Extrapolation of AR models using cubic splines for damage progression evaluation in composite structures

Samuel da Silva[1], **Jessé Paixão**[1], **Marc Rébillat**[2] and **Nazih Mechbal**[2]

Abstract

This paper presents the potentiality of the use of extrapolation of a set of Auto-Regressive (AR) models to inspect a future damage sensitive indices based on changes in one-step-ahead prediction errors. The key idea is to use multiple AR models to assess a data-driven model to represent and predict the time-series outputs of the PZT sensors receiving Lamb waves in a composite coupon. Based on some simplified assumptions, after detecting initial damage using some previous classifier, its progression evaluation by interpolating the AR parameters is proposed and examined based on cubic spline functions. After, an extrapolated AR model using this information may verify the future state and to inspect how the damage could progress. An aeronautical composite panel with bonded piezoelectric elements that act both as sensors and actuators is utilized to examine the relationship between the variation of the identified model parameters with various levels of simulated damage. The results have shown a smooth and adequate correlation between the estimates obtained by the extrapolated model and the actual progress of the damage observed. The significant advantage of the proposed procedure is implementing this task without adopting a complicated and costly mathematical-physical model.

Keywords

Auto-regressive models, data-driven system identification, multiple models, composite structures, extrapolated model, damage progression

1. Introduction

Extensive use of composites structures occurs due to their adequate properties combining lightweight and high resistance. However, the existence of complex mechanisms of failures and progression of damage, such as debonding, matrix cracking, fiber failures, or delaminations that are not visible a priori requires full use of structural health monitoring (SHM) techniques for such structures (Mitra and Gopalakrishnan, 2016). These SHM approaches seem to be in a mature stage in the steps of detection and localization of possible structural changes in composite structures. Several robust methods have been proposed and validated in the last decades by extracting features from signals or model-based parameters in time and frequency domain (Fendzi et al., 2016). Among these approaches, one very effective way to address this issue is by using data-driven model identification based on guided Lamb wave propagation or random inputs provided by PZT active-sensing. In particular, Auto-Regressive (AR) models to describe Lamb waves are already abundantly used in the SHM of composite structures (da Silva, 2018; Figueiredo et al., 2012; Nardi et al., 2016; Paixão et al., 2020).

Nardi et al. (2016) using an auto-regressive (AR) model were able to detect delamination in a carbon-fiber-reinforced-plastic laminate plate excited by random input using a couple of piezoelectric patches as actuator and sensors. Kim et al. (2014) also showed a possible data-driven system identification through a state-space model to capture the wave motion in metallic structures. Figueiredo et al. (2012) showed the use of time-series predictive models for piezoelectric

[1]Departamento de Engenharia Mecânica, UNESP - Universidade Estadual Paulista, Ilha Solteira, SP, Brazil
[2]PIMM Laboratory, Arts et Métiers - ENSAM/CNRS/CNAM, Paris, France

Corresponding author:
Samuel da Silva, Departamento de Engenharia Mecânica, UNESP - Universidade Estadual Paulista, Av. Brasil 56, Ilha Solteira, SP 15385-000, Brazil.
Email: samuel.silva13@unesp.br

active-sensing using autoregressive with exogenous input (ARX) models and machine learning to detect damages. da Silva (2018) also applied the ARX model to perform predictions, and a waveform generator in a 10 layers carbon-epoxy plate excited by guided Lamb waves assuming different central frequencies and environmental conditions, such as temperatures changes, and noted some benefits and disadvantages of the possible performance of this strategy for SHM.

All these papers addressed the lower level of the SHM's hierarchy, that is, to detect and, in some cases, to locate the position of the structural change associated with damage. However, to reach higher levels, such as quantification and prognosis, numerous impediments are required to be overcome yet. One of them is that it is vital to have an adequate mathematical model if the user needs to predict a future state based on previous past data to investigate the existence and evolution of damage propagation. These SHM models for damage progression evaluation demand to incorporate information about the damage behavior in its dynamics to gain a comprehensive physical insight into the monitored structure; consequently, this model should be most physically possible. However, to construct numerical models, for example, using finite element methods with damages requires ensuring its local validity and how to model a mechanism of damage adequately. This step is challenging for a real-time monitoring system in the industrial field, even when modeling from its healthy state (Balmes et al., 2016). Another limitation is that the behavior of damage evolution is usually very complicated to model in a real-world application because complex types of damage can appear coincidentally with several confounding effects, like noise, uncertainties, temperature changes, and operational variability etc.

Despite these limitations, various authors proposed methods using advanced signal processing and machine learning algorithms. Larrosa et al. (2014) examined a set of carbon fiber-reinforced polymer (CFRP) laminates with various levels of delamination and other simultaneous structural changes. Fatigue tension-tension tests simulated all these conditions. Several features are extracted from time and frequency signals measured by the piezoelectric sensors receiving guided waves. A machine learning framework using a Gaussian discriminant was trained to classify the structural state and obtain a learning model. A successful test was reached. Corbetta et al. (2018) demonstrated Bayesian processing using a particle filtering algorithm for estimating fatigue life prediction in the same CFRP laminate used by Larrosa et al. (2014) with the occurrence of matrix cracks and delamination. Paixão et al. (2020) utilized a Mahalanobis squared distance calculated using data-driven models to interpolate the features seeking to correlate with the delamination area in composite structures. Amer and Kopsaftopoulos (2019) proposed a similar strategy by interpolating damage indices from non-parametric statistical times series representation using Gaussian Process Regression to identify damage size. The experimental setup was performed in a notched aluminum plate with beneficial results. Notwithstanding these results, any overall strategy must be consolidated to carry out these quantifications and prognosis. In particular, estimating a future model to observe the damage propagation in a more complex material, as composite structures, should be helpful. Fortunately, simple algorithms of system identification can provide this information about damage progression and to have access to a mathematical-physical model to simulate the future state of damage from a test bench.

Identified data-driven model, as suggested by Nardi et al. (2016) or da Silva (2018) using AR family models, could be attractive to be adopted to interpolate the states between the healthy and damaged condition and to extrapolate or to quantify a damage progression. These models can be used as a surrogate model to reach a subsequent application of higher forms of SHM's hierarchy. Thus, the present paper demonstrates that the coefficients of AR models in a healthy state and initial damage conditions are well related to the severity of the structural changes that cause the damage. If an adequate interval between the healthy (reference) and the damaged case is measured, warranting a smooth behavior with the same regressive order, a trend curve of the damage-sensitive feature index can be updated and easily extrapolated show future conditions of the outputs. Goidescu et al. (2013) demonstrated a full damage investigation in composites using different optical techniques and imaging process that allows us to ensure qualitatively this hypothesis involving a local damage phenomenon. Thus, this manuscript is a first effort of the authors in this direction seeking to interpolate and extrapolate AR polynomials through spline functions to extend how damage-sensitive indices could evolute, but quantitatively, based on simplified assumptions.

This paper is organized as follows. First, the problem statement and hypothesis assumed in the implementation of the proposal are presented. Next, the feature extraction and damage detection procedure are reviewed. To classify the structural states, an analysis of variance (ANOVA) with Tukey's multiple comparison test is utilized based on Gonsalez et al. (2015). In the next section, when damage is previously detected in an early stage, a new set of models, named by *initial damaged models*, is captured to extrapolate a projected state using cubic spline polynomials with a trained data. A carbon-epoxy laminated plate with controlled progressive structural change similar to actual damage, as delamination, is described in the results to illustrate its effectiveness. Finally, the concluding remarks are provided in conclusion with recommendations for the use of such a method.

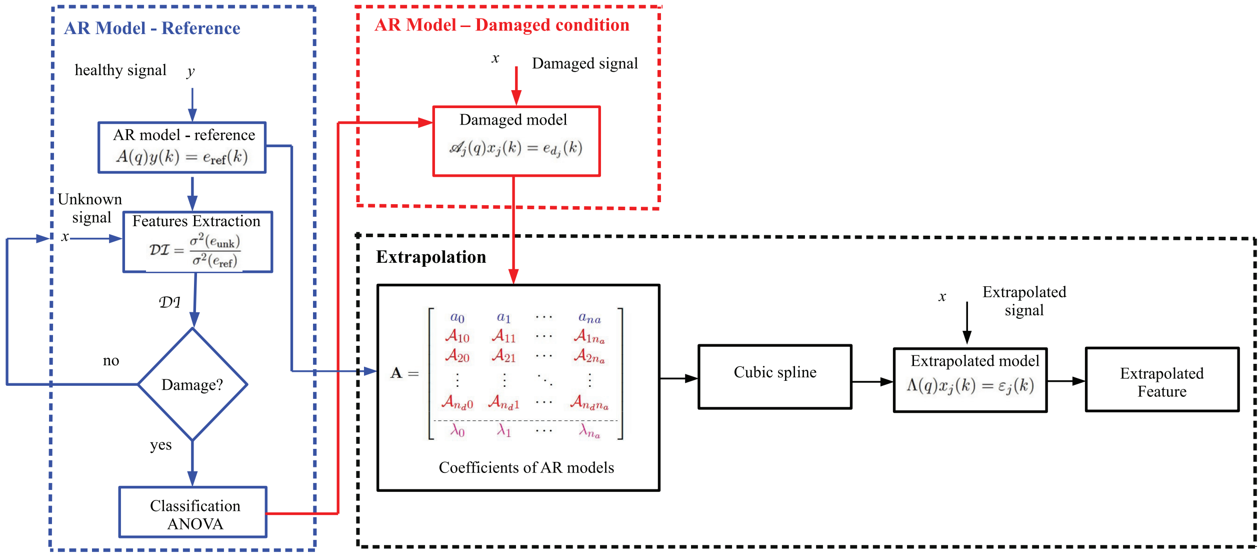


Figure 1. Proposed algorithm.
Please refer to the online version of the article to view the figure(s) in color.

2. Problem statement and proposed algorithm

The proposed procedure has three steps: identification of the reference model, classification, and obtention of the damaged models with the extrapolation, to be implemented, illustrated in Figure 1 for each path of guided wave propagation. First, damage detection and localization using an index extracted by predictions errors filtered with a reference AR model are performed handling conventional tools. Any classifier could be used in this step of the procedure, for example, support vector machine or other probabilistic machine learning algorithms (Mechbal et al., 2015). However, the goal of the present paper is not to study the classifier in itself. For this reason, a previously proposed procedure using an ANOVA scheme is utilized to classify the structural states (Gonsalez et al., 2015). The objective is to find a way to identify a set of initial models adequately and to have a process to understand as these sets of initial damaged AR models appropriately identified can evolute when damage size is increased. Primarily, we intend to show that the coefficients of these AR models may be extrapolated in a controlled form utilizing some splines functions to show a correlation capturing the damage evolution in a hotspot. This work's main contribution is to extrapolate the models to take the progress of located damage in a composite structure when it is initially detected.

Some assumptions are required to apply the strategy proposed in the present paper. These points are described below:

- A network of PZTs active-sensing optimally placed is previously bounded on the composite coupon's surface to be monitored. In our case, the direction of fibers in the composite plate helps us define the sensor's position without using an optimization method to select it or to define the number of required transducers. So, the methodology produced adequate results due to the simplicity of the geometry of the tested structure. Of course, when assuming a complex structure, the definition of the number and placement of the sensors should be addressed carefully to overcome issues related to the effect of reflected waves on the damage index's sensitivity.
- The guided wave signals used to excite the structure with specific central frequencies are also earlier chosen.
- A dataset of time-series is measured and saved in different structural scenarios and compared with a healthy (reference) AR model. The features sensitive to the damage are extracted from one-step-ahead prediction errors. Other features can also be used depending on the application.
- The classification among the structural state is triggered through hypothesis tests, and ANOVA is applied to cluster the initial damage states.
- The set of classified states in different structural conditions are utilized to suggest a family of AR models with the same order.
- A cubic spline polynomial is updated using the set of AR coefficients to observe the changes and to capture the effect caused by the damage. It is challenging to determine an interpolation polynomial because we do not know how the damage can propagate and modify the coefficients of an AR model. The choice is by using a cubic spline because it is smoother comparing with the quadratic spline. This happens because the quadratic spline has only continuous first

derivatives, while the cubic spline also has a second derivative. Thus, a cubic spline is satisfactory to approximate more complex functions, and it is our preference here.
- If the adjusted trend curve is well done, the family of AR model in the reference and initial damage conditions could be interpolated to observe how the occurrence of damage changes the coefficients of the AR models in the initial stages. This evolution of damage is unknown a priori, but with the set of models correctly interpolated could be used to make a prevision of a future model before the occurrence of a propagated damage. Hence, a posteriori damage index can be calculated using the extrapolated model to observe the future state.

The following sections explain the damage feature extraction, damage detection, and, mainly, the essential contribution of this paper using the cubic splines to estimate the effect of the simulated damage in the AR models to validate the extrapolation step. Once the extrapolated AR model is created, it is utilized to evaluate a future state and confronting the effects with the actual progression of the damage.

3. Feature extraction and damage detection

Assuming a discrete time-series $y(k)$ measured by a PZT sensor in a healthy state, a normalization is conducted to remove offset and have mean 0 and scaled to have standard deviation 1:

$$\hat{y}(k) = \frac{y - \bar{y}}{\sigma(y)} \tag{1}$$

where $\hat{y}(k)$ is the normalized signal, $\bar{y}$ is the mean and $\sigma(y)$ is the standard deviation. For simplicity, hereafter, y is used to denote $\hat{y}$. Now, a simple AR model can be described by a compact difference equation (Ljung, 1998):

$$A(q)y(k) = e_{ref}(k) \tag{2}$$

where $e_{ref}(k)$ is the one-step-ahead error prediction in a healthy condition assumed to be a white noise and $A(q) = \sum_{i=0}^{n_a} a_i q^{-i}$ is the healthy AR polynomial[1] with the coefficients a_i with a lag-order n_a, for example, $y(k)a_i q^{-i} = a_i y(k-i)$, where q^{-i} is a lag operator and k is the time sample. The order can be estimated using the Akaike information criterion (AIC), and the polynomial $A(q)$ may be identified through a least-squares or Yule-Walker approach, with codes fully available in Matlab or Octave software. When a new normalized data, $x(k)$, in an unknown state is measured, one can predict using the corresponding reference model:

$$A(q)x(k) = e_{unk}(k) \tag{3}$$

where $e_{unk}(k)$ is the unknown error prediction to be classified in a healthy or damaged state. Various papers have been using a simple damage-sensitive index $\mathcal{DI}$ based on a comparison of the variance $\sigma^2(\cdot)$ of prediction errors (Shiki et al., 2017):

$$\mathcal{DI} = \frac{\sigma^2(e_{unk})}{\sigma^2(e_{ref})} \tag{4}$$

If $\mathcal{DI}$ belongs a $\mathcal{F}-$distribution[2] there is no damage and the unknown condition is associated to healthy state (null hypothesis $\mathcal{H}_0$ is true). On the other hand, if the structure presents a damaged state, the probability distribution of the unknown error changes, and the alternative hypothesis $\mathcal{H}_1$ is true (Kopsaftopoulos and Fassois, 2010; Shiki et al., 2017).

A one-way ANOVA can also be used to test the hypothesis that the samples in the running tests $\mathcal{DI}$ belong to a population with the same means (null hypothesis $\mathcal{H}_0$), that is, the systems are classified as healthy state, against the alternative hypothesis $\mathcal{H}_1$ that the population means are not all the same, that is, damaged state (Gonsalez et al., 2015; Hogg, 1987). Tukey's multiple comparison test to decide whether the ANOVA results are in one particular group with a mean significantly different than another group is also implemented. This enables us visible to distinguish the clusters correlated with different damage sizes to be possible to observe the number of clusters presented in the structural states examined in a simple plot. For more details about this procedure of classification, the reader is encouraged to consult Gonsalez et al. (2015).

4. Extrapolation of AR coefficients using cubic splines

After clustering using ANOVA,[3] a new set of AR models for each initial damage recognized is estimated since the reference model given by equation (2) is not any more accurate. This initial damaged model is described to predict the current state output $x_j(k)$ by:

$$\mathscr{A}_j(q)x_j(k) = e_{d_j}(k) \tag{5}$$

where $\mathscr{A}_j(q) = \sum_{i=0}^{n_a} \mathcal{A}_{ji} q^{-i}$ is the AR polynomial with coefficients $\mathcal{A}_{ji}$ and $e_{d_j}(k)$ is the prediction error (white noise) in the damaged state $j = 1, \cdots, n_d$ classified by ANOVA, where n_d is the number of initial damaged states. An important simplifying assumption is considered here: *this classified initial damage is an early stage and does not change abruptly comparing with reference (healthy state)*. Consequently, the same regressive order n_a and framework (AR model) may be employed and the change is smooth between the coefficients, that is, $A(q) \approx \mathscr{A}_j(q)$, once $|a_i - \mathcal{A}_{ji}| < \delta$ for all i and δ is a limit value to warrant this hypothesis. It is essential to observe that the damage index given by equation (4) is sensitive to the changes to detect damage, but the specific changes in the AR coefficients are not significant when this structural variation is in the initial stages. Paixão et al. (2020) demonstrated this question in two features using prediction errors, and the variations of the AR coefficients were to observe a significant change in the coefficients, the sum of changes in all n_a parameters was used in the damage is in the initial stages.

Rearranging the coefficients of reference model $A(q)$ and damaged $\mathscr{A}_j(q)$ as:

$$\mathbf{A} = \begin{bmatrix} a_0 & a_1 & \cdots & a_{na} \\ \mathcal{A}_{10} & \mathcal{A}_{11} & \cdots & \mathcal{A}_{1n_a} \\ \mathcal{A}_{20} & \mathcal{A}_{21} & \cdots & \mathcal{A}_{2n_a} \\ \vdots & \vdots & \ddots & \vdots \\ \mathcal{A}_{n_d0} & \mathcal{A}_{n_d1} & \cdots & \mathcal{A}_{n_dn_a} \\ \lambda_0 & \lambda_1 & \cdots & \lambda_{n_a} \end{bmatrix} \tag{6}$$

where the first line of the matrix $\mathbf{A} \in \mathbb{R}^{n_d+1 \times n_a}$ is formed by the reference coefficients and the next one by the initial damage states, the last line in magenta is the extrapolated coefficients. To enable performing adequate extrapolations of the AR coefficients in the reference and damaged conditions, the number of states needs to be $n_d \geqslant 3$; otherwise, only linear extrapolation is plausible. So, it was assumed $n_d = 3$ structural states in damage clusters to permit to use cubic splines.

Figure 2 shows the procedure to be used in each coefficient of the AR models. Two steps are required to perform the method. First, an interpolation is obtained using the AR models' coefficients in the healthy and damage conditions classified previously. The key idea is to capture by using $f_i(z)$ the influence of the damage size increasing contained in structural states in the health (H) and three damage conditions D1, D2, and D3. After defining the interpolation function, extrapolation can be proposed to extend the future coefficient λ_i associated with a future state, for example, in a damaged state D4. Here is essential to observe that this can be proposed to a short horizon of time respecting the interval between the damage states identified in the interpolation step. Hence, the effect of evolution should be visible through $\mathcal{DI}$ to show a monotone change correlated with the severity. If a long horizon of extrapolation is applied, for example, for a damaged state D5, a significant error may be manifested because the interpolation function lacks sensibility to the damage's progression. This occurs because the assumptions made are broken, that is, the order and framework of the AR model's structure can change abruptly.

Piecewise polynomials can be used to extend each polynomial coefficient of λ_i associated with a future state through if it is supposed monotone (Wolberg and Alfy, 2002):

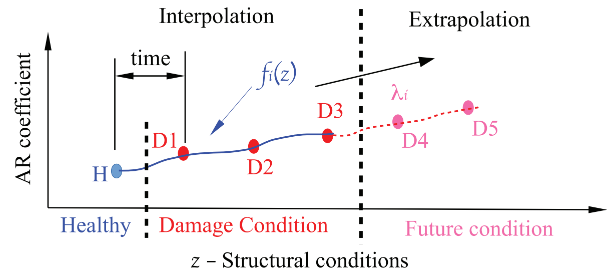


Figure 2. Extrapolation of AR coefficients.
Please refer to the online version of the article to view the figure(s) in color.

$$\lambda_i(z) = f_i(z) + \xi_i, \qquad i = 0, 1, \cdots, n_a \tag{7}$$

where z is the defined intervals associated with the number of structural conditions[4], $f_i(z)$ is a smoothing spline estimate by minimizing a penalized criterion for each coefficient in the AR model, and ξ_i is an independent random error. This function catches the variability of the coefficients in the AR models between the transition of healthy and damaged states in the interpolation steps to permit to apply an extrapolation. Several software packages implementing conventional interpolation and extrapolation methods using cubic splines can be utilized to determine this function using Python or Octave codes, and here is used the Matlab code *interp1*[5]. More details to find the spline interpolation f_i can be seen in Wang (2011).

The critical issue here is to recognize that it is now possible to have a future model to predict the data $x_j(k)$ when $j = n_d + 1$ represents a future state of data on the same scale time captured in the scenario of the previous interpolation. This AR model is given by:

$$\Lambda(q)x_j(k) = \varepsilon_j(k) \tag{8}$$

where $\Lambda(q) = \sum_{i=0}^{n_a} \lambda_i q^{-i}$ is the extrapolated AR model in a future damage condition. The basic premise is that the damage progression occurs as previously captured by evolution in the columns of matrix $\mathbf{A}$. Thus, a new index and hypothesis tests can also be estimated to evaluate the progression and the changes in the distribution of the extrapolated prediction error $\varepsilon_j(k)$ and in the damage index $\mathcal{DI}$. Alternatively, else, the initial damage state can be defined as the new reference for monitoring this damage index to observe evolution.

5. Experimental example

Figure 3 shows a carbon-epoxy laminated with layup containing 10 plies unidirectionally oriented along 0° with four PZTs SMART Layers from Accelent Technologies, with 6.35 mm in diameter and 0.25 mm in thickness with a free-free boundary condition. PZT 1 is used as an actuator with a five-cycle tone burst input signal applied with 35 V of amplitude and center

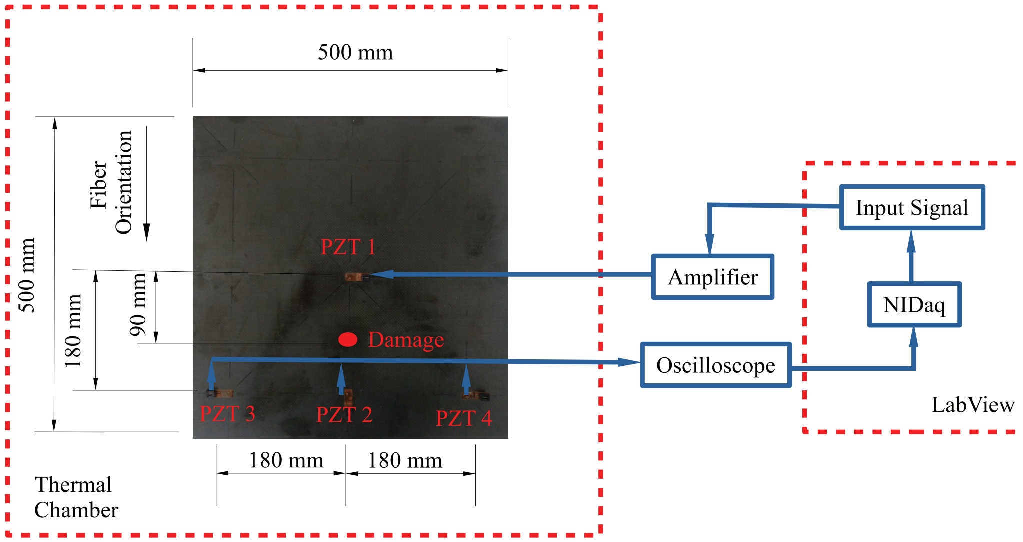


Figure 3. Illustration of the experimental setup with details about the geometry and instrumentation utilized.
Please refer to the online version of the article to view the figure(s) in color.

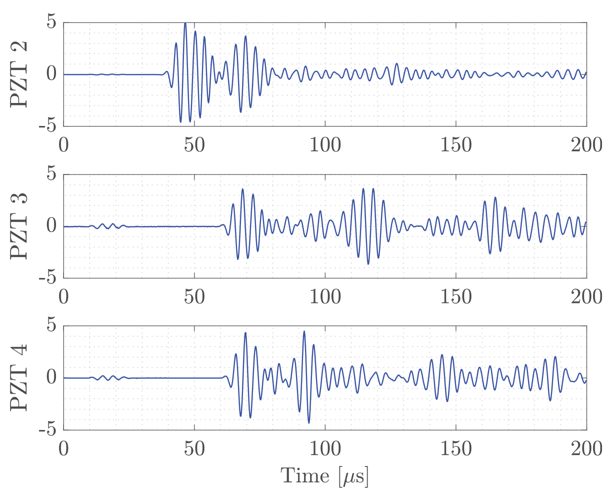


Figure 4. Healthy output time-series when PZT 1 acts as an actuator with a central frequency of 250 kHz.
Please refer to the online version of the article to view the figure(s) in color.

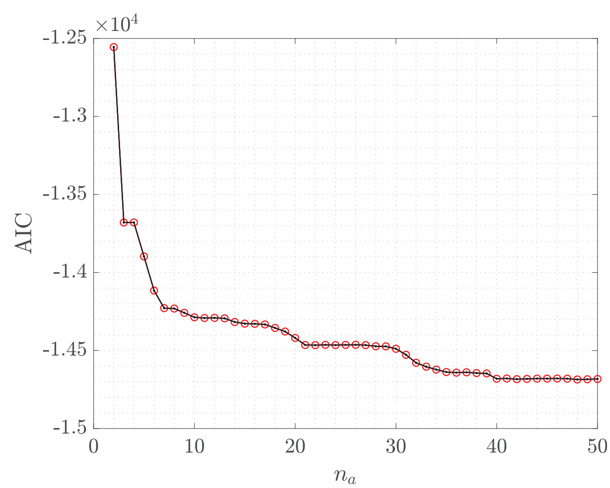


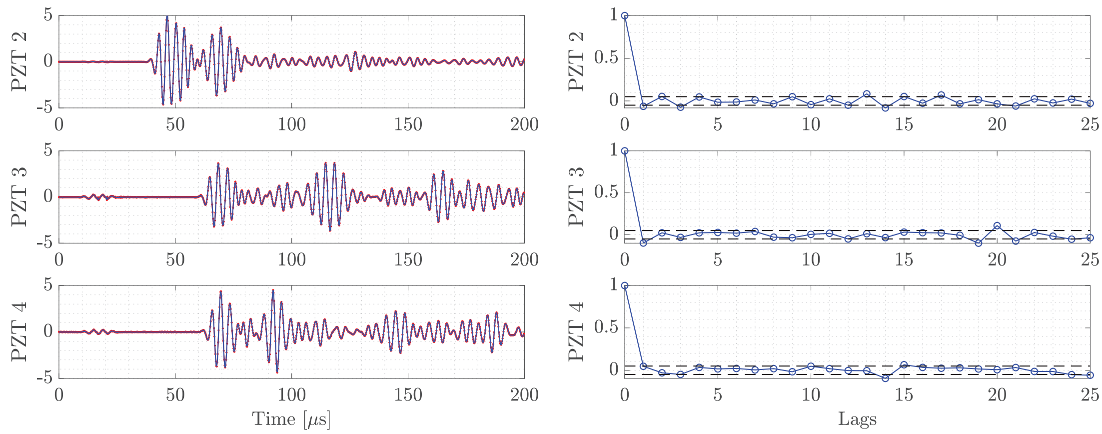
Figure 5. AIC order selection with focus in prediction - for the path PZT 1-2.
Please refer to the online version of the article to view the figure(s) in color.

frequency of 250 kHz. The outputs are collected in PZT 2, PZT 3, and PZT 4 with a sampling rate of 5 MHz and timespan of 200 μs with all channels synchronized measured. Data acquisition was controlled by Labview using a NI USB 6353 from National Instrument (NIDaq) and an oscilloscope DSO7034B Keysight assuming a controlled temperature of 30°C with all tests conducted inside a thermal chamber from Thermotron.

Figure 4 illustrates the output time-series measured by PZTs in a healthy state normalized by equation (1). An industrial adhesive putty was glued on the plate surface to simulate gradual damage by an additional mass, increasing the coverage area in the path between PZT 1 and PZT 2 progressively. This change modifies local material properties with a similar effect on the structural change associated with damage in composites structures, as exemplified by Lee et al. (2011). In each structural state, 100 tests were repeated for an adequate statistical characterization of the proposed damage detection.

5.1. AR model identification

The AIC order selection with a focus of prediction indicates that an order of $n_a = 40$ is sufficient to give an adequate validation for all paths of propagation, as seen in Figure 5 (Figueiredo et al., 2011). A raffle is performed to sort within 100 realizations randomly the signals in the PZT 2, PZT 3, and PZT 4 to be used as a reference and a specific healthy model $A(q)$ is identified in each path using the least square method. This is



(a) — Measured vs ··· Predicted.

(b) Whiteness test criteria within of 95% confidence bound patch.

Figure 6. Healthy output predicted by healthy model AR(40) (a) comparison of Measured versus Predicted, and (b) Whitenesse test criteria.

Please refer to the online version of the article to view the figure(s) in color.

executed to reduce computational processing time. Other realizations were filtered using this reference model by equation (2) to estimate the prediction error in the healthy state.

Figure 6(a) shows the comparison between the measured versus predicted assuming one validation data. The analysis of the autocorrelation function of residuals also indicates that the prediction errors are white noises, that is, the model has identified adequately, as observed in the Figure 6(b).

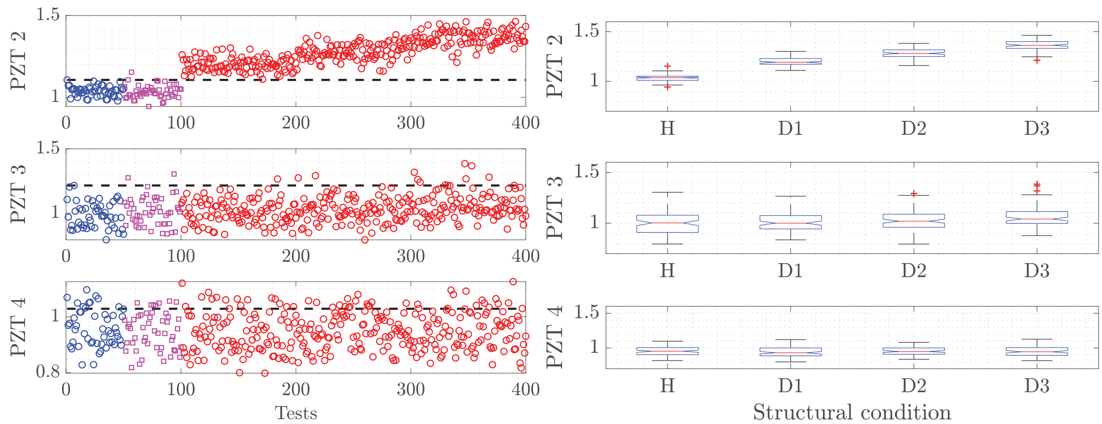
5.2. Damage detection and classification

Once a reference model, named by H, is correctly identified, it is used to detect some possible structural change. Only half of the measured data in the healthy state is utilized, and the next 50 is applied to evaluate the presence of false-positive using the hypothesis test. A set of blind tests using three different structural states associated with damage is performed, named by D1, D2, and D3 with increasing damage severity associated with the area covered by the industrial adhesive given by 490 mm^2 (0.19%), 707 mm^2 (0.28%) and, 962 mm^2 (0.38%), respectively, where the value between parenthesis is the percentage of the ratio of the covered area. Besides this knowledge about the covered area to be known, this is not utilized to implement the method. It is assumed that the damage increases between some states when the damage indices $\mathcal{DI}$ indicate a change.

Each structural condition was also measured with 100 realizations in each path. It is worth noting that the algorithm admits that all these data are assumed in unknown condition to be classified initially in two groups: healthy or damage. The prediction errors of these unknown conditions are computed using equation (3) and then the damage index $\mathcal{DI}$ is estimated by each test using the equation (4). The Lilliefors test is used to ensure that the variance of the prediction errors $\sigma^2(\cdot)$ in healthy states has normal distributions, that is, the $\mathcal{DI}$ trends to have a $\mathcal{F}-$ distribution in the reference condition (Shiki et al., 2017).

Figure 7(a) illustrates the index $\mathcal{DI}$, where it is possible to observe, as presumed, that the sensor PZT 2 is sensible to the presence of damage while other PZTs the distinction between the different situations is not possible. Figure 7(b) displays the evolution of the indices with the increase of the damage using a box-and-whisker plot, assuming known each damage state. A superposition in the upper quartile of the indices computed in the healthy state and the lowest quartile (even the median value) of damage states are also observed in PZT 3 and PZT 4. Figure 8 illustrates the receiver operating characteristics (ROC) curve to detect damage against reference condition considering all PZTs sensors. The line (0,1) designates the correct classification achieved by PZT 2, which is the path where the damage is located.

Additionally, the ANOVA procedure is computed to classify if the means are different or not, combined with Tukey multiple comparisons to see whereby statistically significant the clusters are in a straightforward plot to interpret. This step is performed to classify the number of damage states measured. Figure 9 presents this plot for PZT 2, where the circle is the mean value of the cluster with a 95% confidence interval. Three groups, $n_d = 3$, are distinctly separated and correlated with three possible structural states. After clustering, a new set of AR models in each damage group for PZT 2 is identified for monitoring the structural state and implementing the extrapolation explained in the next section.



(a) $\mathcal{DI}$. (b) Boxplot of $\mathcal{DI}$.

Figure 7. Damage index $\mathcal{DI}$: (a) damage index $\mathcal{DI}$: Healthy – training data (o), Healthy – test data(□) and Damaged (o). −− is the threshold line assuming a significance level of 5%, (b) box plots to observe the median, quartiles and outliers in the index distribution. (a) $\mathcal{DI}$. (b) Boxplot of $\mathcal{DI}$.
Please refer to the online version of the article to view the figure(s) in color.

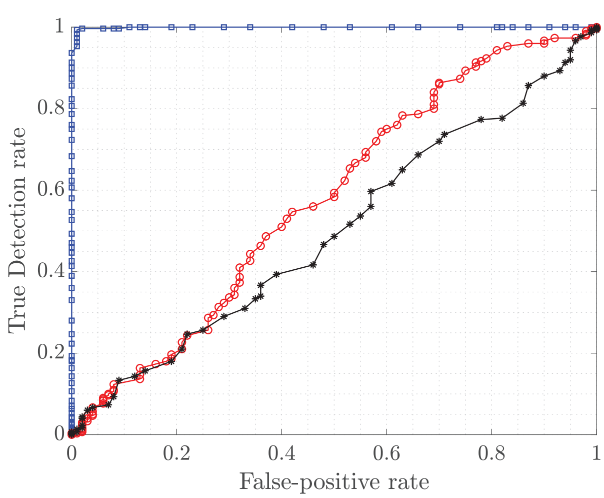


Figure 8. Receiver operating characteristics (ROC) curve: □ PZT 2, o PZT 3, and $*$ is PZT 4.
Please refer to the online version of the article to view the figure(s) in color.

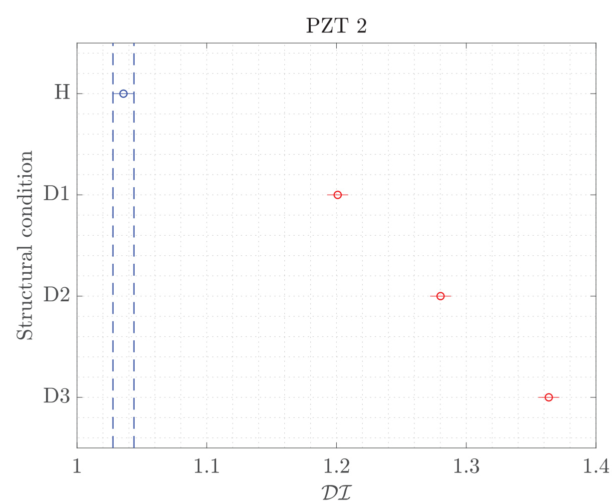


Figure 9. Classification of the clusters of different damage' states using ANOVA of $\mathcal{DI}$. 3 groups (in red) have means significantly different from healthy H.
Please refer to the online version of the article to view the figure(s) in color.

Other PZTs are not used here to identify new models because no variations are observed, permitting the identification of a new polynomial $\mathscr{A}_j(q)$ in equation (5), as confirmed in Figure 7.

5.3. Prevision of future AR damaged model

Three damaged models are estimated using one of the conditions presents in the clusters in Figure 9 after detecting and clustering. Combined with the reference model, a matrix given by equation (6) is formed with $n_d = 3$ and $n_a = 40$ to help the extrapolation to a further coefficient λ_i. A cubic spline polynomial created with the data in matrix $\mathbf{A}$ permits to estimate an extrapolated model $\Lambda(q)$ to use equation (8) to predict future states.

More eight future damage conditions are simulated by increasing the covered area to prove the benefits of the extrapolated model. Each one of these structural conditions is over measured 100 times. These damages are named by D4 to D11 with area covered by 1256 mm^2 (0.5 %), 1963 mm^2 (0.785%), 2375 mm^2 (0.950%), 2827 mm^2 (1.13%), 3848 mm^2 (1.53%), 5026 mm^2 (2.01%), 5674 mm^2 (2.27%), and 6361 mm^2 (2.54%), respectively. All these conditions are assumed not known previously in the procedure. First, the damage D4 with a surface area covered of 0.5% is used to predict the future state before the occurrence of the future

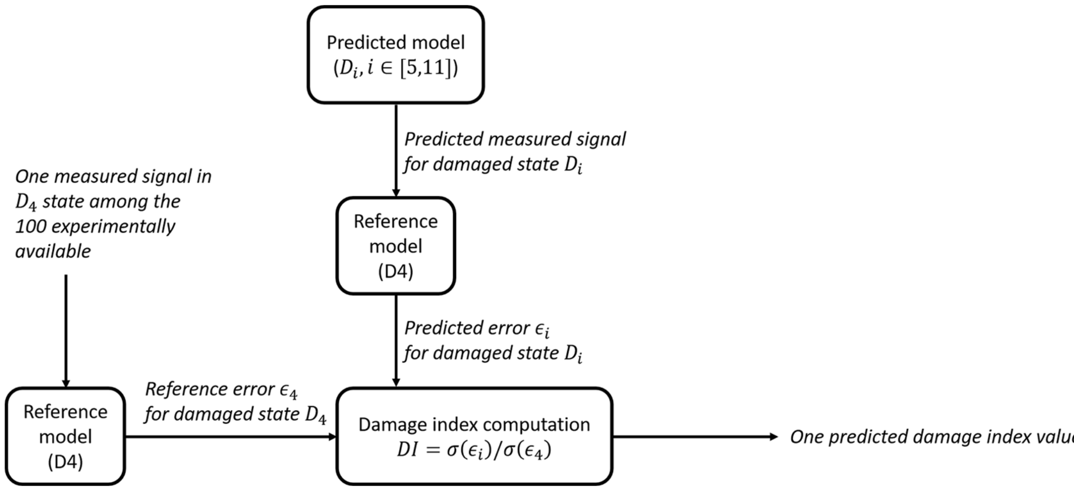


Figure 10. Prediction of damage index $\mathcal{DI}$ using the extrapolated model.
Please refer to the online version of the article to view the figure(s) in color.

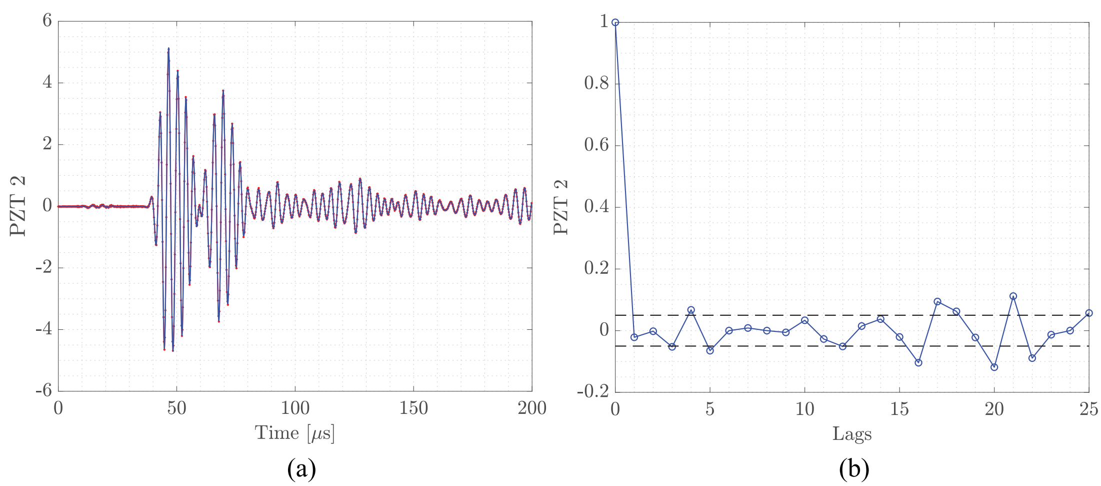
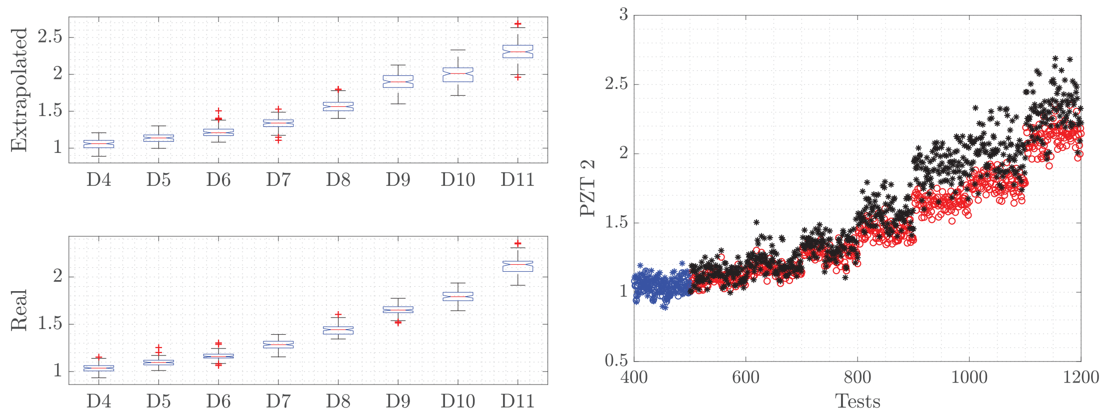


Figure 11. Output predicted by the extrapolated model AR(40) for damage condition D4: (a) measured vs $\cdots$ predicted, (b) whiteness test criteria.
Please refer to the online version of the article to view the figure(s) in color.

damage. Figure 10 demonstrates a schematic with the steps to propose the computation of the indices $\mathcal{DI}$ using the extrapolated model in the condition D_4.

The central idea is to use the coefficients in the matrix from equation (6) with the reference and D1 to D3 as initial damage condition to compute the coefficients for the equation (8) by extrapolating by a cubic spline to reach to condition D4. Figure 11(a) presents the comparison between the measured and predicted by the extrapolated model in damage condition D4. Figure 11(b) shows the residual analysis by the autocorrelation function to confirm an adequate prediction. Again it is essential to observe that these predict AR model is identified without the measured signal and obtained by the previous damaged models by extrapolation to an additional coefficient. The extrapolated model demonstrated adequate performance to predict the future state compared with the actual measured when this damage D4 happened.

When the damage condition is measured, a model can be identified, and one can compare the $\mathcal{DI}$ obtained by model identified by extrapolation and with the signals measured when the damage happens, assuming the new reference as D4. With the reference model and damaged models in conditions D1 to D3 is possible to predict the measures signals from condition D4, as shown in Figure 10. Here is assumed D4 as a new reference to compute the damage index $\mathcal{DI}$ because the model of the structure is constructed based on the previous state to observe the trend in the index when the damage progresses. Thus, the index is calculated based on the prediction error estimated by the extrapolated model comparing with the actual measurement when the damage has grown once the last states occurred, and a previous warning was already done.



(a) Boxplot of $\mathcal{DI}$. (b) $\mathcal{DI}$.

Figure 12. Damage index $\mathcal{DI}$ for PZT 2: (a) box plots of the damage index $\mathcal{DI}$ in extrapolated condition D4 to observe the median, quartiles and outliers in the index distribution assuming the extrapolated model and damaged model. (b) o actual damage D4 represents the new reference, $*$ extrapolated damaged D4 (extrapolated new reference), $*$ extrapolated damage, and actual Damaged (o) is the condition when the damage increases its size.
(a) Boxplot of $\mathcal{DI}$. (b) $\mathcal{DI}$.
Please refer to the online version of the article to view the figure(s) in color.

Thus, the current condition D4 is used as a reference. The procedure is performed for all the other damage situations based on the previous damaged models to observe the trend of the $\mathcal{DI}$ computed using the extrapolated information to get the structural condition D4.

The box-plot is also presented in Figure 12(a) for the other damage conditions to observe its relationship associating the actual data measured and extrapolated. It is possible to observe a similarity, where the $\mathcal{DI}$ using the prediction errors filtered by extrapolated model presents a more significant dispersion, but inside the limit assumed reliable for to observe the current state in an adequate classification. Figure 12(b) illustrates the predictions obtained by the filtered model using the actual data measured when the damage happens and the extrapolated model in each situation from damage D4 to D11, or else, from condition test 400 to 1200 (100 tests in each structural change). A more significant dispersion is found to a broad horizon when the damage (surface area covered by the adhesive) is supposed severe, and the model is not validated for this situation. This choice here of the validated horizon of prediction is correlated with the severity of the damage, once the extrapolation is done, considering it does not modify the AR model's framework and order. When the severity is higher, the order of the model and the regressors' structure is not extended certified.

Additionally, it is fundamental to observe that usually, when damage is alerted in a monitored system, a repair or visual inspection can be performed to evaluate the structural safety. Thus, assuming a short horizon with initial damage, this extrapolated model has a nice feature almost equal to an actual model identified when the damage increases and helps make decisions. Figure 13 shows the ROC curve comparing the similar performance between the extrapolated and the actual model identified when the damage occurred to detect all future damage conditions from D4 to D11. The extrapolated model in structural state D4 also allows dividing the clusters associated with the other damages using ANOVA, as seen in Figure 14. Thus, when new data are measured, the algorithm can correctly distinguish the states using the extrapolated model, even with dispersion in the damage index in the future condition as observed in Figure 12(b).

6. Discussion

A set of models is associated with parameters that predict future states or conditions before the occurrence to anticipate some decision or repair. Of course, the basic premise is that the evolution of the initial damage is not abrupt, and it has not happened at different points to assure the same regression order in the set of AR models.

Figure 12 demonstrates that the damage index $\mathcal{DI}$ computed is a monotone function to ensure adequate interpolation in the AR coefficients in the path where the damage is located. It is worth observing that the $\mathcal{DI}$ computed is also quasi-linear. Investigations regarding some limitations are necessary because, in a real-world scenario, the behavior of a damage evolution may be more complicated than the structural change simulated in the present paper. However, this does not invalidate the results and applicability of the proposed strategy, as presented in the results.

To identify other complex damage conditions could be necessary to use other input signals, for example, random input, and with different temperatures and operational conditions can also be required to be able to conduct tests with a focus on simulation and not the

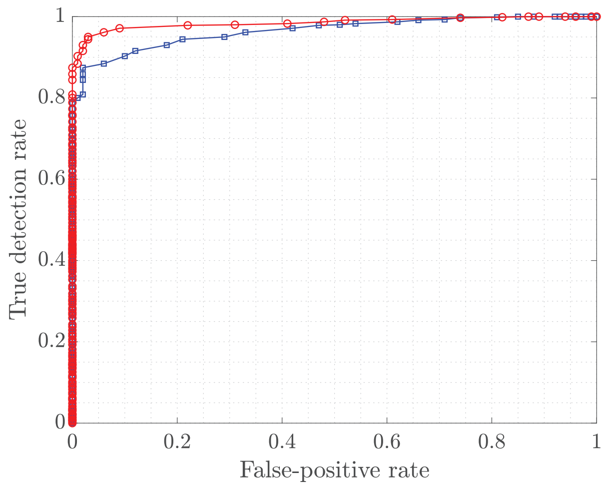


Figure 13. Receiver operating characteristics (ROC) curve: □ Extrapolated, and o Actual.

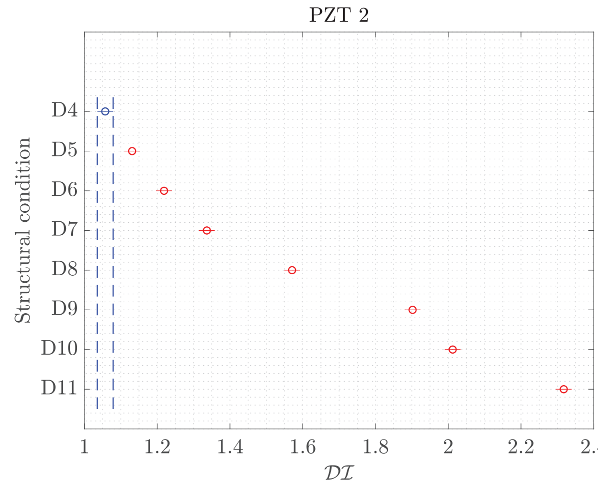


Figure 14. Classification of the clusters of different damage' states using ANOVA of $\mathcal{DI}$ assuming extrapolated model. 7 groups (in red) have means significantly different from healthy H.

only prediction as performed in this paper. Other models can also be employed, in particular, Gaussian Process (GP) combined with NARX models because of the ability of these models to predict with band confidence and admit a priori probability identified to extrapolate a future state. GP-NARX models could also treat the possible nonlinear mechanisms of damage and interactions between guided waves. A combination with numerical simulation using finite element models is further beneficial to correlate more intense and complicated damage progression with an extrapolated data-driven model, and some studies are started in this direction. The smoothness of the damage index extracted allowed us to use a cubic spline. If the damage mechanism presents some singularity, a piecewise of more complex as an interpolation technique could be required. In this case, Gaussian process regression is also feasibility to be used to interpolate the model parameters, as recently discussed by Amer and Kopsaftopoulos (2019), to quantify the extension of damage.

7. Conclusion

This paper demonstrated that multiple AR identified models in the healthy and in the initial damaged conditions could help perform interpolation in the coefficients assuming that the progress of the damage occurred similarly in the identical form for all growth in a given path. This kind of technique, combined with an approach for extrapolating, seeking to reach high levels of SHM's hierarchy, may provide a significant physical insight if we compare the trend curves computed by damage-sensitive index with some machine learning approach.

The extrapolated model estimated by the previous AR models demonstrated to be appropriately correlated with the actual state and well correlated with the severity that, in this case, was the covered area for the adhesive with a quasi-linear behavior. The extrapolated models demonstrate an adequate performance compared with the actual with similar behavior to allow us to calculate a damage-sensitive index before the simulated structural change. After that, the severity of the damage is propagated, an alert trigger. This damage index computed by the extrapolated model can help the user observe how the damage can evolve and simulated by a simple prediction black-box model of extrapolation.

Declaration of conflicting interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: The authors are thankful for the financial support provided from São Paulo Research Foundation (FAPESP) grants 2017/15512-8 and 2018/15671-1 and the Brazilian National Council for Scientific and Technological Development (CNPq) grant number 306526/2019-0.

ORCID iDs

Samuel da Silva https://orcid.org/0000-0001-6430-3746
Jessé Paixão https://orcid.org/0000-0002-2035-0986

Notes

1. Usually the coefficients a_i are normalized such that $a_0 = 1$.
2. That is, e_{unk} and e_{ref} belong to the normal distribution.
3. Different classifiers can provide adequate results; the requirement here is to have the initial damage well classified to estimate a model for extrapolating.

4. It is related to the time when this damage progress.
5. https://www.mathworks.com/help/matlab/ref/interp1.html

References

Amer A and Kopsaftopoulos F (2019) Probabilistic damage quantification via the integration of non-parametric time-series and gaussian process regression models. In: *12th International Workshop on Structural Health Monitoring (IWSHM 2019*. Stanford, CA, USA.

Balmes E, Guskov M and Bianchi JP (2016) Validation and verification of fe models of piezo based shm systems. In: *ISMA 2016 - International Conference on Noise and Vibration Engineering*. Leuven, Belgium.

Corbetta M, Sbarufatti C, Giglio M, et al. (2018) A bayesian framework for fatigue life prediction of composite laminates under co-existing matrix cracks and delamination. *Composite Structures* 187: 58–70.

da Silva S (2018) Data-driven model identification of guided wave propagation in composite structures. *Journal of the Brazilian Society of Mechanical Sciences and Engineering* 40(11): 543.

Fendzi C, Mechbal N, Rébillat M, et al. (2016) A general bayesian framework for ellipse-based and hyperbola-based damage localization in anisotropic composite plates. *Journal of Intelligent Material Systems and Structures* 27(3): 350–374.

Figueiredo E, Figueiras J, Park G, et al. (2011) Influence of the autoregressive model order on damage detection. *Computer-Aided Civil and Infrastructure Engineering* 26(3): 225–238.

Figueiredo E, Park G, Farinholt KM, et al. (2012) Use of time-series predictive models for piezoelectric active-sensing in structural health monitoring applications. *Journal of Vibration and Acoustics* 134(4): 041014.

Goidescu C, Welemane H, Garnier C, et al. (2013) Damage investigation in cfrp composites using full-field measurement techniques: combination of digital image stereo-correlation, infrared thermography and x-ray tomography. *Composites Part B: Engineering* 48: 95–105.

Gonsalez CG, da Silva S, Brennan MJ, et al. (2015) Structural damage detection in an aeronautical panel using analysis of variance. *Mechanical Systems and Signal Processing* 52–53: 206–16.

Hogg RV (1987) *Engineering Statistics*. Michigan: Macmillan Pub Co.

Kim J, Kim K and Sohn H (2014) Subspace model identification of guided wave propagation in metallic plates. *Smart Materials and Structures* 23(3): 035006.

Kopsaftopoulos F and Fassois S (2010) Vibration based health monitoring for a lightweight truss structure: experimental assessment of several statistical time series methods. *Mechanical Systems and Signal Processing* 24(7): 1977–1997.

Larrosa C, Lonkar K and Chang FK (2014) In situ damage classification for composite laminates using gaussian discriminant analysis. *Structural Health Monitoring* 13(2): 190–204.

Lee JS, Park G, Kim CG, et al. (2011) Use of relative baseline features of guided waves for in situ structural health monitoring. *Journal of Intelligent Material Systems and Structures* 22(2): 175–189.

Ljung L (1998) *System Identification*. New Jersey, US: Prentice-Hall PTR.

Mechbal N, Uribe JS and Rébillat M (2015) A probabilistic multi-class classifier for structural health monitoring. *Mechanical Systems and Signal Processing* 60–61: 106–123.

Mitra M and Gopalakrishnan S (2016) Guided wave based structural health monitoring: a review. *Smart Materials and Structures* 25(5): 053001.

Nardi D, Lampani L, Pasquali M, et al. (2016) Detection of low-velocity impact-induced delaminations in composite laminates using auto-regressive models. *Composite Structures* 151: 108–113.

Paixão JAS, da Silva S and Figueiredo E (2020) Damage quantification in composite structures using autoregressive models. In: Wahab MA (ed.) *Proceedings of the 13th International Conference on Damage Assessment of Structures*. Singapore: Springer Singapore, pp. 804–815.

Shiki SB, da Silva S and Todd MD (2017) On the application of discrete-time volterra series for the damage detection problem in initially nonlinear systems. *Structural Health Monitoring* 16(1): 62–78.

Wang Y (2011) *Smoothing Splines: Methods and Applications*. Chapman Hall/CRC Monographs on Statistics and Applied Probability. Chapman and Hall/CRC.

Wolberg G and Alfy I (2002) An energy-minimization framework for monotonic cubic spline interpolation. *Journal of Computational and Applied Mathematics* 143(2): 145–188.