Reducing false alarms in structural health monitoring systems by exploiting time information via Binomial Distribution Classifier

Reducing false alarms in structural health monitoring systems by exploiting time information via Binomial Distribution Classifier

S. Kamali[a,b,*], S. Quqa[a], A. Palermo[a], A. Marzani[a]

[a] *Department of Civil, Chemical, Environmental and Materials Engineering, University of Bologna, Bologna, Italy.*
[b] *Department of Civil and Environmental Engineering, Shiraz University of Technology, Shiraz, Iran.*

Abstract

This paper proposes an approach for anomaly/damage detection via structural health monitoring (SHM) systems based on a Binomial Distribution Classifier (BDC). The approach consists of two monitoring levels, labeled as alert and alarm states, respectively, where a damage index (DI) is computed and tracked over time. The alert state is reached when the DI exceeds a given threshold. At this stage, the BDC algorithm starts counting the number of DI values above the threshold within an observation window and computing the related probability of occurrence by using the binomial probability distribution. If the probability falls below a desired limit, the alarm state is triggered. Conversely, the SHM system returns in the non-alert condition. The proposed approach is discussed and evaluated through case studies involving both simulated and experimental data. In the examples, the DI is computed using the Mahalanobis distance of the monitored modal frequencies. The results demonstrate the capability of the BDC to reduce false alarms while preserving the probability of detection.

Keywords: Structural health monitoring, Anomaly detection, Binomial distribution, False alarms, Alarm threshold

*Corresponding author
Email address: soroosh.kamali2@unibo.it (S. Kamali)

List of acronyms

BDC	Binomial Distribution Classifier
DI	Damage Index
FPR	False Positive Rate
OMA	Operational Modal Analysis
PL	Probability Limit
PFA	Probability of False Alarms
PFA*	Effective PFA
PFA$_a$	Allowable PFA
POD	Probability Of damage Detection
POD*	Effective POD
TPR	True Positive Rate
$w_{\max}$	Maximum time window

1. Introduction

Civil infrastructures are the backbones of any modern society, and their aging is a serious and global matter of concern. Indeed, several infrastructures have already reached the end of their design life and present signs of degradation. Thereby, the real-time assessment of their conditions is of paramount importance, motivating the development and use of structural health monitoring (SHM) systems [1]. In SHM systems, permanent sensors are deployed on the structures to enable multiple and frequent measurements during operation [2, 3]. Continuous monitoring allows recording the structural response under different loads, for instance, passing vehicles and ambient vibrations [4].

In vibration-based SHM systems, the monitored data typically include natural frequencies, mode shapes, and damping ratios, identified via operational modal analysis (OMA) algorithms from ambient vibration measurements [4–6]. The collected data are then processed to identify changes in their statistical pattern over time, which might indicate structural modifications and, thus, the presence of potential damage. To this purpose, a common approach consists of defining a scalar parameter out of processed data, *i.e.*, the damage index (DI) [7, 8], and monitoring it over time. During the monitoring, the DI values are reported on a control chart and compared with a user-defined threshold. The latter is estimated on the DIs related to the reference state of the structure, the so-called "baseline", assuming that when the DI exceeds the threshold, the structure is likely to present an anomaly.

For this reason, the threshold assumes a pivotal role in discriminating non-anomalous vs. anomalous data.

The threshold should minimize false anomaly detection, *i.e.*, the number of DIs above the threshold for non-anomalous data, and ensure the desired sensitivity of the SHM system to detect anomalous data, especially when the anomaly is minor. This task is complicated by the stochastic nature of the DI. For instance, in vibration-based SHM, DIs computed on modal data vary over time due to environmental factors (environmental operational variability - EOV), such as temperature and humidity [7, 9–13], and unavoidable measurement noise, leading to a distribution of the DI for an identical structural state.

As such, strategies to set proper thresholds and assess the reliability of an SHM system are nowadays of primary importance. Commonly available approaches rely on statistical arguments like the false positive rate (FPR) and the true positive rate (TPR), which define the ratios of DIs exceeding the threshold for non-anomalous and anomalous data, respectively. In the Nondestructive Evaluation (NDE) domain [14, 15], where these arguments were originally introduced, every DI value exceeding the threshold is interpreted as an "alarm". The FPR is thus also termed "probability of false alarm" (PFA), while TPR coincides with the "probability of detection" (POD). Note that, while in NDE applications these two ratios are computed by testing once multiple "identical" components/structures, in SHM, a single monitored component/structure is repeatedly tested over time. Hence, some care is needed to adapt and exploit the classical concepts of FPR and TPR in the SHM context.

In this regard, some recent works propose suitable strategies to activate an "alarm" state accounting for the time dependence of the observations. For instance, Falcetelli et al. [16] discussed the challenges to developing proper reliability metrics for SHM starting from concepts originally used for NDE. The authors addressed the time dependency of the data highlighting the importance of the observation window and the frequency of the test in the assessment of the structural state. Similarly, Meeker et al. [17] reviewed the basic statistical concept of the POD by considering repeated measures. Ni et al. [18] proposed a reliability-based anomaly index to evaluate the health condition of expansion joints in bridges accounting for uncertainties from different sources and thus limiting false positives. Soleimani et al. [19] applied reliability analysis to define detection window length and threshold in a fully data-driven approach. Martucci et al. [20] extended the extreme

value statistics-based approach for outlier detection on scalar DIs to vectorial quantities through the extreme function theory. Toshkova et al. [21] proposed the adoption of two different states for monitoring rotating machinery, namely a "warning" and an "alarm" state, depending on the magnitude of the DI. First, the authors proposed a method based on machine learning to identify in which operational condition the machinery is working. Based on the identified condition and an accepted PFA, two thresholds were set for the warning and alarm states using the generalized extreme value theory. The method requires determining the parameters of the data distributions (*e.g.*, location, scale, and shape). Sarmadi and Karamodin [22] proposed a novel anomaly detection method specifically tailored for civil structures and a threshold selection approach based on the extreme value theorem, which does not rely on prior knowledge of the DI distribution. Sarmadi et al. [23] also proposed an anomaly score with automatic threshold selection based on the concepts of semi-parametric extreme value theory and peak-over-threshold.

Although the mentioned research works have demonstrated the possibility of adapting the damage detection theory from NDE to SHM, further studies are desirable to develop simpler, yet effective, algorithms to trigger alarm messages by an SHM system.

In this regard, this paper proposes a fast and intuitive algorithm called Binomial Distribution Classifier (BDC). The method adopts two different states, namely "alert" and "alarm." Specifically, the alert state is activated when the latest observed DI value is over a defined threshold. When this occurs, the SHM system enters a different monitoring level. At this level, the number of DI values over the threshold is counted for a limited time window (*i.e.*, the "observation window"). The number of positive DIs (*i.e.*, above-threshold) observed in this window can be associated with a probability of appearance, here computed using a binomial probability distribution. In doing so, the likelihood of observing a given number of positive DIs in the defined observation window for a healthy structure is estimated. Since a large number of positive DIs in the observation window is associated with a low probability of having the structure in a healthy state, the alarm state is reached when the probability falls below a desired limit.

This approach has the key advantage of drastically reducing the probability of sending false alarms while keeping the probability of detecting structural damage unchanged compared to the traditional threshold-based NDE approach. To quantify this benefit, the "effective PFA" and "effective POD" are determined based on the actual count of alarms activated by the pro-

posed method. In addition, although not investigated mathematically, the algorithm showed robustness in terms of different probability distributions of DI, (*e.g.*, normal, uniform, etc.), as we examined for different cases.

The proposed approach is tested on three case studies. The first is a simulated dataset for a steel truss bridge-like structure where the effects of temperature variation and instrumentation noise are modeled. Damage scenarios are simulated to investigate the effectiveness of the proposed method in damage detection. The second and third cases refer to the KW51 bridge [24] and Z24 bridge [25], for which datasets collected in different structural conditions are publicly available. In the presented examples, the approach was only tested for step-anomalies with modrate and large magnitudes.

2. Damage index, false positive rate and true positive rate

In SHM, the damage index (DI) is typically calculated over time from a set of damage-sensitive features. Without loss of generality, this study considers the natural frequencies as damage-sensitive features and a DI evaluated through the Mahalanobis distance. The latter provides a distance measure of a single observation from a set of observations [7].

As such, for a defined observation period (generally one year of temperature fluctuation) in which the structure is considered in the "healthy" state (baseline), the natural frequencies f_n^m of modes $m = 1, \ldots, M$ identified for $n = 1, \ldots, N$ observations are collected in a baseline dataset $\mathbf{H} \in \mathbb{R}^{M \times N}$. The DI associated to an observation vector $\mathbf{f}_n = \left[f_n^1, f_n^2, \ldots, f_n^M\right]^T$ is computed as:

$$\mathrm{DI}_n = \sqrt{[\mathbf{f}_n - \boldsymbol{\mu}]^T \mathbf{S}^{-1} [\mathbf{f}_n - \boldsymbol{\mu}]} \tag{1}$$

where $\boldsymbol{\mu} \in \mathbb{R}^{M \times 1}$ and $\mathbf{S} \in \mathbb{R}^{M \times M}$ are the empirical mean vector and the covariance matrix of the baseline dataset $\mathbf{H}$, respectively. It is worth mentioning that the larger the sample size N, the more accurate our estimates of statistical parameters and the more accurate the damage index will be.

Due to the presence of environmental factors and measurement noise, the DI varies over time, even if the structural state remains unchanged. For instance, temperature variations may affect the elastic moduli of materials and/or the structure support conditions, thus impacting the structure modal features [10].

Once the SHM system is in operation outside of the baseline, e.g., for $n > N$, a threshold is utilized to determine whether a DI is representative

of a healthy or damaged state. A classical approach consists of defining the threshold τ on the false positive rate (FPR) of the baseline distribution as:

$$\mathrm{FPR}(\tau) = \frac{\mathrm{FP}}{\mathrm{FP} + \mathrm{TN}} \tag{2}$$

where FP and TN indicate the number of false positives (DI values exceeding the threshold) and true negatives (DI values below the threshold), respectively. Fig. 1a shows a schematic of a typical control chart of the DI computed over time. The first year represents the baseline, in which the structure is assumed undamaged. The DIs (blue dots) change randomly, as the frequencies vary due to operational conditions and measurement noise. In the specific case of Fig. 1a, the DI is a realization of the normal distribution $\mathcal{N}(\mu, \sigma^2)$ with $\mu = 2$ and $\sigma = 1$. The threshold $\tau = 3.635$ is set such that the FPR = 5%.

For the selected threshold τ, the performance of the adopted DI in properly classifying an undamaged/damaged scenario can be assessed a posterior (if the structure is known to be damaged) via the so-called true positive rate (TPR). This metric measures the capability of a given methodology to properly classify a damaged component/structure from a series of tests as:

$$\mathrm{TPR}(\tau) = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FN}} \tag{3}$$

where TP is the number of DI values exceeding the threshold (true positives), while FN is the number of DI values below the threshold (false negatives).

For instance, referring to Fig. 1a, the structure is assumed to remain in a healthy condition for one year after the baseline (yellow dots) and to be damaged in the third year, thus generating a DI normally distributed (orange dots) with $\mu = 4$ and $\sigma = 1.2$. While the above-threshold DIs of the first year after the baseline are statistically similar to those of the baseline, in the damaged configuration (third year) more values are above the threshold leading to a TPR = 62.05%. Fig. 1b shows the FPR and TPR on the statistical representation of the damage indices for the data in baseline and the last monitoring year. In the NDE context [14, 15], every above-threshold DI activates an alarm, denoting that the tested component presents anomalies. This concept involves accepting false alarms, quantified in terms of probability of false alarm (PFA), which in this case coincides with the FPR. Similarly, the TPR is typically interpreted as the probability

of detection (POD) for a given damage state. Based on the above, it is evident that a beneficial reduction of PFA leads, for a given damage, to a deleterious reduction of POD, as the two values are strictly correlated. Still, this approach does not consider that in SHM the DIs represent the same structure over time and that, in general, the appearance of damage might not lead to immediate collapse. As such, this work proposes to exploit the evolution of DIs over time to compute an "effective PFA" smaller than the FPR in Eq. (2), while preserving an "effective POD" coinciding with Eq. (3).

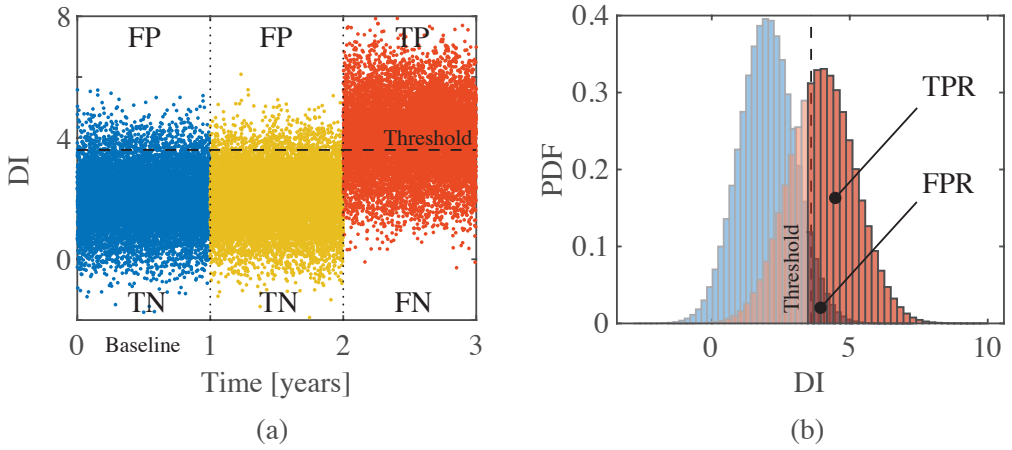


Fig. 1: Threshold-based damage detection: a) DI over time; b) definitions of FPR and TPR from the probability density functions of the damage indexes based on a FPR = 5% (threshold $\tau = 3.635$) that determines a TPR = 62.05%.

3. Proposed methodology

3.1. Overview of the method

The proposed monitoring approach accounts for two distinct monitoring levels, as schematized in Fig. 2.

In "Level 1", the SHM system continuously computes the DI over time and compares it with a user-defined threshold τ. If a DI exceeds the threshold, the system generates an *Alert* and enters the "Level 2" monitoring state. In this level, the SHM system observes consecutive DIs for observation windows with increasing widths ranging from $w = 2$ to a maximum window

$w = w_{max}$. At each width, w, the number of observed alerts is tallied, and the probability of occurrence $\overline{F}$ within w observations is calculated using the binomial probability distribution (the procedure is detailed in Section 3.2). This probability is compared with an acceptable probability level (PL).

If $\overline{F} < \text{PL}$, it is likely that the assumption of the healthy structural state is invalid, and the *Alarm* message is transmitted.

If, instead, the observation window reaches w_{max} without incurring a probability of occurrence below PL, the monitoring system returns to Level 1. PL and w_{max} are two user-selected parameters that can be set depending on the acceptable time between a DI over threshold and the alarm message, and on the acceptable rate of inspections for false alarms. More details on how they can be chosen are provided in Section 3.4.

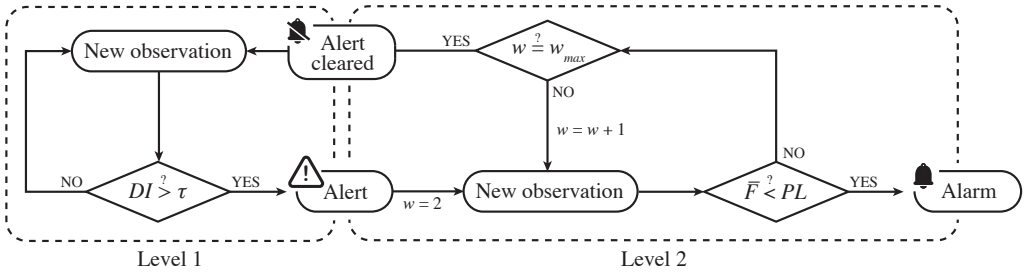


Fig. 2: The flowchart of the two levels of the proposed BDC algorithm.

3.2. Binomial probability distribution

If a Boolean random variable with two outcomes, namely success and failure, has a probability of success equal to p, the binomial probability mass function $P(s, w, p)$ represents the probability that from w independent experiments, s of them are successful. It can be calculated as:

$$P(s, w, p) = \binom{w}{s} p^s \left(1 - p\right)^{w-s} \tag{4}$$

To match the terminology with the present work, the term "success" is equivalent to "exceeding the alert threshold τ." Thus, p represents the FPR, w is the width of the observation window, and s is the number of observed DIs exceeding τ within the observation window. Therefore, $P(s, w, \text{FPR})$ represents the probability of observing s DI values over a threshold defined based on a specified FPR in an observation window of length w.

It is worth highlighting the difference between FPR and P. Fig. 3a shows the probability P of a healthy data set with FPR = 5% with an observation window $w = 100$. On average, it is expected that 5 of the observed DIs are over the threshold. However, although the outcome $s = 5$ has the largest probability of occurrence P, 4 or 6 positive DIs ($s = 4$ or $s = 6$) can also be found with a high probability.

With these premises, let us consider the cumulative distribution function, denoted by $\overline{F}$, defined as the probability that at least s alerts are observed in w consecutive samples, knowing that the first sample is always positive. Such probability can be expressed as:

$$\begin{aligned} \overline{F}\left(s, w, \mathrm{FPR}\right) &= \Pr\left(\left(w-1 \geq x \geq s-1\right) \cap \left(\mathrm{DI}_1 > \tau\right)\right) = \\ &= \left(\sum_{x=s-1}^{w-1} \binom{w-1}{x} \mathrm{FPR}^x \left(1-\mathrm{FPR}\right)^{w-x-1} \right) \mathrm{FPR} = \\ &= \sum_{x=s-1}^{w-1} \binom{w-1}{x} \mathrm{FPR}^{x+1} \left(1-\mathrm{FPR}\right)^{w-x-1} \end{aligned} \tag{5}$$

where DI_1 denotes the DI in the first position of the observation window.

Fig. 3b reports the values of $\overline{F}$ versus s, again computed for an observation window $w = 100$ and an FPR = 5%. Note that the probability $\overline{F}$ of observing at least $s = 2$ positive outcomes in the observation window (including DI_1) is almost equal to FPR (*i.e.*, 4.97%). Conversely, the probability of observing a high number of positives is quite low, *e.g.,* 0.64% for $s \geq 9$. Therefore, when this situation occurs, it is very unlikely that the structure is still in a healthy state. On the basis of this argument, it is possible to set a probability limit PL, for instance, PL = 1%, and issue the alarm when $\overline{F} < \mathrm{PL}$. In the considered scenario, observing at least 9 positive DIs within a window of 100 observations would lead to an alarm message.

In the proposed method, the value of $\overline{F}(w, s, \mathrm{FPR})$ is computed for each window width from $w = 2$ to $w = w_{max}$ and compared with PL to determine whether the structural state is anomalous (according to the algorithm shown in Fig. 2).

For a given value of FPR, the probability $\overline{F}(s, w, FPR)$ can be computed for different w and s via Eq. (5) and stored in a table. For instance, Tab. 1 is generated by considering a FPR = 5%; the values of $\overline{F} < \mathrm{PL} = 1\%$ are marked with a $*$. If any of the marked outcomes occur, the SHM system will

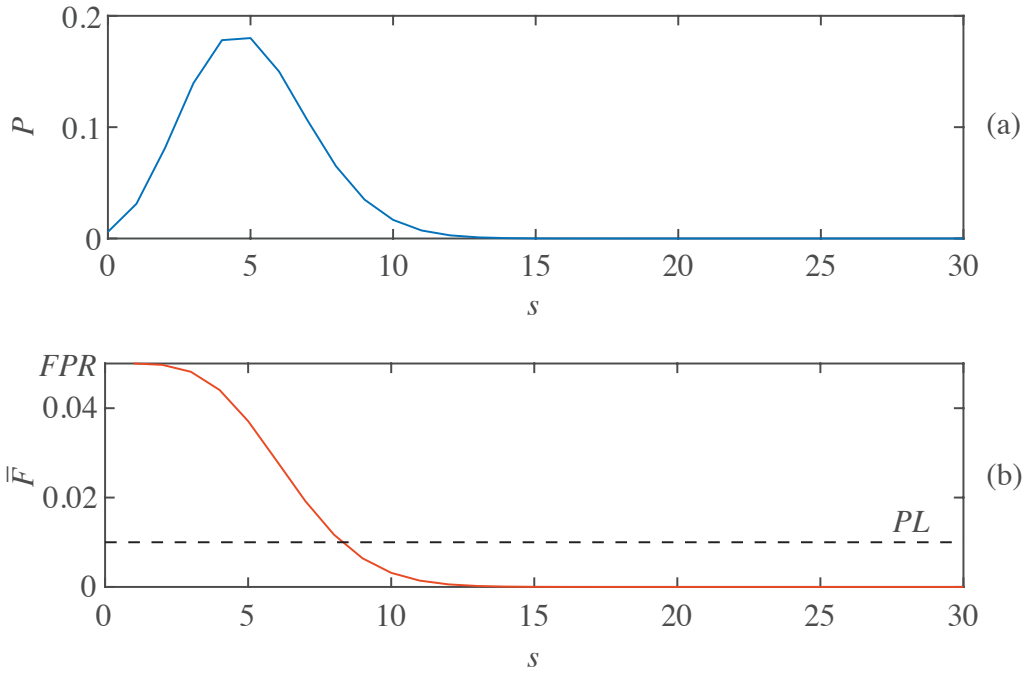


Fig. 3: a) The probability density function (P), and b) the cumulative distribution of Eq. 5 ($\bar{F}$) (plotted for $s = 1, ..., 30$) of the binomial distribution with $w = 100$ and FPR $= 5\%$.

transmit the alarm. It should be noted that Tab. 1 does not depend on the probability distribution of the DIs.

Table 1: $\overline{F}$ distribution of Eq. 5 with FPR $= 5\%$ for different w and s values.

w	$s \geq 1$	$s \geq 2$	$s \geq 3$	$s \geq 4$	$s \geq 5$	$s \geq 6$	$s \geq 7$	$s \geq 8$
2	0.05	0.0025*						
3	0.05	0.004875*	0.000125*					
4	0.05	0.007131*	0.000362*	6.25E-06*				
5	0.05	0.009275*	0.000701*	2.41E-05*	3.13E-07*			
6	0.05	0.011311	0.001130*	5.79E-05*	1.50E-06*	1.56E-08*		
7	0.05	0.013245	0.001639*	0.000111*	4.32E-06*	8.98E-08*	7.81E-10*	
8	0.05	0.015083	0.002219*	0.000188*	9.68E-06*	3.01E-07*	5.23E-09*	3.91E-11*

* indicates values of $\overline{F} <$ PL $= 1\%$

To quantify the performance of the proposed algorithm, new metrics termed as "effective PFA" and "effective POD" are formulated based on the exploitation of the BDC algorithm and discussed in the following.

3.3. Effective PFA and effective POD

A definition of the "effective PFA" and "effective POD", herein denoted as PFA* and POD*, respectively, are first introduced. Then, a closed-form equation is provided to estimate such quantities.

3.3.1. PFA* and POD* definition

For an empirical definition of the PFA* let the subset A represent the space of false alarms using the BDC algorithm, and the subset B the space of total BDC checks within the sample space of a healthy structure. Then, $P(C) = P(A|B)$ is defined as the probability of a false alarm occurring given the BDC algorithm is used. Since A is always a subset of B, this probability can be expressed as:

$$P(C) = P(A|B) \simeq \frac{n(A)}{n(B)} \tag{6}$$

where $n(A)$ denotes the number of false alarms, and $n(B)$ represents the total number of BDC checks. Additionally, an event D is defined as the space of alerts within the sample space of a healthy structure, given by:

$$P(D) = P(DI \geq \tau) = \text{FPR} \tag{7}$$

The PFA* is then defined as the probability of a false alarm occurring within the BDC algorithm, the latter activated by an alert. Therefore, the PFA* can be calculated as:

$$\text{PFA}^* = P(C \cap D) = P(C)P(D) = \frac{n(A)}{n(B)}\text{FPR} \tag{8}$$

Notably, Eq. (8) is valid since the events C and D are independent, *i.e.*, the BDC diagnosis does not rely on the activation alert.

Similarly, the POD* can be defined within the sample space of a damaged structure. POD* represents the probability of correctly detecting damage using the BDC algorithm activated by an alert. This probability can be expressed as:

$$\text{POD}^* = P(C \cap D) = P(C)P(D) = \frac{n(A)}{n(B)}\text{TPR} \tag{9}$$

In Eq. (9), A represents the space of detected damages (*i.e.*, when an alarm message is actually sent out), B denotes the space of total BDC checks, $P(C) = P(A|B)$ is the probability of detecting damage given the BDC algorithm is used, and D represents the space of alerts within the sample space of a damaged structure.

From Eqs. (8) and (9), it can be observed that PFA* of the BDC algorithm is always less than or equal to FPR; similarly, the POD* cannot exceed TPR.

3.3.2. Closed-form equation to predict PFA* and POD*

A closed-form equation to compute the PFA* can be written as:

$$\text{PFA}^*(w_{max}) = \sum_{w=2}^{w_{max}} \text{FPR}^{\bar{s}}(1-\text{FPR})^{w-\bar{s}} \sum_{i=\bar{s}-w+\hat{w}-1}^{\hat{s}-2} \binom{\hat{w}-1}{i} \binom{w-\hat{w}-1}{\bar{s}-i-2} \tag{10}$$

where $\bar{s} = \bar{s}(w)$ represents the minimum number of observations for a given window length w such that $\overline{F} < \text{PL}$. Such values can be determined by using a table similar to that presented in Tab. 1, computed for FPR = 5%. Similarly, $\hat{w} < w$ is the largest window size before the last shift of a cell marked with $*$ to the right in Tab. 1, and $\hat{s}$ is the value of $\bar{s}$ at which it occurred. For instance, if $w = 7$, then $\bar{s} = 3$, $\hat{w} = 5$, and $\hat{s} = 2$.

It should be noted that Eq. (10) is valid for $w_{max} \geq 2$. When $\hat{s} \leq 2$ (*i.e.*, before the first shift of the marked cell), the values $\hat{s} = 2$ and $\hat{w} = 1$ should be used in the formula.

Eq. (10) can also be employed to calculate the POD* by substituting the value of FPR with the one of TPR. In this case, $\bar{s}$, $\hat{s}$, and $\hat{w}$ can still be determined using Tab. 1.

Fig. 4 shows the PFA* and POD* for different maximum window sizes ($w_{max} = 2, ..., 20$) computed for the normally-distributed DIs shown in Fig. 1, obtained by setting a FPR of 5% ($\tau = 3.635$ and TPR = 62.05%) and a probability limit PL = 1%. The blue stars denote the result obtained by applying the proposed BDC algorithm to the data, with PFA* and POD* computed by counting the number of false alarm and detected damages of the BDC algorithm, respectively. For such computation, 1 million DI values

are considered for both the damaged and the undamaged states. The orange lines, representing the values obtained through the application of the proposed Eq. (10), are in perfect agreement with those computed by the BDC algorithm.

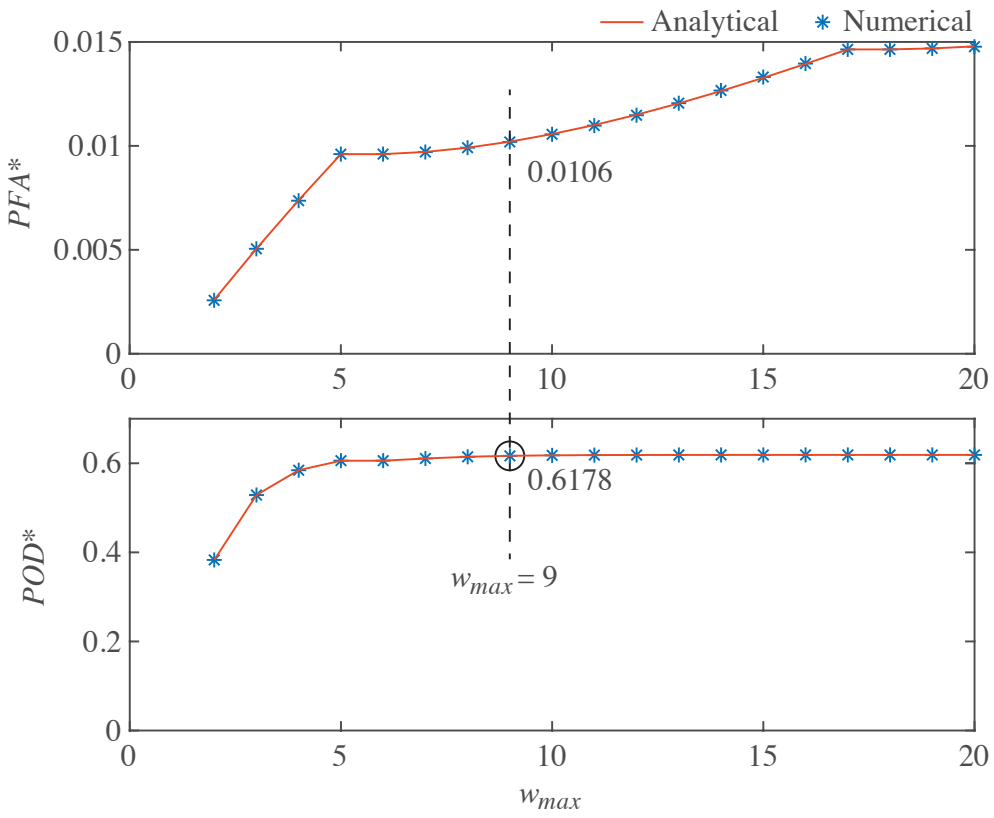


Fig. 4: Effective PFA (PFA*) and effective POD (POD*) in the case of the normally distributed DI shown in Fig. 1 with $\tau = 3.635$ and PL $= 1\%$ (blue stars). The analytical values are predicted using Eq. (10) considering PL $= 1\%$, FPR $= 5\%$ and TPR $= 62.05\%$ (orange lines).

As shown in Fig. 4, for given values of FPR and PL, a different selection of w_{max} leads to different values of PFA* and POD*. In particular, the PFA* is substantially lower than the PFA traditionally used in the NDE field for all the analyzed cases, e.g. the FPR. Conversely, the POD* quickly converges to TPR of 62.05% when increasing w. Therefore, an optimal w_{max} can be chosen as the value for which the increase of POD* with w becomes practically

negligible, here $w_{max} = 9$. Increasing w over w_{max} is disadvantageous as it only leads to higher PFA*.

Fig. 4 shows that the proposed approach offers the advantage of substantially reducing alarm messages, bringing the "original" PFA = 5% to PFA* = 1.06% while maintaining POD* $\simeq$ TPR, at the cost of waiting for up to 9 observations of the DI (*i.e.*, w_{max}) before transmitting the alarm message. In contrast to traditional methods, this approach does not depend on a single DI, but it assesses the DIs dynamically, taking into account consecutive observations.

In the example discussed, the proposed method is exploited to substantially reduce the PFA. Similarly, it can also be used to increase the POD. Indeed, let the threshold discussed in Section 2 be $\tau = 3.2$ instead of the original value $\tau = 3.635$. In this case, the dataset would lead to an FPR = 11.52% and TPR = 74.79%. Using the proposed approach with these values and PL = 1% provides the PFA* and POD* shown in Fig. 5. In particular, using the same $w_{max} = 9$ as the previous case leads to a slight increase in PFA* from 1.06% to 1.25%, still well below the FPR, and to a considerable improvement in POD* from 61.78% to 74.70%.

It is important to remark that the proposed method also works for DIs that are not normally distributed. For instance, let the DI be uniformly distributed such that the baseline DIs are included in $\mathcal{U}(2, 3)$, while the DIs representative of the damage state are included in $\mathcal{U}(2.7, 3.7)$. In this case, a FPR = 5% leads to a threshold $\tau = 2.95$ and to a TPR = 75%. Applying the proposed method, it is possible to observe that POD* stabilizes at $w_{max} = 7$ (Fig. 6), bringing PFA* down to 0.95% while keeping the POD* $\simeq$ TPR.

3.4. Selection of w_{max} and PL

The selection of w_{max} determines the maximum time delay from the initial identification of an abnormal DI to the alarm message being sent. Hence, it is important to choose an appropriate w_{max} to ensure the SHM system can promptly detect damage with a satisfactory level of confidence. It is worth highlighting that the optimal value of w_{max} to maximize POD* can be selected as discussed in Section 3.3. However, a smaller value can be selected based on the specific time requirements of the infrastructure manager/owner.

Having an estimation on w_{max}, it is possible to find a suitable value for PL. Suppose that the infrastructure owner can tolerate false alarms every d days and the SHM system transmits DI data h times a day, the maximum

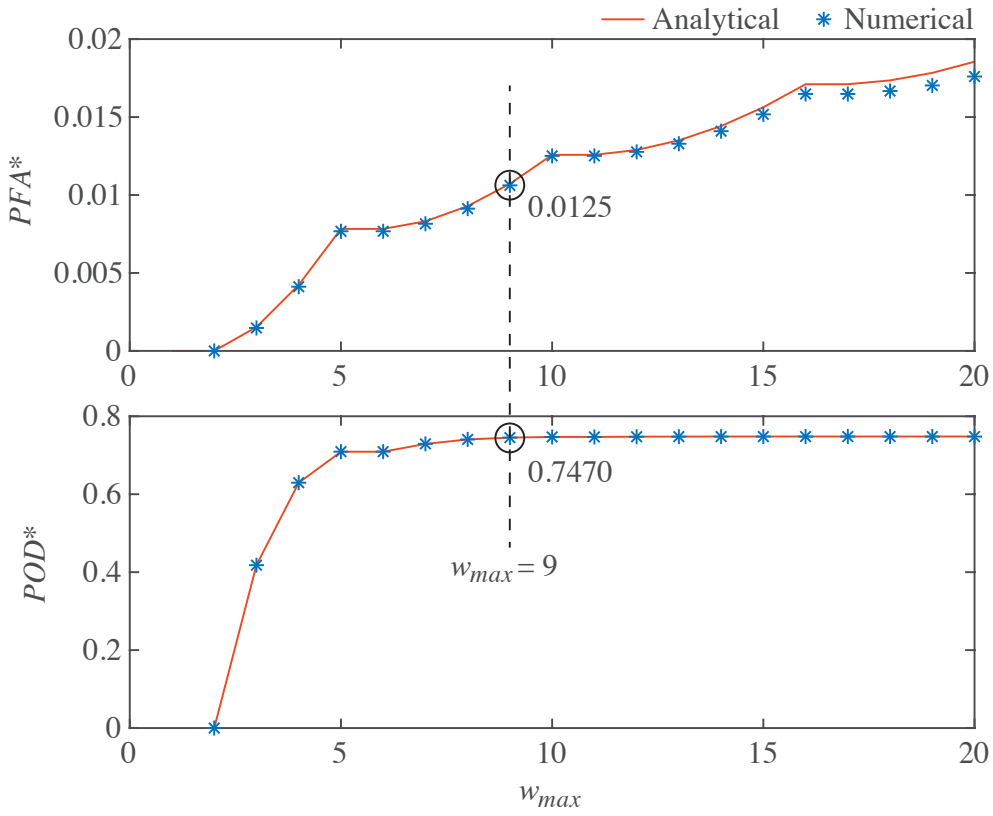


Fig. 5: Effective PFA (PFA*) and effective POD (POD*) in the case of the normally distributed DI shown in Fig. 1 with $\tau = 3.2$ and PL = 1% (blue stars). The analytical values are predicted using Eq. (10) considering PL = 1%, FPR = 11.52% and TPR = 74.79% (orange lines).

allowable probability of false alarm (namely PFA_a) of the SHM system can be determined as:

$$\text{PFA}_a = \frac{1}{dh} \tag{11}$$

For instance, if the owner is willing to tolerate false alarm inspections with a return period of 100 days and the SHM system transmits DIs once a day, PFA_a would be $1/(100 \times 1) = 0.01$.

At this point, the value of PL can be chosen based on a parametric analysis requiring that PFA* is below the specified value of PFA_a. For this purpose, having the baseline DIs and a selected w_{max}, the following steps can

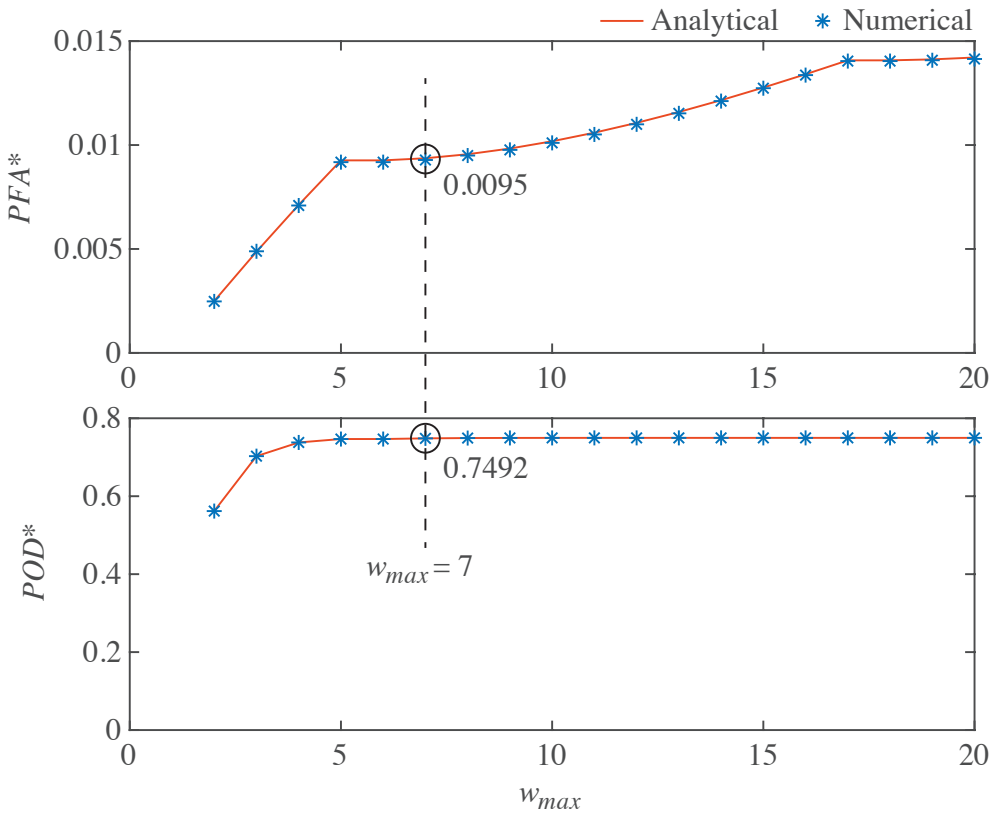


Fig. 6: Effective PFA and POD in the case of uniformly distributed DI with $\tau = 2.95$ and PL $= 1\%$.

be followed:

- the alert threshold τ is set based on FPR $= 5\%$ on the healthy DIs and alerts are detected;
- for each PL $\in [\text{PL}_{min}, \text{PL}_{max}]$, the BDC algorithm is applied to diagnose false alarms and the PFA* is computed;
- a value of PL satisfying PFA$^*(PL) \leq \text{PFA}_a$ is selected.

For instance, considering the healthy dataset of 3650 DIs shown in Fig. 1 and a FPR $= 5\%$, Fig. 7 shows the values of PFA* computed for different values of PL and maximum window widths $w_{max} = [4, 8, 12, 16, 20]$. It is evident that increasing PL leads to a corresponding increase in PFA*. Likewise,

expanding w_{max} also raises PFA*, albeit with a different degree of impact, as discussed in Section 3.3. Different combinations of w_{max} and PL can be chosen with the aim of keeping PFA* below a certain value. In this example, if the condition PFA$^* < 0.01$ is explicitly required by the infrastructure manager/owner, $w_{max} = 8$ and PL $= 0.008$ could be a suitable choice.

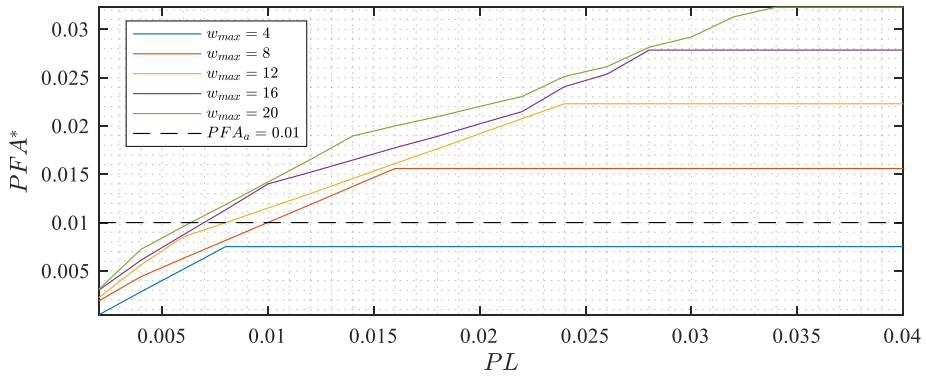
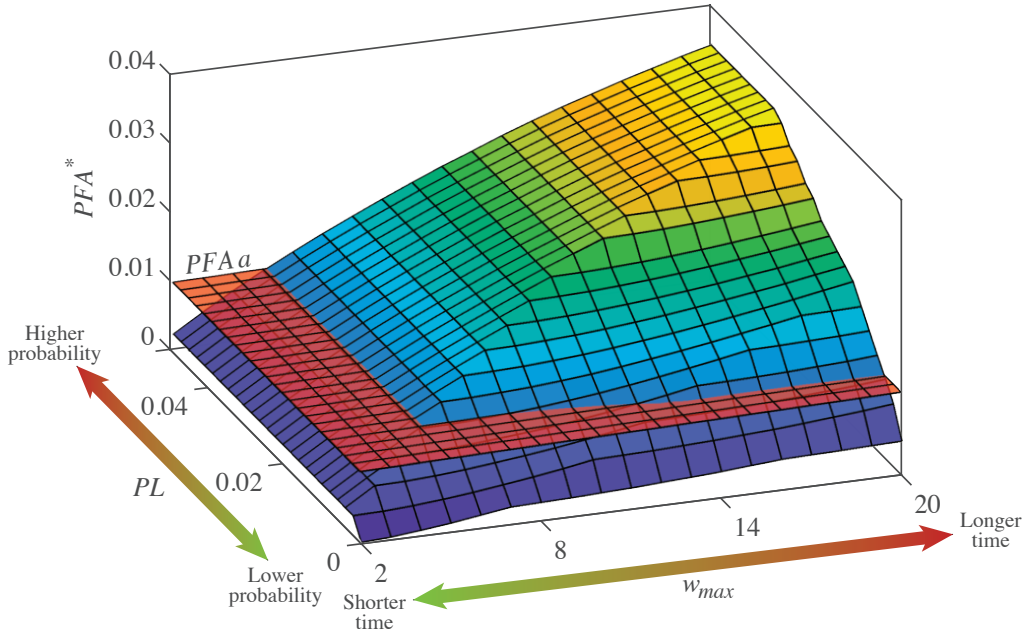
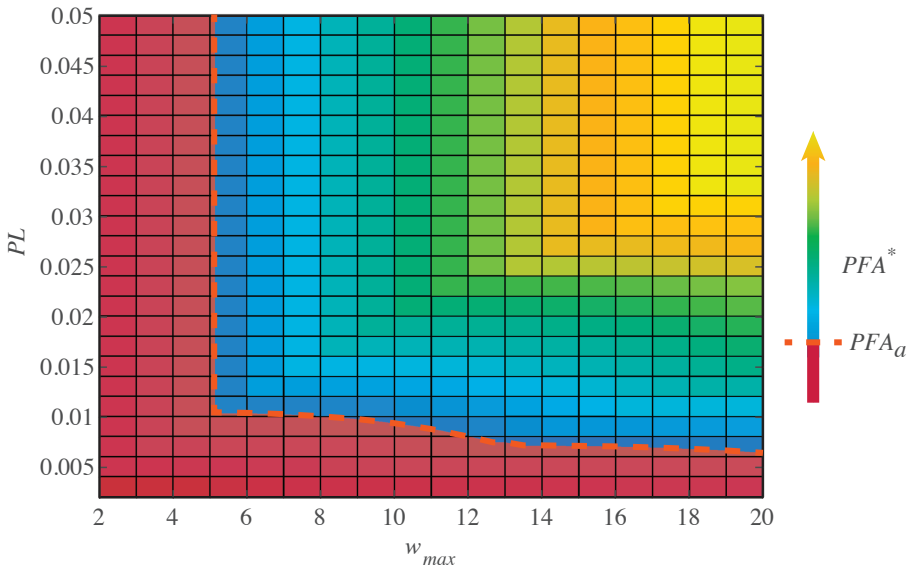


Fig. 7: PFA* vs PL computed on the healthy DIs of Fig. 1 for an assumed FPR=5%.

Alternately to the analysis in Fig 7, the selection of the BDC parameters is discussed in Fig. 8a where PFA* is shown as a function of PL and w_{max}. For a given PFA$_a$ (red plane), all the combinations of parameters lying on the dashed red line of Fig. 8b meet the owner/manager requirements. While selecting the minimum w_{max} among the possible configurations ($w_{max} = 5$) would also minimize the delay in alarm transmission, it may lead to a sub-optimal situation in terms of POD*, as can be seen in Fig. 9. In fact, it is possible to observe that POD* grows fast with w_{max} and the minimum w_{max} that fulfills the manager requirements may lead to lower POD*. Therefore, selecting a wider w_{max} may be convenient, compatibly with the manager time constraints.



(a)



(b)

Fig. 8: (a) Sensitivity analysis of the PFA* vs PL and w_{max} using the healthy data of Fig. 1, with FPR = 5%. (b) Top view where the intersection PFA* = PFA$_a$ = 1% is highlighted as a dashed red line.

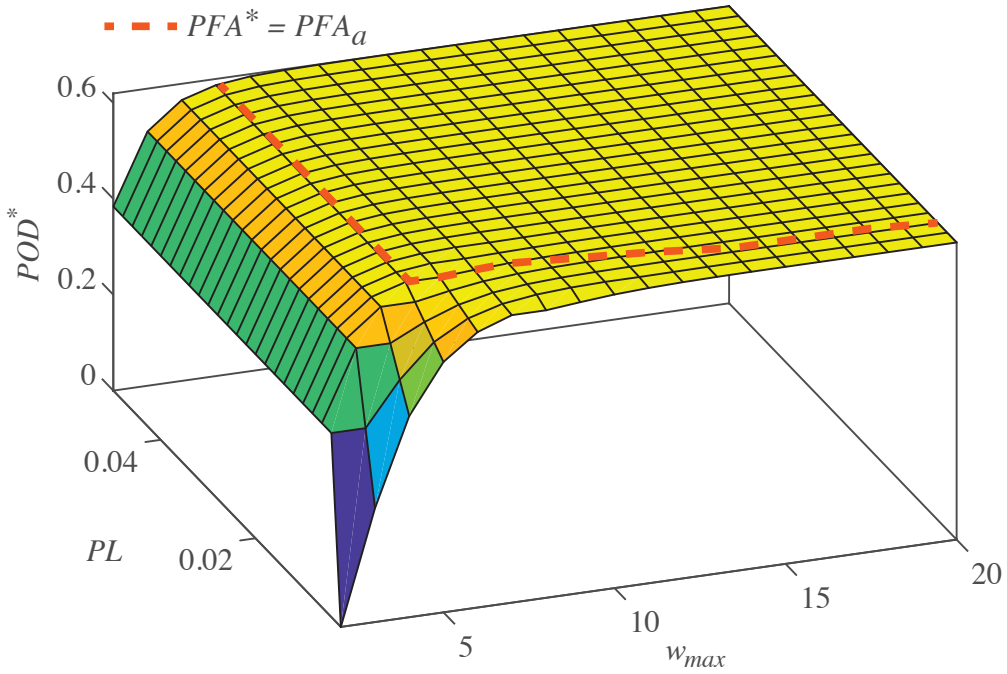


Fig. 9: POD* in the cases analyzed in the sensitivity analysis, using the damaged data of Fig. 1, with FPR = 5%.

4. Applications

4.1. Simulated scenario

A statically determinate truss structure with 9 bar elements resting on two vertical supports (elements 10 and 11) is considered (see Fig. 10a).

Each bar element has a cross-section area $A = 0.0025\ m^2$, Young's modulus $E_0 = 200\ GPa$, and a material density $\rho = 7850\ kg/m^3$. Using a finite element model of the structure, the first modal frequencies of the truss are determined as $f_1 = 36.71\ Hz$, $f_2 = 78.91\ Hz$, $f_3 = 82.00\ Hz$, $f_4 = 145.97\ Hz$, and $f_5 = 219.22\ Hz$.

To introduce environmental operational variability (EOV), a temperature distribution is built, including seasonal, daily, and random fluctuations (see Fig. 10b). In particular, ten samples per day per two years are considered, resulting in 7300 samples. The temperature effect is applied to the structure assuming a non-linear temperature-dependent axial stiffness $k(T)$ for the two

supports (elements 10 and 11) as [10]:

$$k(T) = \frac{E_0 A}{l}[1 - 0.005(T - T_0)^{(-0.01T)} + 0.05r] \tag{12}$$

where $l = 1$ m and $T_0 = 20$ °C, and r is a random variable with a normal distribution $\mathcal{N}(0, 1)$. Fig. 11 illustrates the stiffness of the supports $k(T)$ with respect to the time samples and to the corresponding temperatures. The structure is considered healthy during the first year; in the second year, the cross-section of element 1 is damaged with a stiffness reduction of 10%.

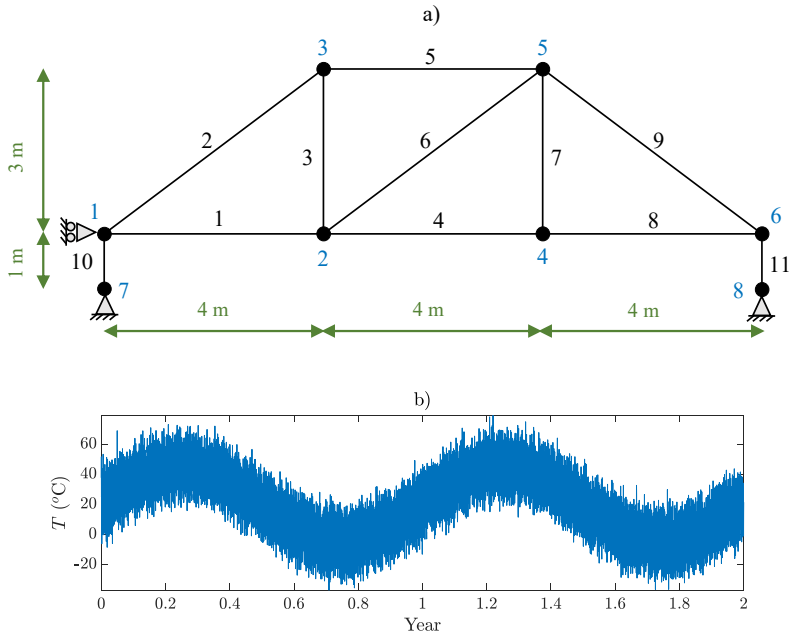


Fig. 10: Numerical case study: a) Scheme of the structure b) Simulated temperature variation

The natural frequencies of the undamaged structure, computed for the 3650 temperature samples of the first year, are subsequently contaminated with 1% noise, a common occurrence in vibration-based SHM systems [26, 27]. These healthy frequencies (first 3650 time samples of Fig. 12) are used to form the baseline matrix $\mathbf{H}$. Then, 3650 DIs are calculated as the Mahalanobis distance between $\mathbf{H}$ and its own distribution. The computed DIs are shown as blue dots in Fig. 13 (first 3650 samples). Assuming a

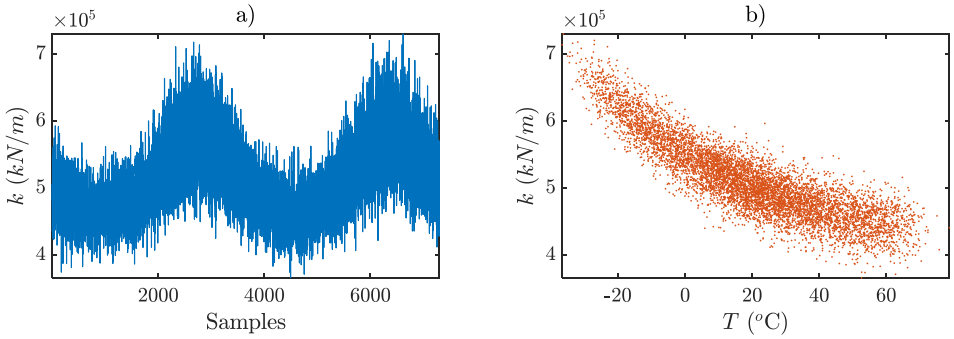


Fig. 11: Variations of the supports stiffness according to Eq. (12) with respect to a) samples; b) temperature.

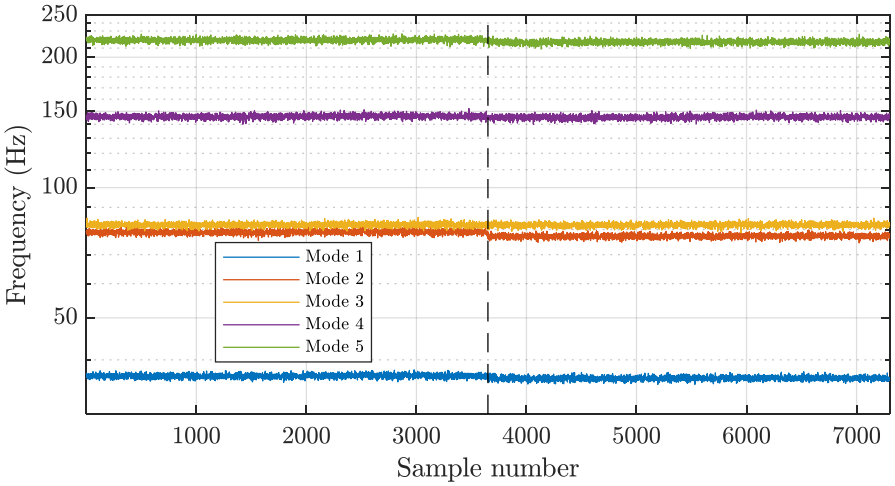


Fig. 12: Natural frequencies of the healthy (first year) and the damaged (second year) structure for damage $\kappa = 0.10$ on element 1.

FPR = 5% leads to an alert threshold of $\tau = 3.2946$ (reported as a dashed horizontal line in Fig. 13).

The natural frequencies of the damaged structure, computed assuming the temperature distribution of Fig. 10b and an added noise of 1%, are shown in Fig. 12 (samples 3651 to 7300). The DIs, computed using the mean vector and covariance matrix of $\mathbf{H}$, are shown in Fig. 13 (samples from 3651 to 7300).

For the considered threshold $\tau = 3.2946$, a TPR = 51.59% is obtained for the damaged configuration.

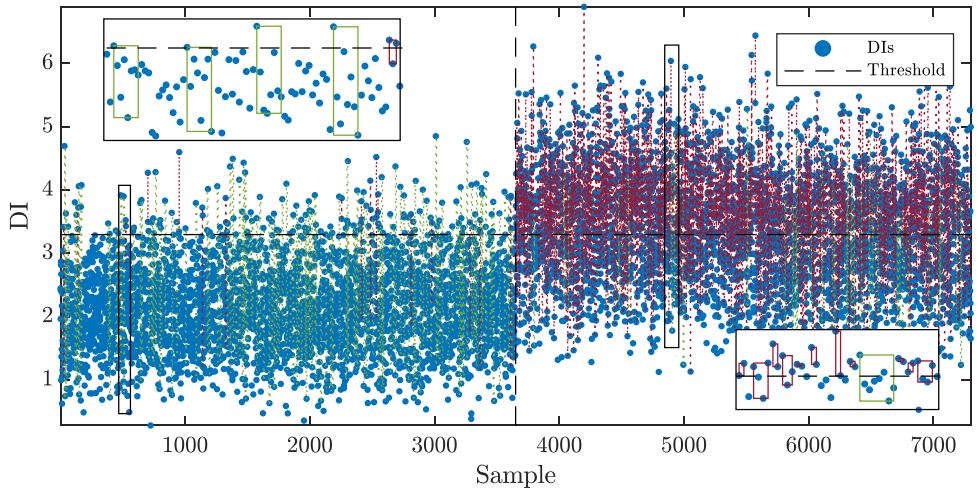


Fig. 13: DI computed over time and application of the BDC algorithm to the numerical case study using $w_{max} = 8$ and PL $= 0.008$.

The proposed BDC algorithm is thus applied to the simulated DIs. For this purpose, a maximum window width of $w_{max} = 8$ and a probability limit of PL $= 0.008$ is adopted based on the discussion in Section 3.4.

The rectangular windows in Fig. 13 indicate the samples included in each BDC check. Each window begins with an alert (the DI is above τ), where the algorithm switches to Level 2. If the algorithm returns to Level 1 without transmitting alarms, the window border is colored green, while a red border is used to label the occurrence of an alarm.

In the case of the undamaged structure (first year), out of 3650 samples, 182 DIs above τ are found. This would suggest an occurrence of a misleading alarm (false positive) every two days in the classical context

(refer to Fig. 13). In the same time interval, the application of the BDC algorithm yields only 19 wrong alarms out of 141 BDC checks thus being highly effective in avoiding false alarms ($\text{PFA}^* = (19/141) \times \text{FPR} = 0.67\%$, FPR $= 5\%$). Thus, with $w_{max} = 8$, PL $= 0.008$, and the limitation of $\text{PFA}_a = 0.01$, the designed system would satisfy the requirement of $\text{PFA}^* < \text{PFA}_a$.

Furthermore, on the damaged samples of the second year, the method is noticeably accurate in diagnosing 872 true damages out of 906 BDC checks ($\text{POD}^* = (872/906) \times \text{TPR} = 49.65\%$).

It should be noted that all green observation windows contain 8 samples, which is equivalent to w_{max} since the false alarm is reported if no $\bar{F} < \text{PL}$ is observed up to w_{max}.

Conversely, the red windows frequently contain a number of samples less than w_{max}, as the alarm is sent out as soon as the condition $\bar{F} < \text{PL}$ occurs.

4.2. KW51 Bridge

In this section, real data obtained from the KW51 bridge (Leuven, Belgium) [24] are used to test the proposed procedure.

The bridge was monitored before (October 2, 2018 – May 15, 2019), during (May 15, 2019 – September 27, 2019), and after (September 27, 2019 – January 15, 2020) retrofitting interventions. The publicly available database provides access to the 14 natural frequencies of the bridge, which were determined using the stochastic sub-space identification (SSI) method [28] in the different health conditions of the bridge.

It is worth mentioning that the KW51 database is characterized by some missing data (see Fig. 14a). To address this issue, missing data were replaced by the average values of the preceding ten available data points for each missing frequency. While we are aware that the chosen data imputation technique can influence the quality of the diagnosis, a detailed discussion of missing data imputation techniques, while important, falls outside the scope of this paper. Finally, the resulting filled database is illustrated in Fig. 14b.

The 3.5-month period after retrofitting is here considered the "healthy state" of the bridge (baseline). For this purpose, the DIs are calculated as the Mahalanobis distance of the entire frequency dataset with respect to the baseline. Having the baseline DIs, the alert threshold is set to provide a FPR $= 0.05$ resulting in a $\tau = 5.4518$ (see Fig. 14c).

The performance of the BDC algorithm is assessed herein for (i) limiting false alarms during the post-retrofitting stage, and (ii) detecting anomalies in the other two stages. For this purpose, PL $= 0.05^3$ and $w_{max} = 10$ are assumed. The results obtained via application of the BDC procedure are reported in Fig. 15.

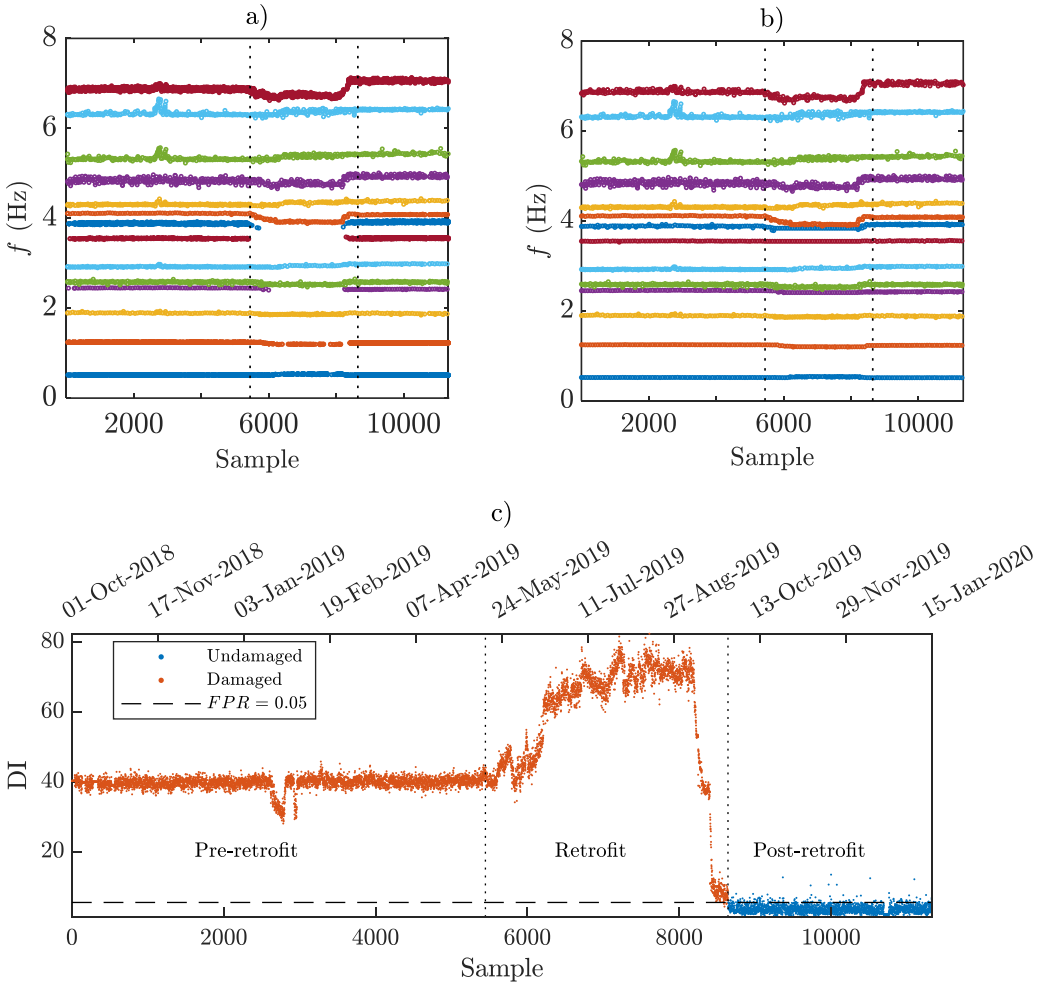


Fig. 14: a) Natural frequencies of the bridge with missing points, b) Filled database, c) Resulting DI distributions of the KW51 bridge.

Remarkably, the BDC algorithm provides excellent diagnostic capabilities, without sending out any alarms in the post-retrofit state and accurately detecting the anomaly in the entire pre-retrofit and during-retrofit intervals. Note that 134 alerts are issued during the post-retrofitting period, resulting in daily false alarms for traditional threshold-based methods.

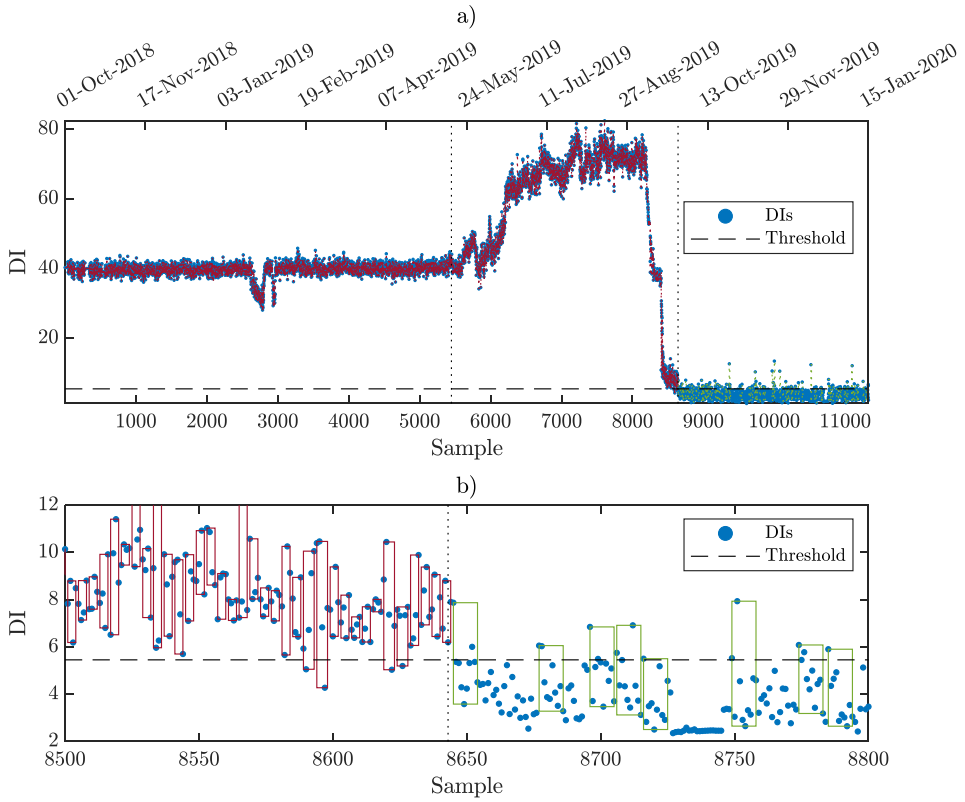


Fig. 15: DI computed over time and application of the BDC algorithm to the KW51 bridge using $w_{max} = 10$ and PL $= 0.05^3$: a) Entire monitoring period; b) Close-up of the interval when the anomaly occurs.

Fig. 15b provides a clear visualization of the boundary between the anomalous and non-anomalous states of the bridge and the ability of BDC to differentiate them without errors, confirming the effectiveness of the proposed algorithm as a reliable anomaly classification tool.

4.3. Z24 Bridge

The Z24 bridge was situated in the canton of Bern near Solothurn, Switzerland. It served as an important part of the roadway connection linking the villages of Koppigen and Utzenstorf, over the A1 highway connecting Bern and Zurich. Constructed in 1963, the bridge was a conventional, post-tensioned concrete structure comprising a two-cell box-girder design, with the primary span measuring 30 m. Due to the construction of a new railway

adjacent to the highway and the need for a bridge with a larger lateral span, the Z24 was demolished at the end of 1998. In particular, starting from August 10, 1998, a series of damages were imposed on the bridge (see Tab. 2) [29, 30]. The dynamic response of the bridge was recorded over a one-year period, by using 16 accelerometers placed at various locations, including the period of the demolition phase.

In this study, the first four modal frequencies of the bridge are employed (illustrated in Fig. 16a), obtained by Favarelli et al. [31] from recorded accelerations.

Table 2: Scenario timeline of the Z24 bridge.

Date (1998)	**Scenario**
4 August	Reference condition
9 August	Installation of pier settlement system
10 August	Lowering of pier, 20 mm
12 August	Lowering of pier, 40 mm
17 August	Lowering of pier, 80 mm
18 August	Lowering of pier, 95 mm
19 August	Lifting of pier, tilt of foundation
20 August	New reference condition
25 August	Spalling of concrete at soffit, 12 m^2
26 August	Spalling of concrete at soffit, 24 m^2
27 August	Landslide of 1 m at abutment
31 August	Failure of concrete hinge
2 September	Failure of 2 anchor heads
3 September	Failure of 4 anchor heads
7 September	Rupture of 2 out of 16 tendons
8 September	Rupture of 4 out of 16 tendons
9 September	Rupture of 6 out of 16 tendons

From a total of 4107 samples considered, the first 3253 come from the reference structure, and the last 854 (starting from August 9) were obtained from the anomalous scenarios. Among the reference samples, in the present work, the first 2379 are used as the baseline, namely the "training" dataset (to define the mean vector and the covariance matrix used in the Mahalanobis distance), and the rest 874 samples were considered for testing purposes. Fig. 16b shows the DIs of the entire dataset, calculated with respect to the

baseline.

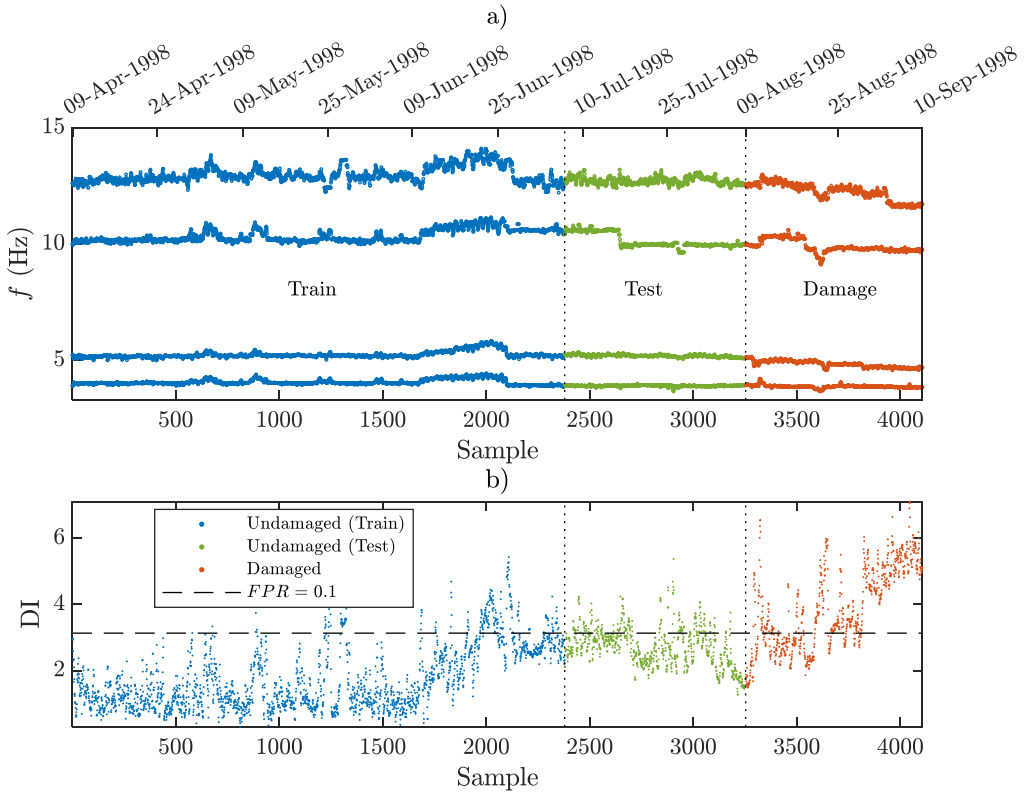


Fig. 16: Dataset processed for the Z24 bridge: a) Natural frequencies, b) DI distribution.

The BDC algorithm is then applied using the parameters given in Tab. 3. This table also reports the main results of the BDC performance analyses, depicted in Fig. 17 for the entire monitoring period and a close-up interval at the anomaly occurrence.

Table 3: Parameters and results summary of the Z24 bridge.

Parameter	**Value**
PL	1×10^{-5}
$w_{\max}$	15
FPR on baseline	0.1
$\tau(\text{FPR} = 0.1)$ on baseline	3.1304
Resulting FPR on test dataset	0.2311
PFA* on test dataset	0.1095
TPR on damaged dataset	0.6663
POD* on damaged dataset	0.5899

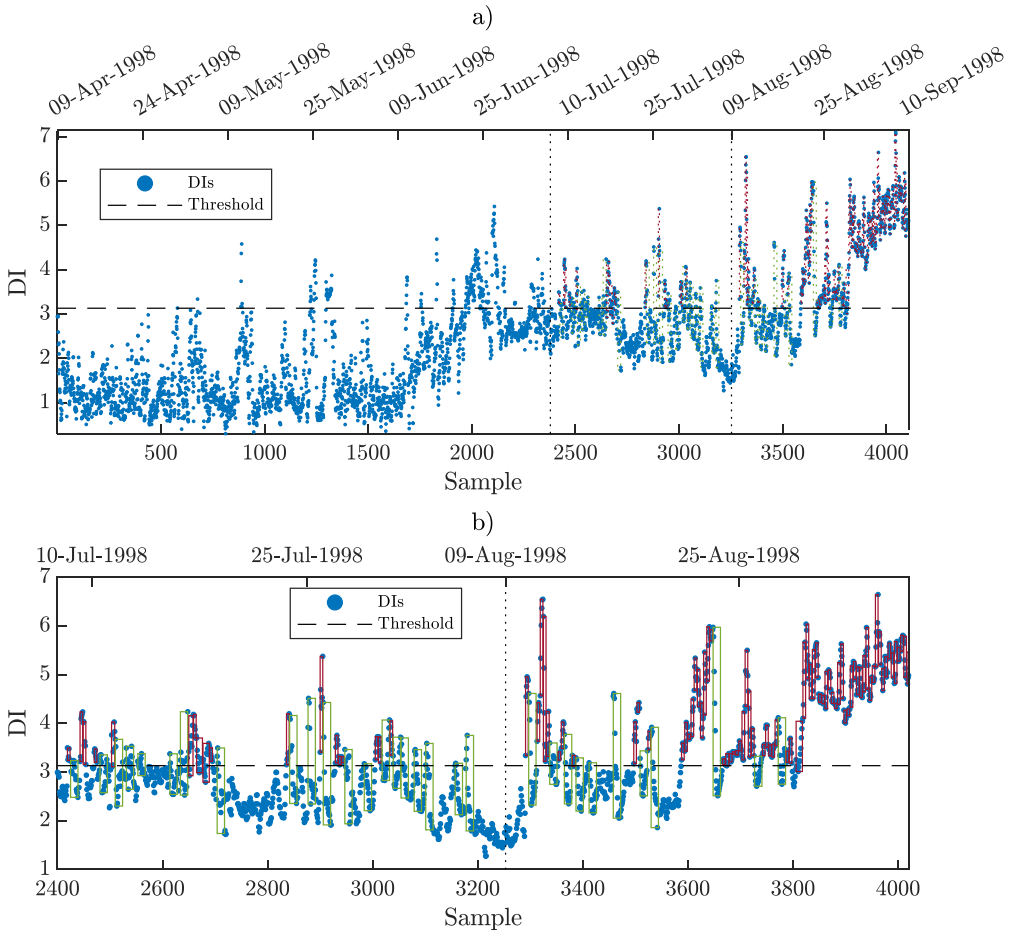


Fig. 17: DI computed over time and application of the BDC algorithm to the Z24 bridge using $w_{max} = 15$ and PL $= 10^{-5}$: a) Entire monitoring period; b) Close-up of the interval when the anomaly occurs

It is worth mentioning that in order to challenge the BDC algorithm, no prior temperature compensation was applied to the database. For this reason, it can be observed that the statistics of the "testing" period – in terms of mean value and variation – are different from the baseline, while they are both related to the reference state (see Fig. 17), thus resulting in a different FPR = 0.2311 (while FPR = 0.1 on the baseline). Nevertheless, the BDC algorithm is still capable of obtaining a lower $PFA^* = 0.1095$. It is also worth mentioning that the obtained $POD^* = 0.5899$ is close to TPR = 0.6663 for the damaged samples.

5. Conclusions

This study presented an approach for damage detection aimed at minimizing false alarms without compromising the ability to detect anomalies. The approach was applied here to the context of vibration-based structural health monitoring, where the damage index is built on the frequencies of vibration of the structures.

The method considers two distinct states, namely "alert" and "alarm," and a Binomial Distribution Classifier (BDC) to step from the former to the latter based on the concept of Binomial probability distribution. An "effective probability of false alarm" (PFA^*) and "effective probability of detection" (POD^*) were defined, and a closed-form formula was presented for their calculation. Moreover, the criteria for selecting the algorithm parameters were discussed based on the requirements of the structural manager/owner.

Both numerically simulated studies and experimental data clearly indicated that the BDC achieved a significantly lower PFA^* compared to traditional threshold-based algorithms used in most monitoring applications, highlighting the possibility of improving the performance of SHM systems in reducing false alarms while maintaining a high detection ability.

References

[1] Z. He, W. Li, H. Salehi, H. Zhang, H. Zhou, P. Jiao, Integrated structural health monitoring in bridge engineering, Automation in Construction 136 (2022) 104168.

[2] A. B. Noel, A. Abdaoui, T. Elfouly, M. H. Ahmed, A. Badawy, M. S. Shehata, Structural health monitoring using wireless sensor networks: A comprehensive survey, IEEE Communications Surveys & Tutorials 19 (3) (2017) 1403–1423.

[3] G. Webb, P. J. Vardanega, C. R. Middleton, Categories of shm deployments: technologies and capabilities, Journal of Bridge Engineering 20 (11) (2015) 04014118.

[4] R. Brincker, C. Ventura, Introduction to operational modal analysis, John Wiley & Sons, 2015.

[5] C. Rainieri, G. Fabbrocino, Operational modal analysis of civil engineering structures, Springer, New York 142 (2014) 143.

[6] S. Kamali, M. A. Hadianfard, Spectral optimization-based modal identification: A novel operational modal analysis technique, Mechanical Systems and Signal Processing 198 (2023) 110445.

[7] A.-M. Yan, G. Kerschen, P. De Boe, J.-C. Golinval, Structural damage diagnosis under varying environmental conditions—part i: A linear analysis, Mechanical Systems and Signal Processing 19 (4) (2005) 847–864.

[8] M. Jahangiri, A. Palermo, S. Kamali, M. A. Hadianfard, A. Marzani, A procedure to estimate the minimum observable damage in truss structures using vibration-based structural health monitoring systems, Probabilistic Engineering Mechanics 73 (2023) 103451.

[9] A. Kita, N. Cavalagli, F. Ubertini, Temperature effects on static and dynamic behavior of consoli palace in gubbio, italy, Mechanical Systems and Signal Processing 120 (2019) 180–202.

[10] M. Seif, L. Choe, J. Gross, W. E. Luecke, J. A. Main, D. McColskey, F. Sadek, J. M. Weigand, C. Zhang, Temperature-dependent material modeling for structural steels: formulation and application, US Department of Commerce, National Institute of Standards and Technology, 2016.

[11] M. A. Wahab, G. De Roeck, Effect of temperature on dynamic system parameters of a highway bridge, Structural Engineering International 7 (4) (1997) 266–270.

[12] X. Hua, Y. Ni, J. Ko, K. Wong, Modeling of temperature–frequency correlation using combined principal component analysis and support vector regression technique, Journal of Computing in Civil Engineering 21 (2) (2007) 122–135.

[13] H. Zhou, Y. Ni, J. Ko, Structural damage alarming using auto-associative neural network technique: Exploration of environment-tolerant capacity and setup of alarming threshold, Mechanical Systems and Signal Processing 25 (5) (2011) 1508–1526.

[14] C. R. Farrar, K. Worden, Structural health monitoring: a machine learning perspective, John Wiley & Sons, 2012.

[15] F. Falcetelli, N. Yue, R. D. Sante, D. Zarouchas, Probability of detection, localization, and sizing: The evolution of reliability metrics in structural health monitoring, Structural Health Monitoring 21 (6) (2022) 2990–3017.

[16] F. Falcetelli, N. Yue, R. Di Sante, D. Zarouchas, Probability of detection, localization, and sizing: The evolution of reliability metrics in structural health monitoring, Structural Health Monitoring 21 (6) (2022) 2990–3017.

[17] W. Meeker, D. Roach, S. Kessler, Statistical methods for probability of detection in structural health monitoring, IOVA State University (2019).

[18] Y. Ni, Y. Wang, C. Zhang, A bayesian approach for condition assessment and damage alarm of bridge expansion joints using long-term structural health monitoring data, Engineering Structures 212 (2020) 110520.

[19] M. H. Soleimani-Babakamali, R. Sepasdar, K. Nasrollahzadeh, R. Sarlo, A system reliability approach to real-time unsupervised structural health monitoring without prior information, Mechanical Systems and Signal Processing 171 (2022) 108913.

[20] D. Martucci, M. Civera, C. Surace, Bridge monitoring: Application of the extreme function theory for damage detection on the i-40 case study, Engineering Structures 279 (2023) 115573.

[21] D. Toshkova, M. Asher, P. Hutchinson, N. Lieven, Automatic alarm setup using extreme value theory, Mechanical Systems and Signal Processing 139 (2020) 106417.

[22] H. Sarmadi, A. Karamodin, A novel anomaly detection method based on adaptive mahalanobis-squared distance and one-class knn rule for structural health monitoring under environmental effects, Mechanical systems and signal processing 140 (2020) 106495.

[23] H. Sarmadi, A. Entezami, C. De Michele, Probabilistic data self-clustering based on semi-parametric extreme value theory for structural health monitoring, Mechanical Systems and Signal Processing 187 (2023) 109976.

[24] K. Maes, G. Lombaert, Monitoring railway bridge kw51 before, during, and after retrofitting, Journal of Bridge Engineering 26 (3) (2021) 04721001.

[25] J. Maeck, G. De Roeck, Description of z24 benchmark, Mechanical Systems and Signal Processing 17 (1) (2003) 127–131.

[26] Y. Xia, H. Hao, J. M. Brownjohn, P.-Q. Xia, Damage identification of structures with uncertain frequency and mode shape data, Earthquake engineering & structural dynamics 31 (5) (2002) 1053–1066.

[27] Z. Ding, J. Li, H. Hao, Z.-R. Lu, Structural damage identification with uncertain modelling error and measurement noise by clustering based tree seeds algorithm, Engineering Structures 185 (2019) 301–314.

[28] P. Van Overschee, B. De Moor, Subspace identification for linear systems: Theory—Implementation—Applications, Springer Science & Business Media, 2012.

[29] E. Reynders, G. De Roeck, Vibration-based damage identification: the z24 benchmark (2014).

[30] E. Reynders, G. Wursten, G. De Roeck, Output-only structural health monitoring in changing environmental conditions by means of nonlinear system identification, Structural Health Monitoring 13 (1) (2014) 82–93.

[31] E. Favarelli, A. Giorgetti, Machine learning for automatic processing of modal analysis in damage detection of bridges, IEEE Transactions on Instrumentation and Measurement 70 (2020) 1–13.