-off threshold (b), and χ^2 distribution-based threshold (c).

Indeed, the proximity of a point to the top-left corner of the ROC space – representing a low false positive rate and a high true positive rate – serves as a qualitative measure of the model classification performance. In addition, the distance of each threshold point from the bisector (corresponding to the line of no-discrimination) quantified by the Youden index provides a quantitative measure for the selection of the most effective threshold value.

The Youden index values have been computed for each threshold setting method, and they are reported in Table 3. The results indicate that the threshold derived through the BM-GEV method lies very close to the optimum, corresponding to the maximum Youden index. Furthermore, it outperforms the other considered approaches, confirming an excellent balance between sensitivity and specificity.

Table 3. Youden index values corresponding to different thresholds values.

Approach	Youden index
Max	0.987
BM-GEV	0.981
Cut-off	0.950
χ^2 CDF	0.970

DOI: 10.3217/978-3-99161-057-1-143

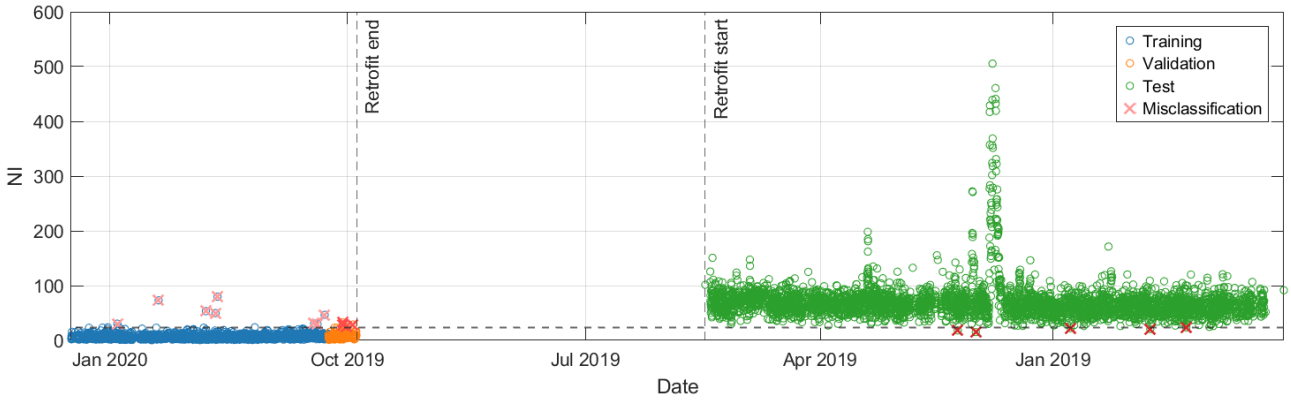


Figure 8. NIs time series during the entire monitoring period, considering the post-retrofitting condition as the healthy state for model training and validation, and EVT threshold-based novelty detection.

The analysis presented thus far follows the chronological order of data acquisition, corresponding to the subdivision of the dataset into training, validation, and testing sets. Since the retrofitting was intended to correct a construction defect, it led to an improvement in the structural condition of the bridge. As such, the post-retrofitting state can be regarded as the actual healthy condition.

An additional analysis is conducted by applying the proposed methodology using post-retrofitting data for model training and validation, and subsequently testing the model on pre-retrofitting observations. Once again, the novelty analysis results (Figure 8) confirm the model's ability to compensate for EOVs effects and demonstrate strong generalization performance, as evidenced by the NIs of validation data aligning closely with the training trend. The distinct structural condition characterizing the pre-retrofitting period is clearly revealed. Furthermore, a noticeable spike in the NIs time series during the freezing period highlights the presence of a transient condition within the pre-retrofitting state. In this case as well, applying EVT-based threshold yields an excellent balance between FPs (0.5%) and FNs (0.1%).

CONCLUSIONS

The present study has focused on the comparative performance assessment of different methods for determining a reliable threshold for anomaly detection to be applied in the context of data-driven, modal-based SHM. The natural frequency time series of the KW51 bridge served as the benchmark dataset for the performance assessment. It has been processed using a method that combines GMM and MSD to compute the NI time series and compensate the normal variability of the selected DSFs due to changing environmental and operational conditions. An approach based on EVT has been applied and compared with alternative approaches for threshold setting. The EVT-based approach started from the identification of a set of maxima in the NI time series at the training stage according to the BM method; the GEV distribution was afterwards fitted to the collected maxima and used to define the alarm threshold. Comparing the performance of the EVT-based method with other approaches for threshold setting has shown that, for the considered dataset, the BM-GEV approach appears to be the most precise, with an optimal balance between FPs and FNs, as confirmed by the value of the Youden index derived from the ROC curve.

ACKNOWLEDGMENTS

The present study is part of the research activities developed by the authors within the framework of the PNRR Program, CN00000023 National Center for Sustainable Mobility, SPOKE 7 "CCAM, Connected Networks and Smart Infrastructure" - WP4 (CUP B43C22000440001), and by the last author in the context of the PE00000005_1 "MITIGATE - Monitoring buIlT-up envIronment throuGh dinsAr Time series" Research Project (CUP E63C22002000002). Additional support from the STRIVE – INOSTRI FOE Project is also gratefully acknowledged. Finally, the KU Leuven Structural Mechanics Section is kindly acknowledged as the source of the data.

REFERENCES

[1] C. R. Farrar and K. Worden, Structural Health Monitoring: A Machine Learning Perspective. John Wiley & Sons, Ltd, 2012.
[2] Q. Chen, J. Cao, and S. Zhu, Data-Driven Monitoring and Predictive Maintenance for Engineering Structures: Technologies, Implementation Challenges, and Future Directions, IEEE Internet of Things Journal, vol. 10, n. 16, pp. 14527–14551, 2023.
[3] E. Figueiredo and J. Brownjohn, Three decades of statistical pattern recognition paradigm for SHM of bridges, Structural Health Monitoring, vol. 21, n. 6, pp. 3018–3054, 2022.
[4] M. A. F. Pimentel, D. A. Clifton, L. Clifton, and L. Tarassenko, A review of novelty detection, Signal Processing, vol. 99, pp. 215-249, 2014.
[5] Z. Wang, D. H. Yang, T. H. Yi, G. H. Zhang, and J. G. Han, Eliminating environmental and operational effects on structural modal frequency: A comprehensive review, Structural Control and Health Monitoring, vol. 29, 2022.
[6] H. Sohn, D. W. Allen, K. Worden, and C. R. Farrar, Structural Damage Classification Using Extreme Value Statistics, Journal of Dynamic Systems, Measurement, and Control, vol. 127, n. 1, pp. 125–132, 2005.
[7] E. Figueiredo and E. Cross, Linear approaches to modeling nonlinearities in long-term monitoring of bridges, Journal of Civil Structural Health Monitoring, vol. 3, n. 3, pp. 187–194, 2013.
[8] S. Coles, An Introduction to Statistical Modeling of Extreme Values. London: Springer London, 2001.
[9] J. Prawin and G. S. Vijaya Bhaskara, Outlier analysis combined with Gaussian mixture model for structural damage detection, Materials Today: Proceedings, 2023.
[10] R. G. Brereton and G. R. Lloyd, Re-evaluating the role of the Mahalanobis distance measure, Journal of Chemometrics, vol. 30, n. 4, pp. 134–143, 2016.
[11] E. Figueiredo, G. Park, C. R. Farrar, K. Worden, and J. Figueiras, Machine learning algorithms for damage detection under operational and environmental variability, Structural Health Monitoring, vol. 10, n. 6, pp. 559–572, 2011.
[12] V. Giglioni, E. García-Macías, I. Venanzi, L. Ierimonti, and F. Ubertini, The use of receiver operating characteristic curves and precision-versus-

DOI: 10.3217/978-3-99161-057-1-143

recall curves as performance metrics in unsupervised structural damage classification under changing environment, Engineering Structures, vol. 246, 2021.
[13] K. Maes, L. Van Meerbeeck, E. P. B. Reynders, and G. Lombaert, Validation of vibration-based structural health monitoring on retrofitted railway bridge KW51, Mechanical Systems and Signal Processing, vol. 165, n. 108380, 2022.
[14] C. Rainieri and G. Fabbrocino, Operational Modal Analysis of Civil Engineering Structures, New York: Springer, 2014.
[15] B. Peeters and G. De Roeck, One-year monitoring of the Z24-Bridge: environmental effects versus damage events, Earthquake Engineering and Structural Dynamics, vol. 30, n. 2, pp. 149–171, 2001.