



City Research Online

City St George's, University of London

Citation: Quiroz, M. E., Kuncham, E., Ponnusami, S. A. & Bravo-Haro, M. A. (2026). Data-Driven Damage Detection on a Temperature-Varying Turbine Blade Using Multiple Sensor Modalities. Structural Control and Health Monitoring, 2026(1), 7943746. doi: 10.1155/stc/7943746

This is the accepted version of the paper.

This version of the publication may differ from the final published version. To cite this item please consult the publisher's version.

Permanent repository link: <https://openaccess.city.ac.uk/id/eprint/37515/>

Link to published version: <https://doi.org/10.1155/stc/7943746>

Copyright and Reuse: Copyright and Moral Rights remain with the author(s) and/or copyright holders. Copies of full items can be used for personal research or study, educational, or not-for-profit purposes without prior permission or charge, unless otherwise indicated, provided that the authors, title and full bibliographic details are credited, a hyperlink and/or URL is given for the original metadata page and the content is not changed in any way. For full details of reuse please refer to [City Research Online policy](#).

DATA-DRIVEN DAMAGE DETECTION ON A TEMPERATURE-VARYING TURBINE BLADE USING MULTIPLE SENSOR MODALITIES

María E. Quiroz¹, Eshwar Kuncham¹, Sathiskumar A. Ponnusami¹, Miguel A. Bravo-Haro^{1,2}

*Corresponding author: miguel.bravo-haro@city.ac.uk

¹City St George's, University of London, UK

²University of Cambridge, UK

Abstract

This study investigates vibration-based structural health monitoring (SHM) of a small-scale wind turbine blade using multiple sensor types to evaluate the performance of an output-only, data-driven, and semi-supervised damage detection framework. The primary objective is to assess how different sensing modalities influence damage detection and localisation, and to demonstrate the robustness of the proposed SHM approach across diverse sensor configurations in a wind turbine composite material blade. To this end, experimental accelerometer and strain-gauge time series are considered in this study. The method integrates multivariate autoregressive (MAR) modelling, principal component analysis (PCA), Mahalanobis Distance-based dissimilarity metrics, and K-means clustering. This hybrid framework requires no labelled damage data and minimal user input, offering a transparent, interpretable, and computationally efficient solution suitable for real-world applications. The blade was experimentally tested under progressive damage, which was introduced as sequential cracks at three locations, and under varying temperature conditions. Results show that the method reliably detects damage across sensor configurations, highlighting its potential as a practical and cost-effective tool for continuous SHM. Notably, both sensor types were able to detect damage using only one sensor as input in the model, however strain gauges proved more effective than accelerometers for damage localisation.

Keywords: *Structural health monitoring, Data-driven methods, Semi-supervised learning, Damage detection, Sensor modalities*

1. Introduction

Wind turbine blades (WTBs) are among the most critical and costly components of a wind energy system, both in terms of manufacturing and maintenance (Carroll et al., 2016; Dao et al., 2019). Their failure can lead to significant operational losses. Common damage types in WTBs—such as surface and internal cracks, debonding, and delamination—are especially concerning given the composite, anisotropic, and heterogeneous nature of blade materials, which complicates both detection and prediction of damage progression.

Structural Health Monitoring (SHM) aims to address these challenges by enabling early detection and prognosis of damage through sensor-based surveillance systems (Du et al., 2020; Li et al., 2015). While several non-destructive testing (NDT) techniques—including acoustic emission, ultrasonic testing, machine vision, and vibration-based methods—have been proposed for blade inspection, their widespread application to WTBs is limited due to accessibility issues, complex geometries, and the influence of variable environmental and operational conditions (Han et al., 2021; Sohn, 2007). As a result, hybrid SHM approaches that combine multiple methods are increasingly recommended to improve robustness and fault-detection accuracy (Du et al., 2020).

Among data-driven approaches, vibration-based SHM methods have shown great promise by using operational response data to infer structural condition (Eltoumy et al., 2023). Autoregressive (AR) models and their variants (e.g., ARMA, ARX) are widely used in this context as damage-sensitive feature extractors, generally in the time-domain (Entezami et al., 2020; Entezami & Shariatmadar, 2018; Gul & Necati Catbas, 2009; Meixedo et al., 2021, 2022; Panagiotopoulos et al., 2021). However, univariate AR models treat each sensor independently and do not account for cross-sensor interactions that naturally arise in multi-sensor SHM systems. In contrast, Multivariate Autoregressive (MAR) models (also referred to as vector autoregressive models in the literature) provide a natural extension by capturing temporal dependencies and spatial correlations between multiple sensor signals. These inter-sensor correlations can enhance both the sensitivity to damage and the ability to localise it, particularly in distributed sensing networks. As a result, MAR models often outperform their univariate counterparts in detecting subtle structural changes, especially in cases where damage manifests globally across multiple locations. Despite their advantages, most existing applications of MAR models have focused on relatively simple structures such as beams, steel truss bridges, and small-scale buildings (Achilli et al., 2021; Goi & Kim, 2017; Mosavi et al., 2012). Their application to more complex, composite systems, particularly wind turbine blades, remains limited. Moreover, the influence of environmental variability and the integration of different sensing modalities in MAR-based SHM frameworks are still underexplored.

While studies like Goi & Kim (2017) have shown that MAR models outperform univariate AR models when applied to accelerometer data, the generalisability of this improvement to other sensor types, such as strain gauges, is not yet well understood. Moreover, few SHM studies have systematically compared the effectiveness of different sensing modalities within vibration-based frameworks. Notably, Yang et al. (2015) proposed a frequency response transmissibility analysis using both strain gauges and accelerometers in a lab setting, though without considering environmental effects. As Li et al. (2015), emphasised, both sensor types represent cost-effective and technologically mature options for SHM, each with distinct advantages and limitations depending on the application context. However, they are also susceptible to electrical disturbances and noise, underscoring the need for careful evaluation of sensing strategies in real-world conditions.

In this study, we use experimental data from a 1.75 m small-scale composite WTB tested by Ou et al. (2021) to evaluate damage detection and localisation under varying temperature conditions. The damage was simulated in the experiments as cracks of increasing length (5, 10, and 15 cm) at three locations along the blade's longitudinal axis (17%, 30%, and 50%). Previous work on this dataset employed deep learning methods such as variational autoencoders (Z. Yang et al., 2024) and neural extended Kalman filters (Liu et al., 2024). While these studies demonstrated the potential of advanced data-driven techniques, they required large sensor arrays, five strain gauges and eight accelerometers respectively, which may not be practical for many SHM deployments. Furthermore, these approaches did not explicitly address temperature variability and overlooked progressive damage scenarios, instead performing detection only against the initial undamaged baseline. Such constraints limit their applicability to real-world continuous monitoring, where the current structural state may already be damaged.

To overcome these challenges, the present work develops a hybrid, semi-supervised SHM framework that compares the performance of accelerometers and strain gauges while tracking changes sequentially between damage stages and accounting for temperature variability in both the training and testing phases. The methodology integrates MAR and AR models for feature extraction, principal component analysis (PCA) for temperature compensation, and Mahalanobis distance for sensor-wise damage quantification, enabling progressive and environmentally robust damage detection in composite WTBs. However, given that the proposed damage detection framework relies on extracted strain or acceleration response features without assumptions on sensing modality, it is well-suited for integration with other sensing technologies. Most noticeably, fibre-optic cables for these also measure delta strains (Sierra-Pérez et al., 2013, 2016).

To the best of the authors' knowledge, this is the first study to apply multivariate feature extraction techniques for SHM of composite blades under environmental variation using both accelerometers and strain gauges. The main contributions of this work are: (i) Develop and evaluate a multivariate feature extraction method for detecting and localising damage in composite wind turbine blades and comparing it with univariate approaches. (ii) Comparing the performance of the methodology across different sensor networks, including accelerometers and strain gauges. (iii) Accounting for environmental variability, specifically temperature changes, in the damage detection framework. (iv) Enable progressive damage detection, comparing each damage state with the blade's immediately preceding structural condition, including already damaged states.

2. Description of the experimental wind turbine blade benchmark

The experimental small-scale wind turbine blade benchmark consists of a 1.75 m long and 5.0 kg blade part of a Windspot 3.5 kW wind turbine model. It has a layered sandwich structure, consisting of: (i) a 0.93 mm-thick outer shell of double-layered composite material, composed of plain-weave fabric and chopped strand mat made of E-glass fibres, and (ii) a low-stiffness polyurethane foam core for internal reinforcement. This small-scale blade was tested as part of an experimental campaign by Ou et al. (2021) for the development of damage detection algorithms. The experiment setup, temperature variations and damage considerations are briefly described in the next paragraphs, and a general overview is provided in **Figure 1**.

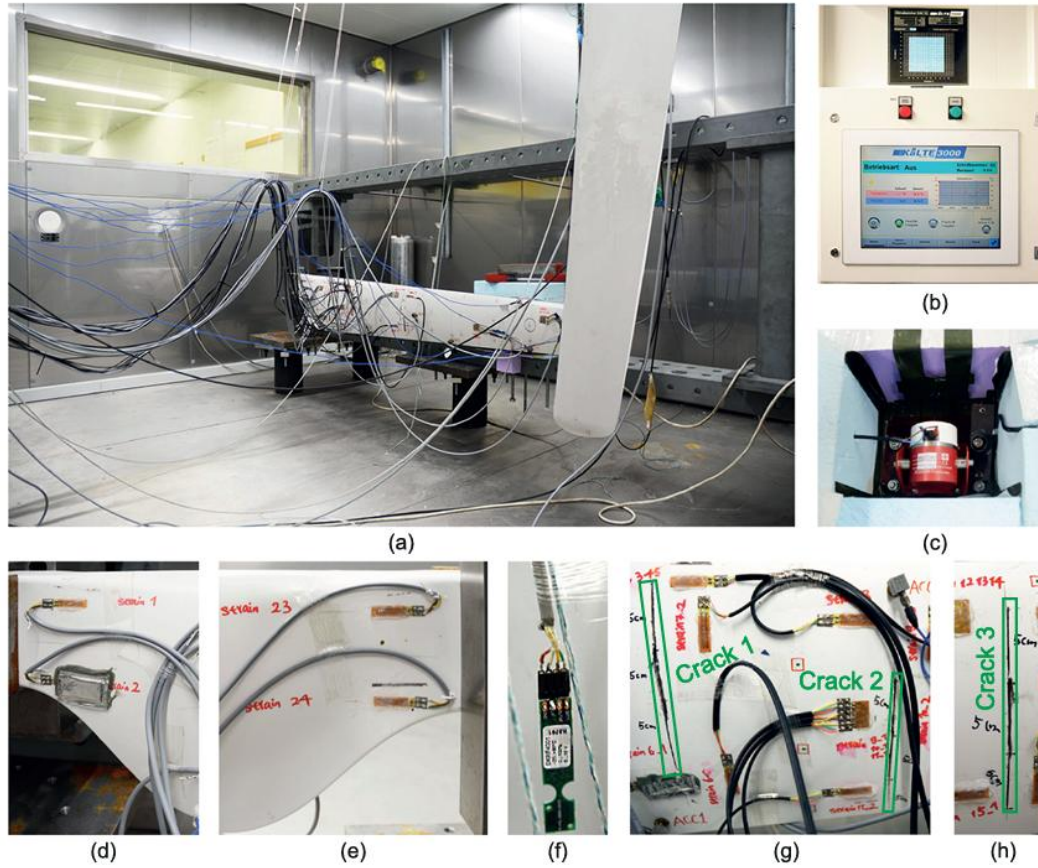


Figure 1: Summary of the experimental campaign elements: (a) blade inside the climate chamber; (b) climate controller; (c) electromechanical shaker; (d–e) strain gauges on the low- and high-pressure sides; (f) climate sensor; (g–h) introduced cracks (Ou et al., 2021).

The blade was fixed at one end using a steel frame secured by four bolts, subjected at the same time by three concrete columns. The tests were conducted in a climate-controlled chamber, maintaining a constant humidity of 60%, while temperature varied between -15°C to $+40^{\circ}\text{C}$. Experimental dynamic tests were carried out for measuring the different responses of the blade. The blade was excited using an electromechanical shaker, applying two types of excitations: (i) white noise signals with an effective frequency range of 0-400 Hz and (ii) sine sweeps with frequencies from 1-300 Hz. Noticeably, this pseudo white noise was intended to mimic a scenario where Operational Modal Analysis (OMA) techniques can be applied for these require a broadband stationary excitation, usually referred to as “ambient” excitation. The laboratory setup included a total of eight PCB Piezotronics accelerometers ($\pm 10\text{g}$ working range) and 21 strain gauges arranged unidirectionally and in form of rosettes, as shown in **Figure 2**. Two distinct strain gauges configurations were tested, each one indicated through the superscripts (1) and (2) in **Figure 2**, setting them along the different axes in each case.

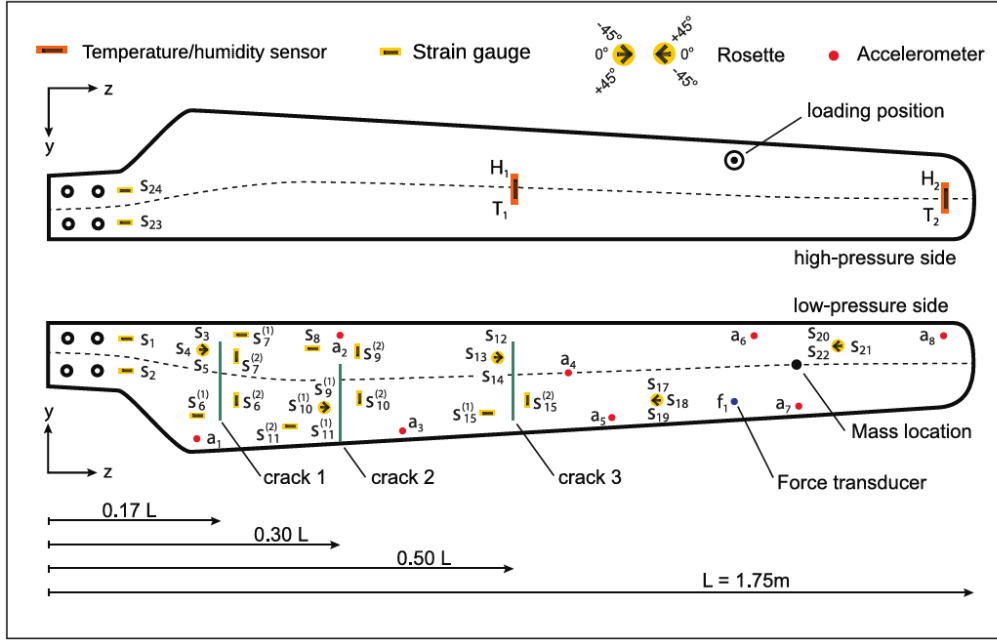


Figure 2: Experimental setup of accelerometers and strain gauges on the tested blade (Ou et al., 2021).

Progressive damage was introduced into the blade in the form of cracks, induced as surface cuts, as outlined in **Table 1**. Additionally, point masses were added to simulate icing loads; however, these are not considered in this analysis, as the primary focus is on structural damage. The healthy state is denoted as *R*, while damaged cases (*D-L*) correspond to increasing crack lengths and additional crack's introduction. Each crack had a uniform depth of 4 mm and width of 1.5 mm. In the healthy (undamaged) state, a total of 20 white noise tests and one sine sweep excitation performed, while each damage case was tested with one sine sweep and five white noise excitations, while temperature varied between -15°C to +40°C in 5°C increments for each case. Data was sampled at 1666.7 Hz in 2-minutes windows.

Table 1: Damage configuration tested by Ou et al. (2021).

Case	Crack 1	Crack 2	Crack 3	# of Experiments
R (Undamaged)	-	-	-	504 (20 WN + 1 SS)
D	5 cm	-	-	144 (5 WN + 1 SS)
E	5 cm	5 cm	-	144 (5 WN + 1 SS)
F	5 cm	5 cm	5 cm	144 (5 WN + 1 SS)
G	10 cm	5 cm	5 cm	144 (5 WN + 1 SS)
H	10 cm	10 cm	5 cm	144 (5 WN + 1 SS)
I	10 cm	10 cm	10 cm	144 (5 WN + 1 SS)
J	15 cm	10 cm	10 cm	144 (5 WN + 1 SS)
K	15 cm	15 cm	10 cm	144 (5 WN + 1 SS)
L	15 cm	15 cm	15 cm	144 (5 WN + 1 SS)

WN: White noise, SS: sine sweep.

3. Damage detection algorithm

The damage detection algorithm is an output-only, data-driven, semi-supervised approach with four main steps following the collection of acceleration data from sensors installed along the structural

element under monitoring: (i) feature extraction, (ii) feature modelling, (iii) feature fusion, and (iv) feature discrimination and clustering, as schematised in **Figure 3**. The proposed algorithm builds upon the methods introduced in Meixedo et al. (2021) which relied exclusively on AR models, but significantly extending them to account for the interdependency of measured signals by employing MAR models for feature extraction. This approach leverages the correlation between multiple signals (i.e., multiple sensors in the network of the installed monitoring system) to describe the behaviour of the system, detecting damage through variations in their correlation rather than changes in individual signal behaviour.

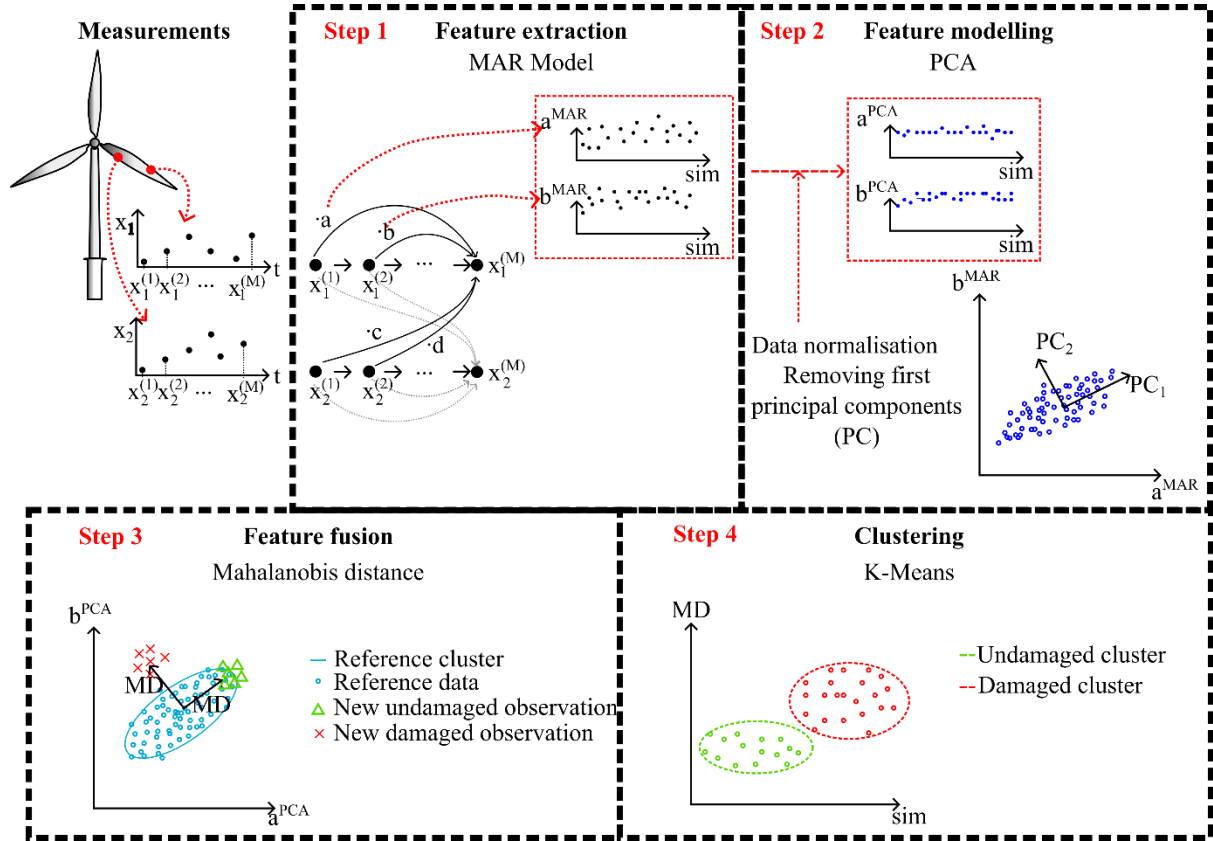


Figure 3: Flowchart of the proposed damage detection algorithm.

The MAR predicts the next value in a D -dimensional time series, y_i , as a linear combination of the N previous vector values. The model is expressed as follows:

$$y_i = \sum_{j=1}^N y_{i-j} A_j + e_i \quad (1)$$

where D represents the number of sensors, N is the MAR model order, $y_i = [y_i^{(1)}, y_i^{(2)}, \dots, y_i^{(D)}]$ is the i th sample of a D -dimensional time series, $A_j \in \mathbb{R}^{D \times D}$ are coefficient matrices capturing interdependencies between signals, and e_i is the residual error vector. The coefficients A_j are estimated using the least-square method, minimizing the residual error across all samples. In contrast, univariate autoregressive (AR) models operate on scalar time series. In this case, each coefficient A_j reduces to a scalar a_j , and both y_i and e_i become scalar values, modelling only the temporal relationship within a single signal without accounting for cross-sensor interactions.

Environmental variability, such as changing wind and wave conditions, affects the structural response, demanding data normalisation. Principal Component Analysis (PCA) addresses this by filtering environmental effects without requiring additional data. PCA transforms standardised feature space X into orthogonal principal components based on the covariance matrix C , retaining components that capture the desired variance threshold:

$$X' = XV_m \quad (2)$$

where V_m contains the eigenvectors corresponding to the largest m eigenvalues, retaining a predefined variance threshold. To normalise the features, the first components associated with the environmental variability are removed (Santos et al., 2013). This reduces false positives caused by environmental changes when applying the damage detection algorithm (Tu et al., 2021). Nevertheless, the algorithm must still be trained with data covering a wide range of temperatures to ensure robustness.

Following normalisation, the Mahalanobis distance (MD) quantifies deviations of each sensor's behaviour from the baseline condition. The MD for a sensor is defined as:

$$MD_s = \sqrt{(x_s - \bar{x}) \cdot S_x^{-1} \cdot (x_s - \bar{x})^T} \quad (3)$$

Here, x_s is the vector of features for a given sensor, $\bar{x}$ and S_x are the mean vector and covariance matrix of baseline data, respectively. The MD values for all sensors are combined to compute the dissimilarity between simulations using the Euclidean distance:

$$dis_{ij} = \sqrt{\sum_{d=1}^D (MD_{s=i} - MD_{s=j})^2} \quad (4)$$

The dissimilarity matrix serves as input for a K-means clustering algorithm, partitions simulations into K clusters by minimizing the intra-cluster dissimilarity:

$$W(P_K) = \frac{1}{2} \sum_{k=1}^K \sum_{c(i)=k} \sum_{c(j)=k} dis_{ij} \quad (5)$$

where $c(i)$ assigns simulation i to cluster k . The algorithm iteratively assigns simulations to the nearest centroid and updates the centroids until convergence. The optimal number of clusters is determined using the silhouette score, enabling the method to operate autonomously and making it suitable for real-time monitoring applications. In real-time applications, as new structural conditions emerge, the baseline conditions can be updated to improve the detection of subsequent changes.

4. Model order selection

To determine the optimal model order for the AR and MAR models used in subsequent analyses, data from the healthy reference condition (state R) of the blade is employed as a reference. Models are fitted to this data, and performance is evaluated using the Normalised Root Mean Squared Error (NRMSE), which quantifies prediction error relative to the signal variance:

$$NRMSE = \sqrt{\frac{\sum_{i=1}^M (y_i - \hat{y}_i)^2}{\sum_{i=1}^M (y_i - \bar{y})^2}} \quad (6)$$

where y_i is the measured data, $\hat{y}_i$ is the simulated data from the AR or MAR model, $\bar{y}$ is the mean of the measured data, and M is the total number of samples in the time window.

For accelerometers data (**Figure 4**), the NRMSE values of AR and MAR models (with two sensors) are shown across different model orders. Note that, unlike AR models, the number of parameters in MAR models scales with the number of sensors; for instance, a MAR model of order 5 with two sensors yields 10 features in total. In both subplots, the error curves generally decrease until convergence at approximately $N \geq 30$, with MAR models consistently exhibiting lower median values than AR models, indicating superior accuracy in signal representation. A full convergence analysis across varying sensor counts is warranted but omitted here for brevity.

The same procedure applied to strain gauges (**Figure 5**) reveals a broadly similar trend, though with higher variability. In this case, both AR and MAR models converge at approximately $N \geq 40$. Despite the relatively higher fitting errors observed, filtering was intentionally omitted to preserve the high-frequency content of the signals. Preliminary testing showed that filtering substantially reduced the performance of the algorithm. These tests involved the application of a fourth-order low-pass Butterworth filter to remove high-frequency components potentially associated with measurement noise. The comparisons between filtered and unfiltered signals showed that filtering led to a noticeable reduction in the algorithm's accuracy. Although these results are not shown here for the sake of brevity, they suggest that high-frequency components may carry information relevant to subtle changes in structural dynamic behaviour—information that are often overlooked or attenuated in traditional OMA, which typically fails to resolve higher-order modes. Therefore, all results presented in this study are based on full-bandwidth signals sampled at 1666 Hz.

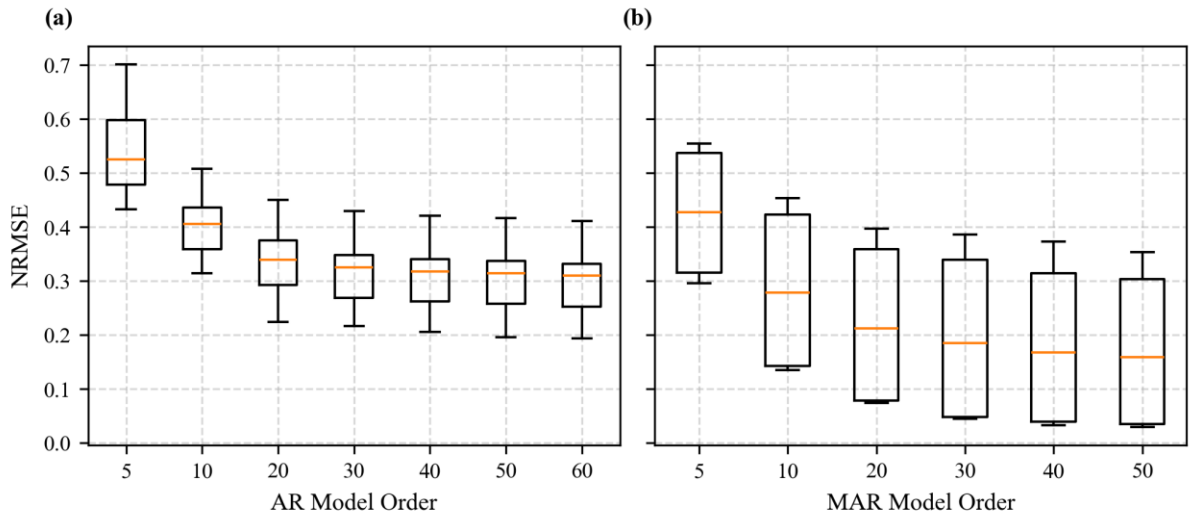


Figure 4: Boxplots of NRMSE values using accelerometers for (a) AR model orders and (b) MAR model orders with two sensors.

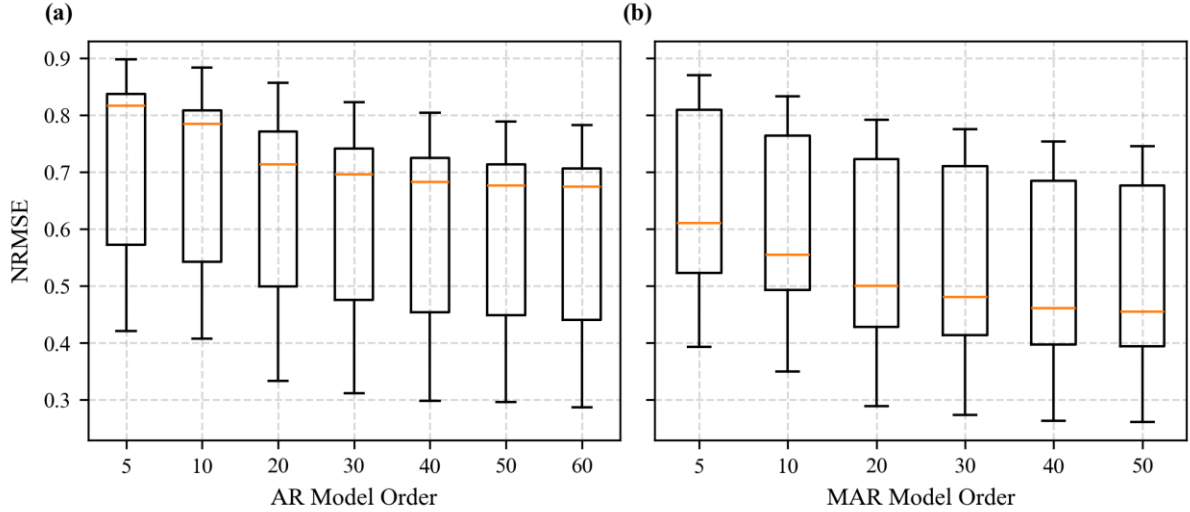


Figure 5: Boxplots of NRMSE values using strain gauges for (a) AR model orders and (b) MAR model orders with two sensors.

Although several methods for selecting model order have been proposed in the literature—such as the Akaike Information Criterion (AIC), the Bayesian Information Criterion (BIC), or direct analysis of model errors (Shumway & Stoffer, 2017)—no universally optimal method exists. These criteria primarily serve as heuristics to balance model accuracy with parsimony, aiming to minimise error while avoiding overfitting. Consequently, the selected model order may not correspond to the absolute best-performing solution for damage detection applications. Some attempts to overcome this issue have been presented in the literature (Hoell & Omenzetter, 2016; Long et al., 2013), but they still rely on a primary order selection based on AIC or NRMSE convergence. Since the primary goal of this paper is to evaluate the capabilities of the proposed methodology rather than to identify the optimal model order or the most informative features (a topic reserved for future work), the final choices are made based on trial and error based on the accuracy of the model and better algorithm’s performance.

5. Detectability

This subsection evaluates the detectability of progressive structural damage using time-series features derived from AR and MAR models applied to accelerometer and strain data. By analysing MD, dissimilarity measures, and clustering performance, we assess the sensitivity of these features to different damage stages. Both single- and multi-sensor configurations are examined, with a focus on quantifying how model complexity and sensor redundancy affect the ability to detect and track evolving damage conditions. The training of the damage detection algorithm is made with 75% of the baseline data, while testing is made with the rest 25% together with samples from the new damage condition, with both stages incorporating samples across all temperature conditions in the range of -15°C to $+40^{\circ}\text{C}$.

5.1. Damage detection based on one sensor measurements

Recent studies on the same benchmark dataset have explored deep learning-based approaches (Calderano et al., 2022; Liu et al., 2024; Z. Yang et al., 2024). For instance, Z. Yang et al. (2024) proposed a variational autoencoder with a neural ordinary differential equation, while Liu et al. (2024) employed neural extended Kalman filters. Although these methods achieved promising results, they require large sensor arrays and generally neglect environmental variability, such as temperature effects.

Moreover, they primarily focus on detecting deviations from the initial undamaged state without addressing the progression of damage. In contrast, the transition-based MD analysis presented here provides a framework for detecting incremental damage stages with fewer sensors while maintaining sensitivity to structural changes.

For visualisation, one accelerometer (a_8) and one strain gauge (s_{13}) are selected for the subsequent figures. Results are later extended to all sensors to allow a comprehensive performance comparison. An offline comparison is performed in first place. **Figure 6** presents a boxplot of Mahalanobis Distance (MD) values for all damaged cases, using the undamaged condition R as a baseline. **Figure 6** subplot (a) shows the results while using one accelerometer, while subplot (b) shows results using the single strain gauge. In both cases, MD values generally increase as the level of damage in the blade structure progresses. The most pronounced increase occurs during the first damage transition $R \rightarrow D$, whereas subsequent transitions show more gradual rises. This trend indicates that the MD metric effectively captures progressive structural degradation, particularly when assessed relative to the initial undamaged state.

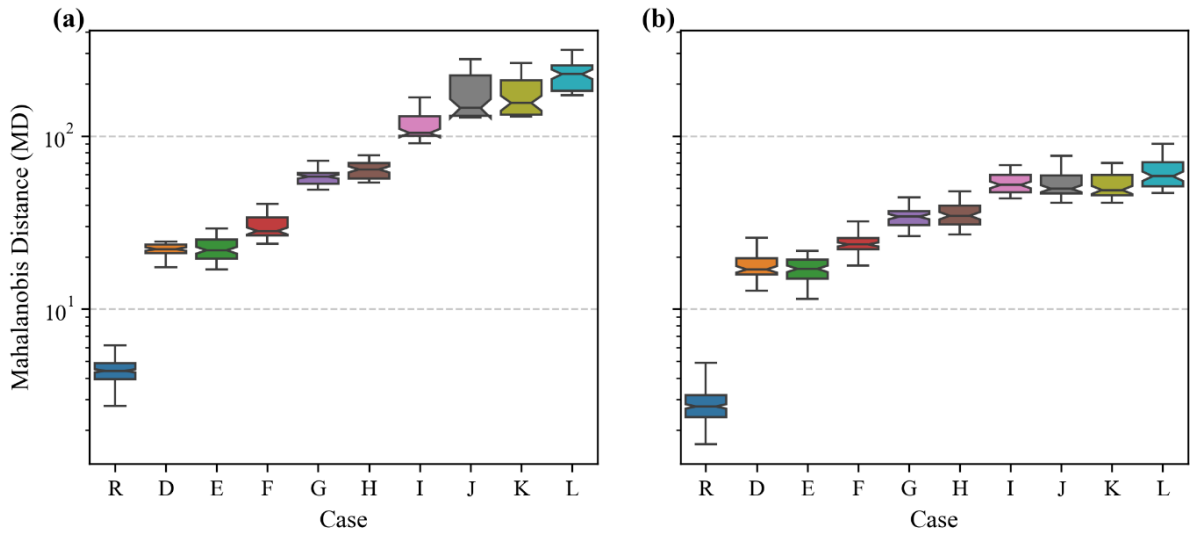


Figure 6: Boxplot of the MD values across all damage cases using AR-40 models for (a) accelerometer a_8 and (b) strain gauge s_{13} .

However, in practical scenarios, a damage detection algorithm aims to identify changes progressively. Therefore, each stage is assessed relative to its immediate predecessor rather than the original undamaged condition. This transition-based framework aligns more closely with real-time monitoring needs. **Figure 7** shows MD values for each damage subtransition, using the preceding state as the baseline: subplot (a) corresponds to the accelerometer, and subplot (b) to the strain gauge. Black bars represent the baseline (previous) condition resulting MD values, while coloured bars indicate the new damaged state MD values.

Among the transitions, $J \rightarrow K$ —one of the final stages, where the crack at 30% of the blade length is enlarged to 15 cm—produces the highest indices, indicating the greatest detectability. By contrast, the early stages $R \rightarrow D$, $E \rightarrow F$ (introducing 5 cm cracks at 17% and 50% of the blade's length, respectively) yield the lowest MDs, making them more challenging to detect with both sensor types. For both sensors, the closest crack is located at 50% of the blade length, yet the associated damage scenarios ($E \rightarrow F$, $H \rightarrow I$, $K \rightarrow L$) are not consistently easier to detect. This suggests that detectability is influenced more by the severity and progression of the damage, as well as experimental factors, than by simple proximity to the sensor.

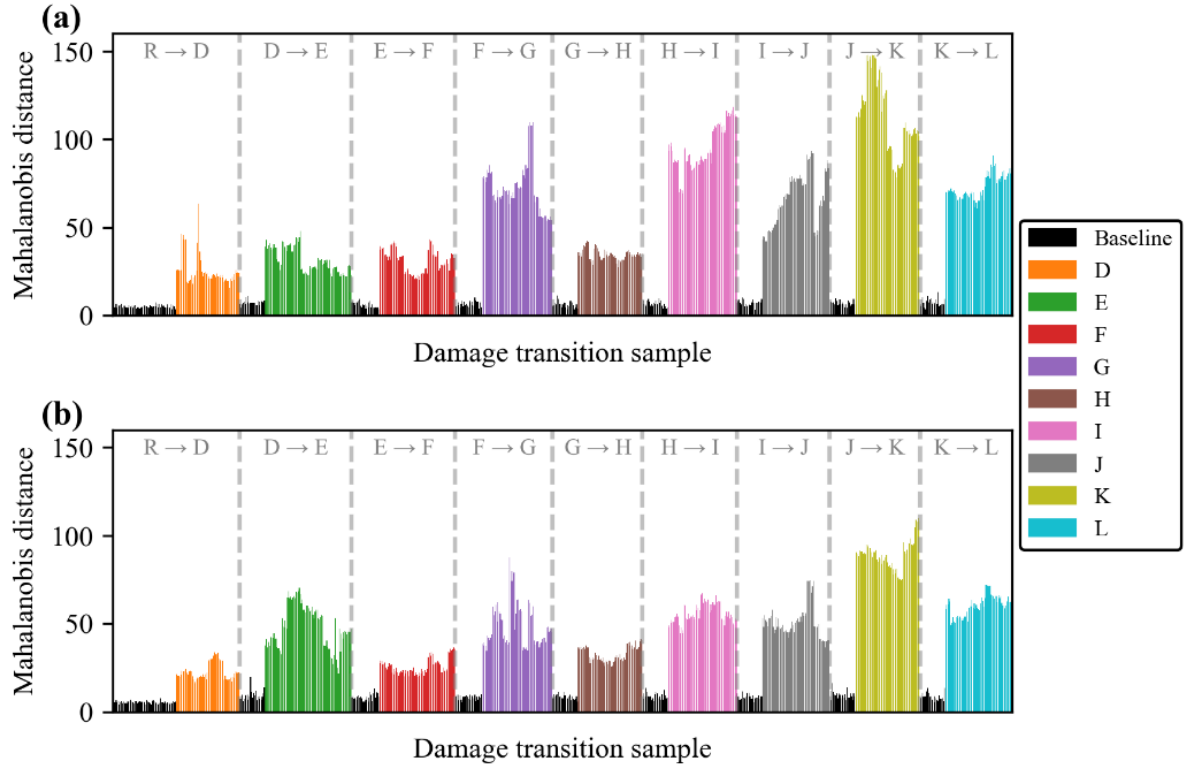


Figure 7: Mahalanobis distance across damage stage transitions using AR-40 models for (a) accelerometer a_8 and (b) strain gauge s_{13} .

The dissimilarity data exhibited an exponential trend and, consequently, was logarithmically scaled before being used as input for K-means clustering. The results, presented in **Figure 8**, show clear separation between two clusters corresponding to pre- and post-damage transition conditions, (a) for the accelerometer and (b) for the strain gauge. This indicates that even a single sensor - accelerometer or strain gauge - provide sufficient information to distinguish structural states. Similar performance was observed across other one-sensor configurations (not shown for brevity), highlighting the effectiveness of the approach even with minimal sensing.

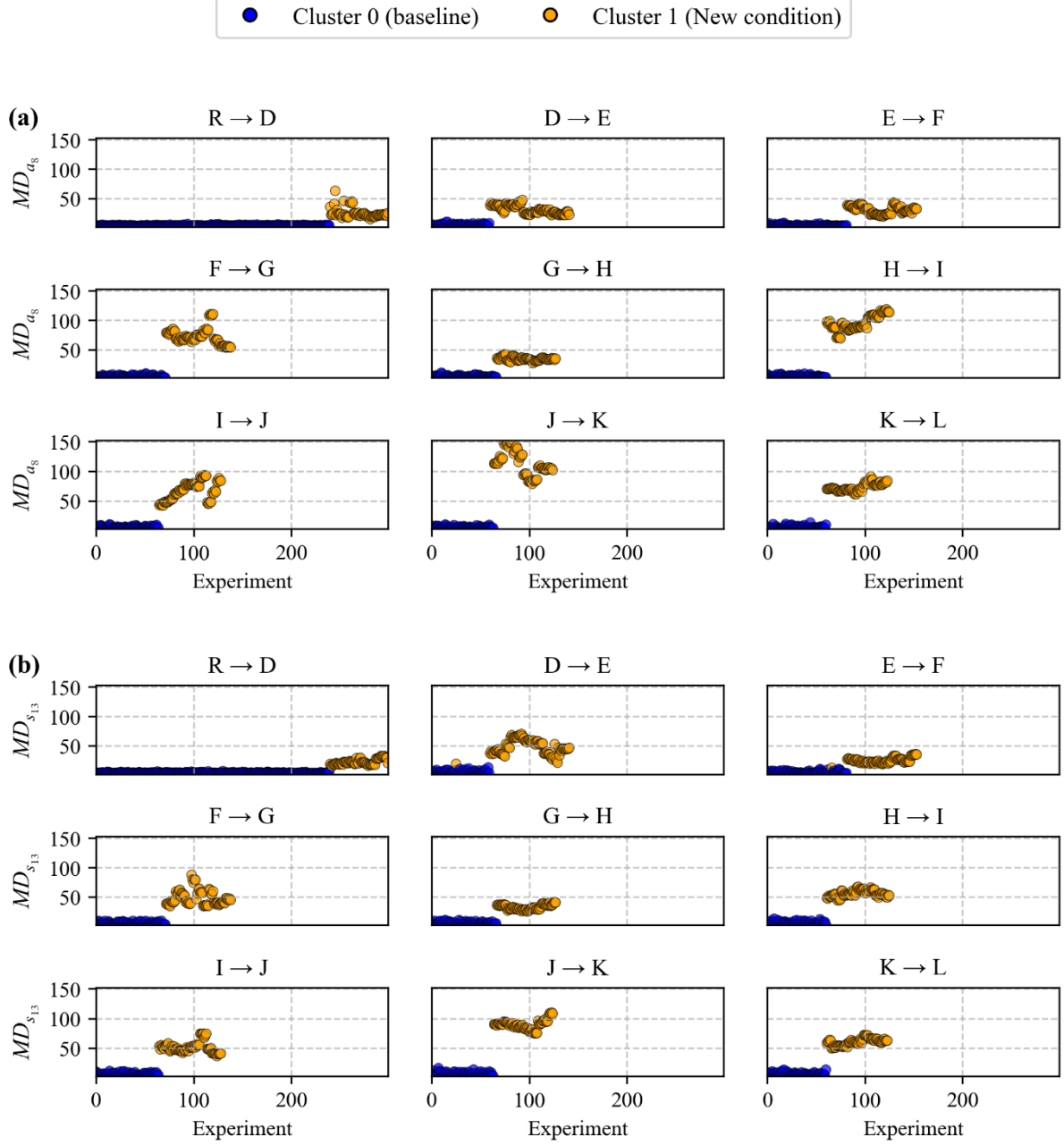


Figure 8: Clustering performance of K-means using AR-40 models for (a) accelerometer a_8 and (b) strain gauge s_{13} .

Mean specificity (E) and sensitivity (S), which are the probabilities of correctly identifying negative and positive cases, were computed over 20 random trials for AR-40 models using 25% test data from a single sensor at various locations. For clarity, **Figure 9** and **Figure 10** present the equivalent false positive rate (1-E) and false negative rate (1-S) for the accelerometers and strain gauges, respectively.

Accelerometer results show consistent detection of structural transitions, with only minor false positives during the $D \rightarrow E$ and $E \rightarrow F$ stages (approximately 10–11% error in both 1-S and 1-E), consistent with the low detectability observed in **Figure 7**. In contrast, strain gauge results show more frequent misclassifications, particularly for transitions $D \rightarrow E$, $E \rightarrow F$ and $H \rightarrow I$, with errors reaching up to 37% for 1-S and 30% for 1-E.

Strain gauge performance also exhibits a strong dependence on sensor placement: gauges located near the blade root yield poorer results, whereas those positioned farther away achieve more reliable detection. This trend may be explained by the larger deformations occurring farther from the root, which make structural changes more pronounced and thus easier to detect.

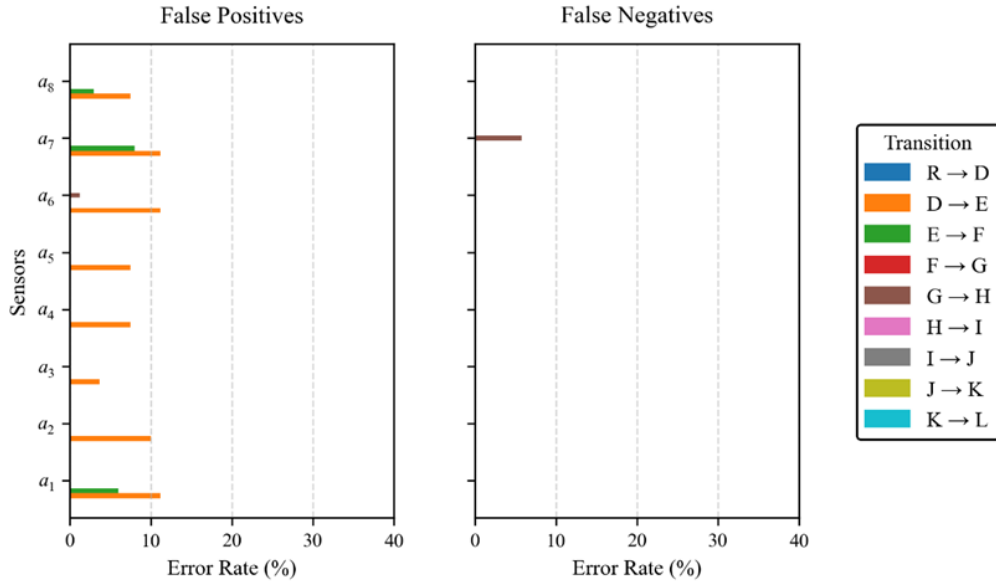


Figure 9: False positive ($1-E$) and false negative ($1-S$) rates across structural condition transitions using a single accelerometer with AR-40 models.

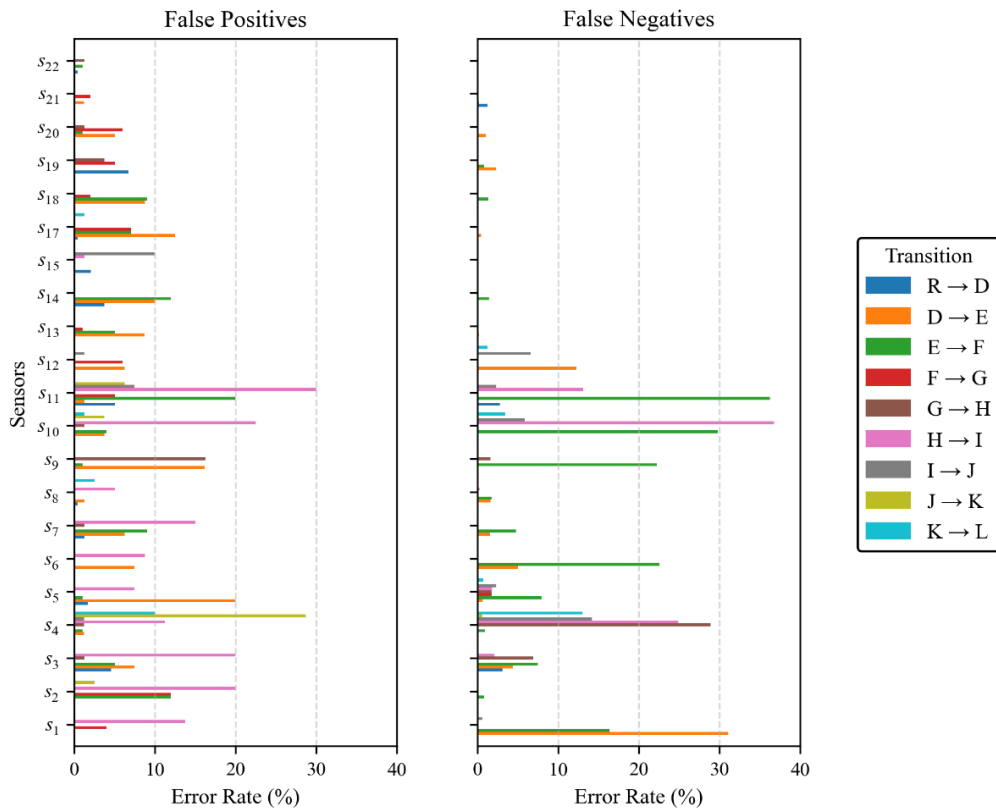


Figure 10: False positive ($1-E$) and false negative ($1-S$) rates across structural condition transitions using a single strain gauge with AR-40 models.

5.2. Two-sensor analysis

The algorithm's performance is next evaluated using two arbitrarily selected sensors. **Figure 11** compares the MDs computed from AR-40 and MAR-40 models for accelerometers a_1 and a_8 . The MAR model produces MD higher values than those from the AR model —up to 600% higher during the $J \rightarrow K$ transition (enlargement of the crack at 30% of the blade length to 15 cm) and up to 700% higher in early transitions such as $E \rightarrow F$ with first incorporations of cracks. These results highlight the improved feature quality achieved by incorporating inter-sensor correlations. For accelerometers, this demonstrates that exploiting correlations between multiple sensors substantially enhances the detectability of structural transitions.

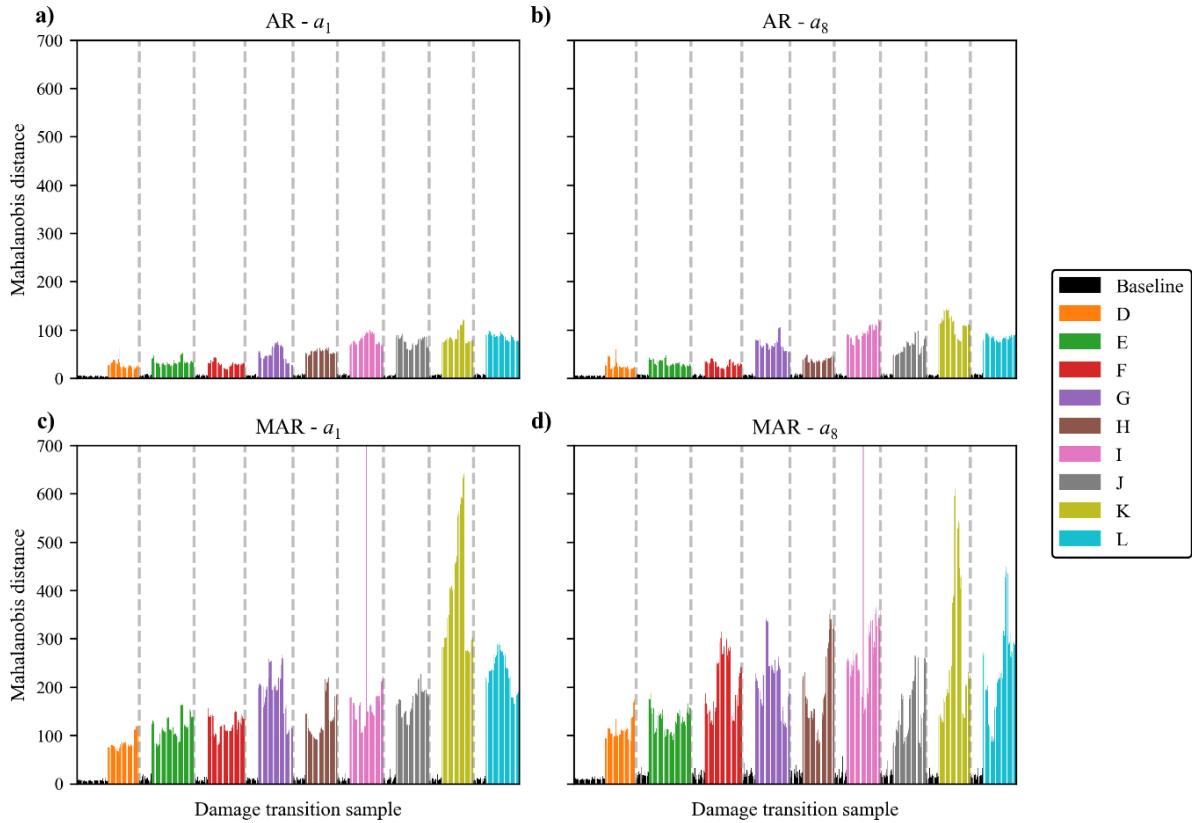


Figure 11: Mahalanobis distances between consecutive structural conditions for (a) AR-40 with accelerometer a_1 , (b) AR-40 with accelerometer a_8 , (c) MAR-40 with accelerometer a_1 , and (d) MAR-40 with accelerometer a_8 .

Figure 12 presents MDs computed using AR-40 and MAR-20 models for strain gauges s_4 and s_{13} . Unlike the accelerometer case, the MAR model does not significantly improve feature quality for strain gauges (excepting transition $F \rightarrow G$ for sensor s_{13} , by 500%). This suggests that inter-sensor correlations among strain gauges are less informative, likely because these sensors capture more localised strain effects compared to the broader vibration patterns (i.e., global dynamic) detected by accelerometers. This observation is further discussed in the following sections.

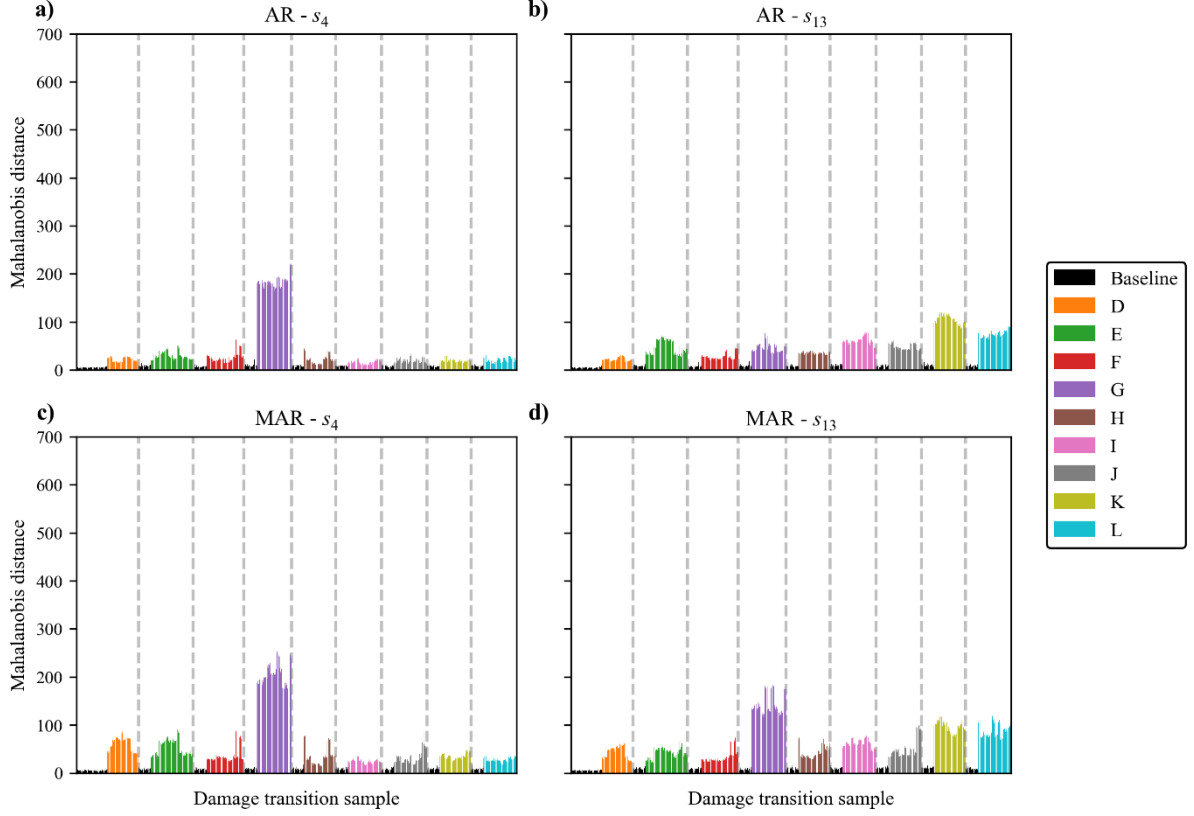


Figure 12: Mahalanobis distances between consecutive structural conditions for (a) AR-40 with strain gauge s_4 , (b) AR-40 with strain gauge s_{13} , (c) MAR-20 with strain gauge s_4 , and (d) MAR-20 with strain gauge s_{13} .

The improvement in feature quality is quantified using the change in mean dissimilarity of the experiments across damage stages, defined as:

$$\Delta \overline{\text{dis}} = \overline{\text{dis}}^{(z)} - \overline{\text{dis}}^{(z-1)} \quad (7)$$

where $z \in \{R, D, E, \dots, L\}$ denotes the structural condition. In this formulation, transitions are penalised when the MD values for the baseline condition are inflated, thus enhancing the robustness of the indicator.

The mean dissimilarity change $\Delta \overline{\text{dis}}$ across transitions is shown in **Figure 13** (a) for two accelerometers and (b) for two strain gauges, comparing AR and MAR models. This provides a more compact representation of the behaviours observed in **Figure 11** and **Figure 12**. For accelerometers, the MAR model consistently produces larger dissimilarity changes, indicating that incorporating inter-sensor correlations improves feature quality and enhances discrimination of structural transitions. Because similar levels of new damage are introduced at each stage, comparable differences between transitions would be expected; deviations from this trend likely reflect experimental variability. Nevertheless, the overall pattern supports the robustness of the dissimilarity metric. In contrast, the strain-gauge data show no clear trend to suggest that incorporating correlation information improves detectability within the algorithm.

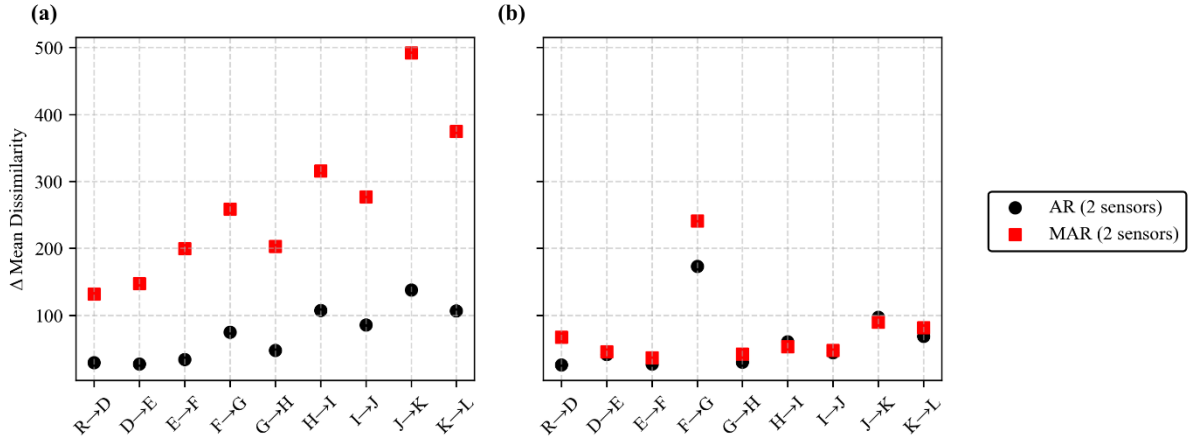


Figure 13: Change in mean dissimilarity for each structural transition, comparing results from (a) AR-40 and MAR-40 models with accelerometers a_1 and a_8 , and (b) AR-40 and MAR-20 models with strain gauges s_4 and s_{13} .

5.3. Dissimilarity vs. number of sensors

To investigate the influence of sensor count on performance, **Figure 14** presents the evaluation of $\overline{\Delta \text{dis}}$ for AR and MAR models as additional sensors are included: (a) accelerometers and (b) strain gauges. In the MAR models, the total number of features is fixed at 80 for accelerometers and 40 for strain gauges, regardless of sensor count. Each boxplot contains nine data points, corresponding to the values for each transition shown in **Figure 13** for the two-sensor case. For consistency, the same sensor combinations are used for both MAR and AR models; results may vary slightly with different combinations.

As shown in **Figure 14**, the mean dissimilarity increases with the addition of sensors for both models, but performance tends to plateau after 5–6 sensors in both scenarios (accelerometers and strain gauges), indicating diminishing returns in feature quality beyond this point. For accelerometers, MAR models consistently outperform univariate models when using two or more sensors. In contrast, for strain gauges no clear trend is observed—while MAR models show a slight performance gain with 2–12 sensors, the difference is not statistically significant. This plateau suggests that, in practical terms, deploying more than six sensors may offer limited diagnostic benefit relative to the added cost and complexity for damage detection in the analysed structure.

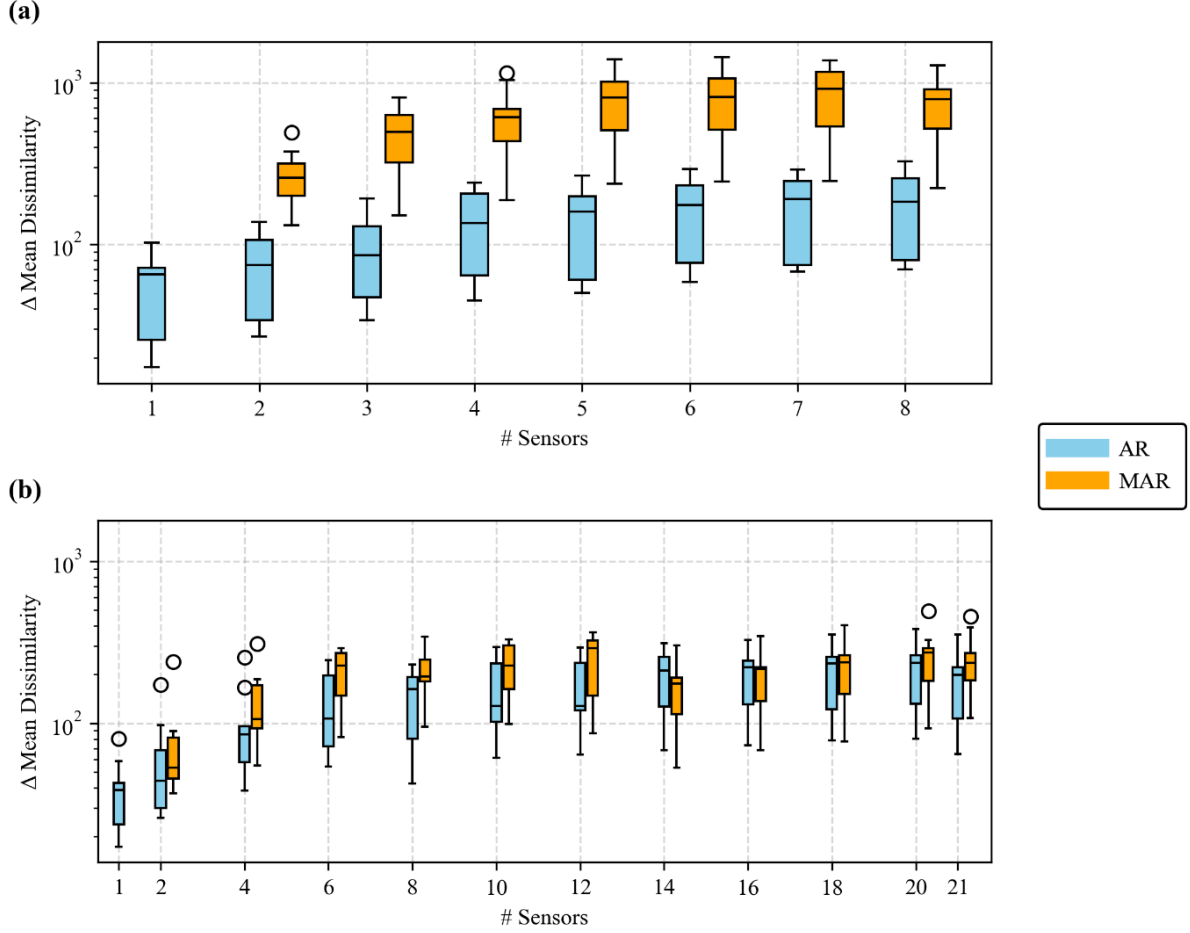


Figure 14: Boxplots of $\Delta \overline{\text{dis}}$ for AR-40 and MAR models across different numbers of sensors: (a) accelerometers and (b) strain gauges.

6. Locatability

This section examines the ability to localise structural damage using acceleration and strain gauge measurements from sensors distributed along the blade. MD values derived from AR and MAR models are analysed to determine whether the spatial distribution of damage can be inferred from sensor responses. Comparisons between model types and sensing modalities identify configurations that support accurate localisation across different damage transitions. For this analysis, all sensors along the blade are used to perform localisation. While results from AR models can be directly extrapolated to other sensor combinations due to their independence, the localisation capability of MAR models may vary with the number of sensors employed.

To evaluate the contribution of each sensor to damage detection across different damage stages, we compute the change in MD values between a damaged condition and its corresponding baseline (or previous condition) using the Fisher criterion (Bishop & Nasrabadi, 2006), as applied in Mosavi et al. (2012). This approach allows the locatability of damage signals to be assessed at the sensor level.

For each damage stage z , and each sensor i , the Fisher criterion $FC_i^{(z)}$ is computed as:

$$FC_i^{(z)} = \frac{(m_i^{(z)} - m_i^{(z-1)})^2}{\sigma_i^{(z)^2} + \sigma_i^{(z-1)^2}} \quad (8)$$

where:

- $m_i^{(z)}$ and $m_i^{(z-1)}$ are the means of the MD value for sensor i under the damaged condition at stage z and $z - 1$, respectively.
- $\sigma_i^{(z)}$ and $\sigma_i^{(z-1)}$ are the corresponding standard deviations.

This metric quantifies the increase in anomaly indication for each sensor due to damage progression. High values of $FC_i^{(z)}$ suggest that sensor i is sensitive to the damage introduced in stage z , hence contributing potentially to its localisation.

6.1. Accelerometers

Figure 15 presents heatmaps of FC values obtained from accelerometers placed along the blade for each damage transition, with the red line marking the location of newly introduced damage (i.e., cracks). Results from AR-40 and MAR-10 models are shown, alongside a quantitative summary in **Table 2**. For direct comparison with the following section, the heatmap colour scale is held invariable.

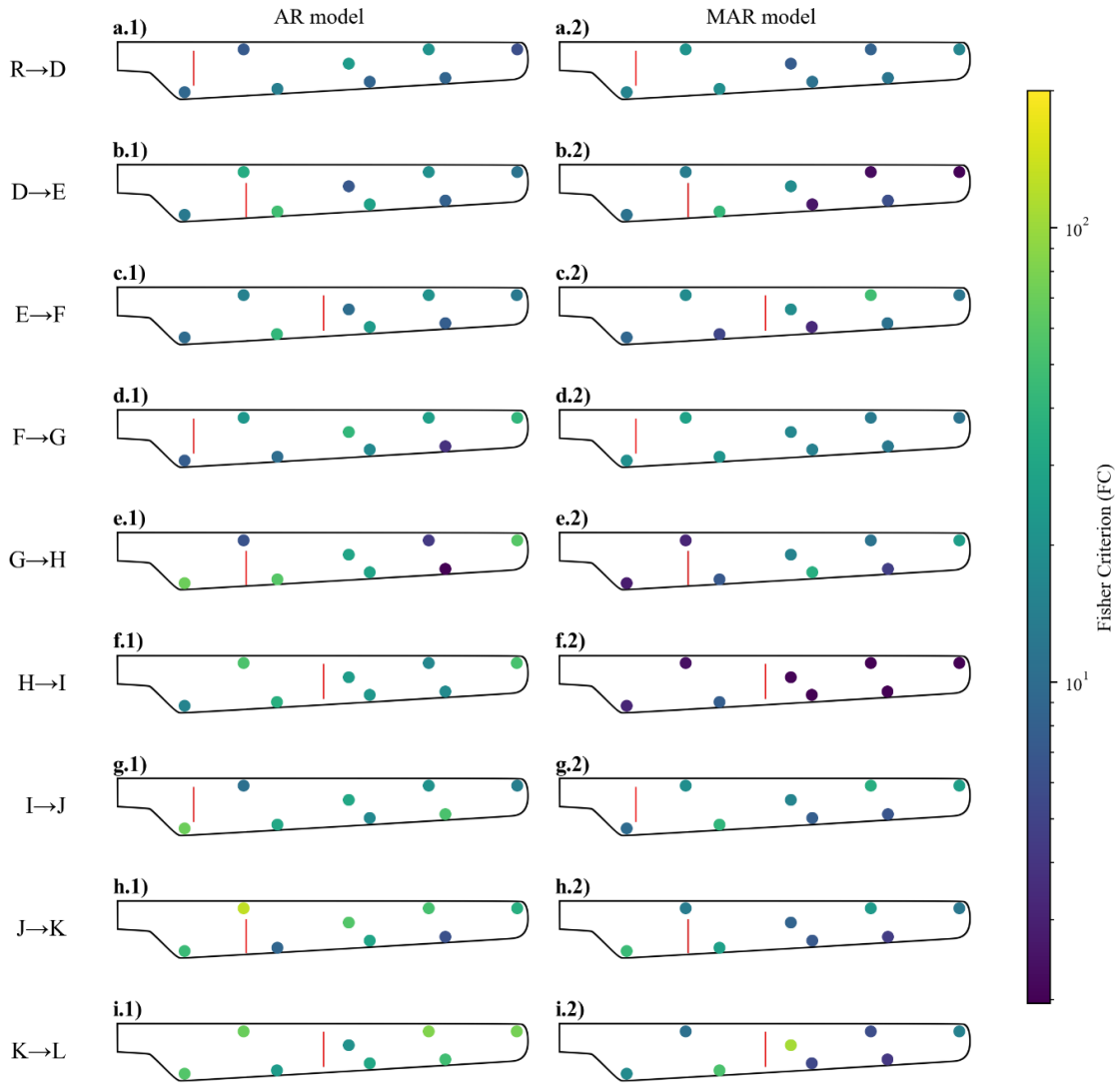


Figure 15: Heatmaps of FC values along the blade using accelerometer measurements, with (a.1)-(i.1) showing AR-40 model results in the left column and (a.2)-(i.2) MAR-10 models in the right column. The red line marks the new damage location for each damage transition.

Table 2: Sensor-wise FC per damage stage using accelerometers with AR-40 and MAR-10 models.

Transition	AR-40 models		MAR-10 models	
	Max FC	Sensor with Max FC	Max FC	Sensor with Max FC
R → D	24.5	a_4	21.0	a_2
D → E	48.7	a_3	42.0	a_3
E → F	41.4	a_3	48.2	a_6
F → G	40.8	a_4	27.4	a_2
G → H	70.3	a_1	35.7	a_5
H → I	54.1	a_2	7.7	a_3
I → J	70.6	a_1	40.6	a_3
J → K	127.3	a_2	45.3	a_1
K → L	83.3	a_6	107.8	a_4

From **Figure 15**, it can be observed that the highest FC values do not consistently correspond to sensors located near the actual damage, indicating limited localisation capability when relying solely on acceleration data. While certain damage transitions, such D → E and K → L, show higher FC values for sensors close to the damage, no clear pattern emerges. This limitation is likely due to the global response characteristics of acceleration signals.

These findings contrast with those of Mosavi et al. (2012), who reported higher FC values for sensors close to the damage when applying multivariate models to acceleration data from a laboratory-tested beam, also using the Fisher criterion. It should be noted, however, that the analysed systems differ substantially: the composite blade exhibits nonlinear and complex behaviour, whereas the beam used in Mosavi et al's study is made from a mostly linear steel material.

6.2. Strain gauges

To assess whether more localised measurements can improve damage localisation, the analysis was repeated using strain gauge data. **Figure 16** presents heatmaps of FC values obtained from AR-40 and MAR-2 models. The red line marks the new damage location for each damage transition. **Table 3** summarises the maximum FC values and corresponding sensors.

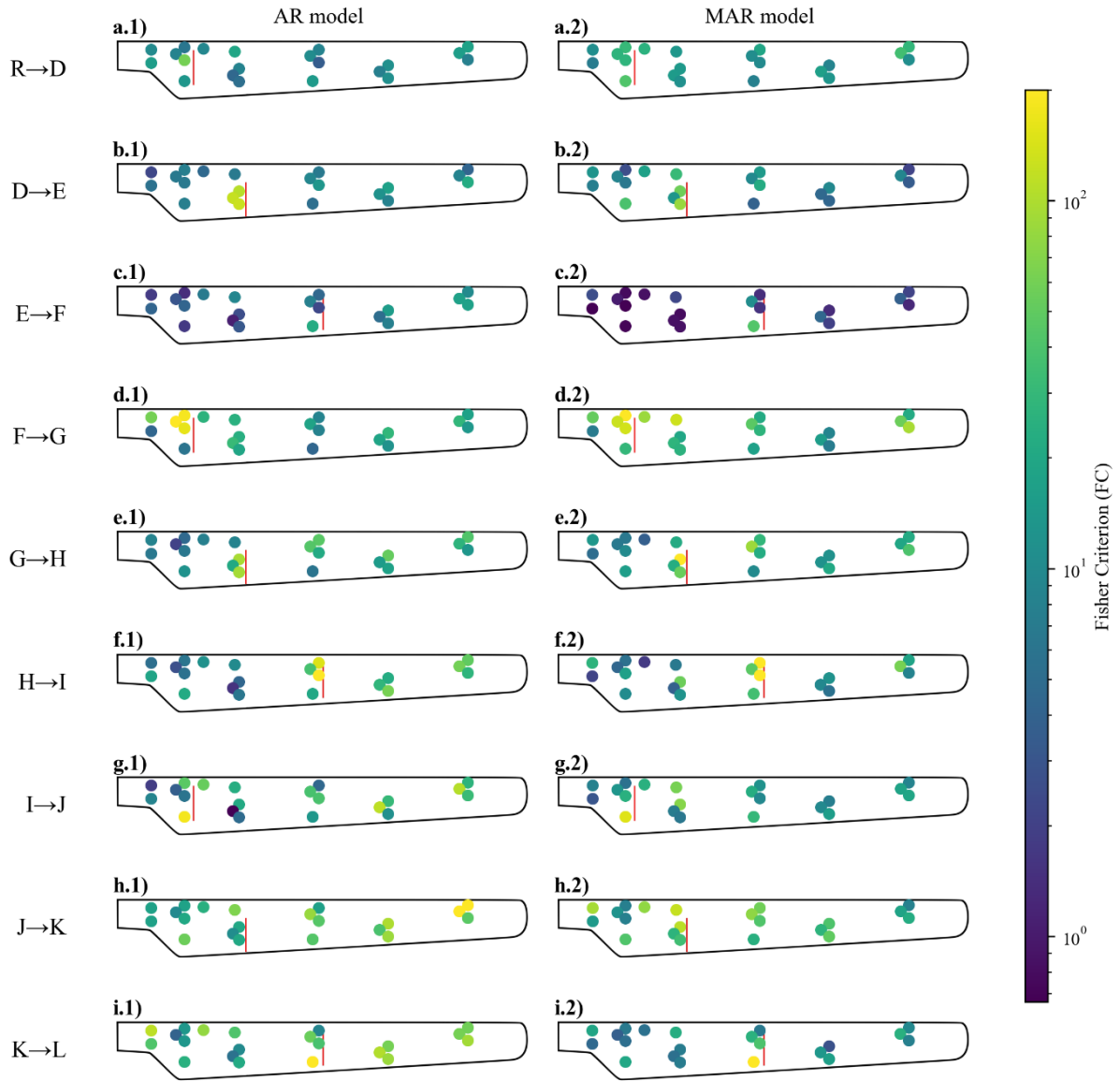


Figure 16: Heatmap of FC values along the blade using strain gauge measurements, with (a.1)-(i.1) showing AR-40 model results in the left column and (a.2)-(i.2) MAR-2 models in the right column. The red line marks the new damage location for each damage transition.

Table 3: Sensor-wise FC per damage stage using strain gauges with AR-40 and MAR-2 models.

Transition	AR-40 models		MAR models	
	Max FC	Sensor with Max FC	Max FC	Sensor with Max FC
$R \rightarrow D$	60.5	s_5	41.4	s_6
$D \rightarrow E$	127.1	s_{11}	83.1	s_{11}
$E \rightarrow F$	21.5	s_{15}	42.1	s_{15}
$F \rightarrow G$	367.7	s_4	389.3	s_3
$G \rightarrow H$	93.9	s_9	203.3	s_9
$H \rightarrow I$	321.6	s_{14}	373.4	s_{14}
$I \rightarrow J$	179.2	s_6	152.5	s_6
$J \rightarrow K$	603.1	s_{21}	189.2	s_8
$K \rightarrow L$	432.9	s_{15}	246.6	s_{15}

As observed, strain-gauge data yield superior localisation performance, with lighter colours appearing closest to the damage site on the heatmaps, indicating higher FC values. This confirms that strain gauges provide a more direct mapping of damage locations than accelerometers, as sensors positioned near the damage consistently register the highest FC values. The results also highlight that the effectiveness of strain gauges depends on their placement relative to the damage site, offering valuable insight for future sensor network design.

While damage transitions with lower detectability —such as $R \rightarrow D$ and $E \rightarrow F$ — produce lower FC values, localisation accuracy is not compromised. In terms of model comparison, MAR models do not substantially improve localisation in this scenario, likely because the high spatial resolution of strain-gauge measurements already provides sufficient information, leaving little added value from multivariate correlations. Nevertheless, MAR models demonstrate greater robustness, as seen in the $J \rightarrow K$ transition, where AR models fail to identify the damage location.

7. Discussion

Based on the results of this study, several key findings and implications for future research can be identified. Regarding the detectability of damage using both MAR and AR models, and with both accelerometers and strain gauges, the main findings are as follows:

1. The proposed algorithm detects damage using both acceleration and strain gauge data from the studied dataset, and a single sensor is sufficient to capture all investigated damage transitions, demonstrating efficiency with minimal instrumentation.
2. Across all examined scenarios, accelerometers show greater robustness for damage detection than strain gauges, achieving the lowest false positive and false negative rates when using a single sensor, and the highest Mahalanobis Distance values when multiple sensors are deployed in the blade's detection network.
3. Increasing the number of accelerometers enhances the accuracy and robustness of detection by providing richer dynamic information from across the structure, although beyond six sensors the benefits diminish due to redundancy.
4. MAR models outperform univariate AR models when two or more accelerometers are used, indicating that incorporating spatial correlations between sensors improves feature quality and sensitivity to damage, whereas for strain gauges MAR models provide no significant improvement over AR.

In contrast, for the locatability of damage, the following conclusions can be drawn:

1. Strain-based Fisher Criterion values exhibit a strong correlation with damage location, highlighting their potential for accurate spatial localisation of anomalies. Acceleration data, by contrast, show weaker direct correlation with location, likely because accelerations capture the global dynamic response, whereas strains are more sensitive to localised deformations in composite blades.
2. MAR models do not substantially enhance localisation accuracy for either accelerometers or strain gauges, although they may improve robustness in certain scenarios.

8. Conclusions

This study presented a semi-supervised, data-driven framework for structural health monitoring (SHM) of a small-scale composite wind turbine blade subjected to progressive damage and varying

temperatures. The methodology combined Multivariate Autoregressive (MAR) modelling, Principal Component Analysis (PCA), Mahalanobis distance metrics, and K-means clustering, and was tested using two sensor types: accelerometers and strain gauges. Experimental data from a blade with sequentially introduced cracks at multiple locations served as the basis for validating the proposed damage detection and localisation approach.

The detectability analysis confirms that both AR and MAR models, particularly when applied to accelerometer data, provide robust indicators of structural degradation. Single-sensor AR models already capture early damage transitions, although with limited sensitivity at initial stages. In contrast, two-sensor MAR models significantly enhance detectability by leveraging inter-sensor correlations, resulting in higher Mahalanobis Distances and clearer dissimilarity changes across damage transitions. These findings suggest that a minimal yet strategically placed sensor array can achieve high sensitivity to damage progression, offering a balanced trade-off between performance and instrumentation cost—supporting the feasibility of lightweight, output-only SHM solutions for in-service monitoring.

The results also indicate that while MAR models improve damage detectability, the capacity for damage localisation is more dependent on the type of measurement. Accelerometers, which reflect global structural dynamics, offer limited spatial resolution, making it difficult to pinpoint damage locations. In contrast, strain gauges, more sensitive to local deformation, enable clearer spatial correlation between damage and sensor response. Notably, the use of MAR models does not significantly improve localisation when applied to strain data, suggesting that strain-based AR models already provide sufficient fidelity for this purpose. These insights underscore the critical role of sensor modality in SHM design, especially when localisation is a key monitoring objective.

Therefore, for detectability aims the selection of accelerometers using MAR models should be the best option, while for locatability, using strain gauges with AR models provide acceptable results. Thus, a combination of those sensor types provides relevant information of the system's health. For detectability specifically, if strain gauges are preferred, farther positions from the blade root might be more suitable, while similar performance is expected using accelerometers over the blade.

Building on the insights from this study, future research should address several remaining challenges and opportunities. First, the maximum effective distance between a sensor and the damage site for reliable localisation remains uncertain, particularly for strain gauges, whose performance is highly dependent on placement — knowledge that could directly guide the optimisation of future sensor network designs. Second, although the optimal selection of AR coefficients has been investigated in the literature (e.g., Hoell & Omenzetter (2016)), its targeted evaluation for damage-detection applications warrants further exploration. Finally, extending the proposed methodology to full-scale wind-turbine blades operating under realistic and variable environmental and operational conditions would provide valuable validation. Such work could also integrate additional environmental variables — including humidity, loading history, and rotational speed — and assess a broader spectrum of damage types, thereby strengthening the robustness, adaptability, and real-world applicability of the approach. Advancing these research directions would not only refine the methodology itself but also contribute to the development of more resilient, efficient, and scalable structural health monitoring systems for the next-generation of offshore wind energy infrastructure.

CRediT authorship contribution statement

MQ: writing–review and editing, writing–original draft, visualisation, validation, software, methodology, investigation, formal analysis, data curation, and conceptualisation. EK: writing–review and editing. SP: writing–review and editing. MB: writing–review and editing, validation, supervision, resources, methodology and conceptualisation.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

The authors gratefully acknowledge the financial support from the City St George’s, University of London through the Research at the Frontiers of Practice PhD Scholarship. Authors would also like to express their sincere gratitude to The Alan Turing Institute and the Lloyd’s Register Foundation for their funding support via the Data-Centric Engineering group. Any opinions, findings, conclusions, or recommendations expressed in this paper are those of the authors and do not necessarily reflect those of the sponsors.

References

- Achilli, A., Bernagozzi, G., Betti, R., Diotallevi, P. P., Landi, L., Quqa, S., & Tronci, E. M. (2021). On the use of multivariate autoregressive models for vibration-based damage detection and localization. *Smart Structures and Systems*, 27(2), 335–350. <https://doi.org/10.12989/sss.2021.27.2.335>
- Bishop, C. M., & Nasrabadi, N. M. (2006). *Pattern recognition and machine learning* (Vol. 4, Issue 4). Springer.
- Calderano, P., De Marins, D. B., & Ayala, H. (2022). A comparison of Feature Extraction Methods for Crack and Ice Monitoring in Wind Turbine Blades: System Identification and Matrix Decomposition. *2022 30th Mediterranean Conference on Control and Automation, MED 2022*, 779–784. <https://doi.org/10.1109/MED54222.2022.9837260>
- Carroll, J., McDonald, A., & Mcmillan, D. (2016). Failure rate, repair time and unscheduled O&M cost analysis of offshore wind turbines. *Wind Energy*, 19(6), 1107–1119.
- Dao, C., Kazemtabrizi, B., & Crabtree, C. (2019). Wind turbine reliability data review and impacts on levelised cost of energy. In *Wind Energy* (Vol. 22, Issue 12, pp. 1848–1871). John Wiley and Sons Ltd. <https://doi.org/10.1002/we.2404>
- Du, Y., Zhou, S., Jing, X., Peng, Y., Wu, H., & Kwok, N. (2020). Damage detection techniques for wind turbine blades: A review. In *Mechanical Systems and Signal Processing* (Vol. 141). Academic Press. <https://doi.org/10.1016/j.ymssp.2019.106445>

555 Eltoumy, K., Gomaa, M., & Liang, X. (2023). Unsupervised Learning Methods for Data-Driven
556 Vibration-Based Structural Health Monitoring: A Review. *Sensors*, 23(6).
557 <https://doi.org/10.3390/s23063290>

558 Entezami, A., Sarmadi, H., Behkamal, B., & Mariani, S. (2020). Big data analytics and structural health
559 monitoring: A statistical pattern recognition-based approach. *Sensors (Switzerland)*, 20(8).
560 <https://doi.org/10.3390/s20082328>

561 Entezami, A., & Shariatmadar, H. (2018). An unsupervised learning approach by novel damage indices
562 in structural health monitoring for damage localization and quantification. *Structural Health*
563 *Monitoring*, 17(2), 325–345. <https://doi.org/10.1177/1475921717693572>

564 Goi, Y., & Kim, C. W. (2017). Damage detection of a truss bridge utilizing a damage indicator from
565 multivariate autoregressive model. *Journal of Civil Structural Health Monitoring*, 7(2), 153–162.
566 <https://doi.org/10.1007/s13349-017-0222-y>

567 Gul, M., & Necati Catbas, F. (2009). Statistical pattern recognition for Structural Health Monitoring
568 using time series modeling: Theory and experimental verifications. *Mechanical Systems and*
569 *Signal Processing*, 23(7), 2192–2204. <https://doi.org/10.1016/j.ymssp.2009.02.013>

570 Han, Q., Ma, Q., Xu, J., & Liu, M. (2021). Structural health monitoring research under varying
571 temperature condition: a review. *Journal of Civil Structural Health Monitoring*, 11(1), 149–173.
572 <https://doi.org/10.1007/s13349-020-00444-x>

573 Hoell, S., & Omenzetter, P. (2016). Optimal selection of autoregressive model coefficients for early
574 damage detectability with an application to wind turbine blades. *Mechanical Systems and Signal*
575 *Processing*, 70–71, 557–577. <https://doi.org/10.1016/j.ymssp.2015.09.007>

576 Li, D., Ho, S. C. M., Song, G., Ren, L., & Li, H. (2015). A review of damage detection methods for
577 wind turbine blades. In *Smart Materials and Structures* (Vol. 24, Issue 3). Institute of Physics
578 Publishing. <https://doi.org/10.1088/0964-1726/24/3/033001>

579 Liu, W., Lai, Z., Bacsá, K., & Chatzi, E. (2024). Neural extended Kalman filters for learning and
580 predicting dynamics of structural systems. *Structural Health Monitoring*, 23(2), 1037–1052.
581 <https://doi.org/10.1177/14759217231179912>

582 Long, B., Xian, W., Jiang, L., & Liu, Z. (2013). An improved autoregressive model by particle swarm
583 optimization for prognostics of lithium-ion batteries. *Microelectronics Reliability*, 53(6), 821–831.
584 <https://doi.org/10.1016/j.microrel.2013.01.006>

585 Meixedo, A., Santos, J., Ribeiro, D., Calçada, R., & Todd, M. (2021). Damage detection in railway
586 bridges using traffic-induced dynamic responses. *Engineering Structures*, 238.
587 <https://doi.org/10.1016/j.engstruct.2021.112189>

588 Meixedo, A., Santos, J., Ribeiro, D., Calçada, R., & Todd, M. D. (2022). Online unsupervised detection
589 of structural changes using train-induced dynamic responses. *Mechanical Systems and Signal*
590 *Processing*, 165. <https://doi.org/10.1016/j.ymssp.2021.108268>

591 Mosavi, A. A., Dickey, D., Seracino, R., & Rizkalla, S. (2012). Identifying damage locations under
592 ambient vibrations utilizing vector autoregressive models and Mahalanobis distances. *Mechanical*
593 *Systems and Signal Processing*, 26(1), 254–267. <https://doi.org/10.1016/j.ymssp.2011.06.009>

- Ou, Y., Tatsis, K. E., Dertimanis, V. K., Spiridonakos, M. D., & Chatzi, E. N. (2021). Vibration-based monitoring of a small-scale wind turbine blade under varying climate conditions. Part I: An experimental benchmark. *Structural Control and Health Monitoring*, 28(6). <https://doi.org/10.1002/stc.2660>
- Panagiotopoulos, A. I., Tcherniak, D., & Fassois, S. D. (2021). Damage Detection on an Operating Wind Turbine Blade via a Single Vibration Sensor: A Feasibility Study. *Lecture Notes in Civil Engineering*, 128, 405–414. https://doi.org/10.1007/978-3-030-64908-1_38
- Santos, J. P., Crémona, C., Orcesi, A. D., & Silveira, P. (2013). Multivariate statistical analysis for early damage detection. *Engineering Structures*, 56, 273–285. <https://doi.org/https://doi.org/10.1016/j.engstruct.2013.05.022>
- Shumway, R., & Stoffer, D. S. (2017). *Time Series Analysis and Its Applications: With R Examples* (Fourth Edition). Springer Texts in Statistics, Springer, Cham.
- Sierra-Pérez, J., Güemes, A., & Mujica, L. E. (2013). Damage detection by using FBGs and strain field pattern recognition techniques. *Smart Materials and Structures*, 22(2). <https://doi.org/10.1088/0964-1726/22/2/025011>
- Sierra-Pérez, J., Torres-Arredondo, M. A., & Güemes, A. (2016). Damage and nonlinearities detection in wind turbine blades based on strain field pattern recognition. FBGs, OBR and strain gauges comparison. *Composite Structures*, 135, 156–166. <https://doi.org/10.1016/j.compstruct.2015.08.137>
- Sohn, H. (2007). Effects of environmental and operational variability on structural health monitoring. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 365(1851), 539–560. <https://doi.org/10.1098/rsta.2006.1935>
- Tu, J. Q., Tang, Z. F., Yun, C. B., Wu, J. J., & Xu, X. (2021). Guided wave-based damage assessment on welded steel I-beam under ambient temperature variations. *Structural Control and Health Monitoring*, 28(4). <https://doi.org/10.1002/stc.2696>
- Yang, W., Lang, Z., & Tian, W. (2015). Condition Monitoring and Damage Location of Wind Turbine Blades by Frequency Response Transmissibility Analysis. *IEEE Transactions on Industrial Electronics*, 62(10), 6558–6564. <https://doi.org/10.1109/TIE.2015.2418738>
- Yang, Z., Xu, M., Wang, S., Li, J., Peng, Z., Jin, F., & Yang, Y. (2024). Detection of wind turbine blade abnormalities through a deep learning model integrating VAE and neural ODE. *Ocean Engineering*, 302. <https://doi.org/10.1016/j.oceaneng.2024.117689>