DATA-DRIVEN AUTOREGRESSIVE MODEL IDENTIFICATION FOR STRUCTURAL HEALTH MONITORING IN ANISOTROPIC COMPOSITE PLATES

SAMUEL DA SILVA*,*, JESSÉ PAIXÃO*, MARC RÉBILLAT† AND NAZIH MECHBAL†

*,*Departamento de Engenharia Mecânica, Universidade Estadual Paulista - UNESP
Av. Brasil 56, 15385-000, Ilha Solteira, SP, Brasil
e-mail: samuel.silva13@unesp.br, jesse.paixao@unesp.br,
Web page: http://www.dem.feis.unesp.br/

† PIMM Laboratory, Arts et Métiers - ENSAM/CNRS/CNAM
151 Boulevard de l'Hôpital, 75013 Paris, France
e-mail: marc.rebillat@ensam.eu, nazih.mechbal@ensam.eu,
Web page: https://pimm.artsetmetiers.fr/

Key words: AR Models, Multiple Models, Extrapolated Model, Quantification, Prognosis

Abstract. A simple data-driven AutoRegressive (AR) model may be used to assess a model to describe and to predict the time-series outputs of the PZT sensors receiving Lamb waves for different operating conditions in composite structures. Thus, this paper presents the potentiality of the use of a set of AR models to detect, locate, and, manly, to extrapolate a damage sensitive index based on changes in one-step-ahead prediction errors. To illustrate this proposal, an aeronautical composite panel with bonded piezoelectric elements, that act both as sensors and actuators, is used to study the relationship between the variation of the parameters of the identified model and the presence of various simulated damage. A damage progression evaluation by extrapolating the AR parameters is also suggested and examined based on cubic spline functions to verify the future state and to observe how the damage could evolute, based on some simplified assumptions. This step could help to make a decision about a possible required repair without adopting a complicated and costly physical model.

1 INTRODUCTION

The structural health monitoring (SHM) approaches seem to be in a mature stage in the steps of detecting and localization of possible damages in structures with several powerful methods proposed and validated in the last decades. One way to adress this issue may be using data-driven model identification based on guided wave propagation or random inputs. For example, Nardi et al. [1] 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. [2] also showed a possible data-driven system identification through a state-space model to capture the wave motion in metallic structures. da Silva [3] applied autoregressive with exogenous input (ARX) model to perform predictions and a waveform generator in a 10 layers carbon-epoxy plate excited by guided waves assuming different central frequencies and environmental conditions[1] and noted some benefits and disadvantages of the possible performance of this strategy for SHM.

However, to reach a high level in an SHM's hierarchy[2] numerous drawbacks need to be overcome yet. One of them is that it is essential to have an adequate mathematical model if the user wants to predict a future state based on previous past data to interrogate about the existence and evolution of damage propagation. So, these models demand to incorporate information about the damage behavior in its dynamics to gain a comprehensive physical insight of the monitored structure; consequently, this model should be most physical possible. However, to construct numerical models, for example, using finite element models with damages in an initial stage, require much time and may have a high cost for a real-time monitoring system in the industrial field, even modeling in its healthy state. Another limitation is that the behavior of damage evolution usually is much complicated to be modeled in a real-world application because complex types of damage can appear coincidentally with several confounding effects, like noise, uncertainties, temperature changes, operational variability, etc..

An identified data-driven model, as suggested by Nardi et al. [1] or da Silva [3], could be attractive to be adopted to extrapolate or to quantify a damage progression as a surrogate model to reach a subsequent application of higher forms of SHM's hierarchy. Thus, this paper is a first effort of the authors in this direction seeking to extrapolate AR polynomials through spline functions to extend how damage-sensitive index could evolute based on simplified assumptions. The proposed procedure has two steps to be implemented. First, damage detection and location using an index extracted by predictions errors filtered with a reference AR model is performed. To classify the structural states an analysis of variance is utilized [4]. Next, when damage is detected a new set of models, named by *initial damaged models*, is captured to extrapolate a projected state. A carbon-epoxy laminated plate with controlled progressive structural change similar to a real damage is used to exemplify the method. Next sections describe these steps and final remarks.

2 DAMAGE DETECTION USING PREDICTION ERRORS

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 [5]:

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

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[3] with the coefficients a_i with a lag-order n_a, e. g., $y(k)a_i q^{-i} = a_i y(k-i)$, where q^{-i} is a lag operator and k is the time sample. The order can be estimated using Akaike information criterion (AIC) and the polynomial $A(q)$ may be identified through a least squares or Yule-Walker approach, fully available in Matlab or Octave software. When a new normalized

[1] temperatures changes.

[2] for example, quantification and prognosis.

[3] Usually the coefficients a_i are normalized such that $a_0 = 1$.

data, $x(k)$, in an unknown state is measured, one can try to 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 [6]:

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

If $\mathcal{DI}$ belongs a $\mathcal{F}-$distribution[4] there is no damage and the unknown condition is associated to healthy state (null hypothesis $\mathcal{H}_0$ is true). On another 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 [7, 6].

To classify the cluster of damages states, a one-way analysis of variance (ANOVA) can be also used to test the hypothesis that the samples in the running tests $\mathcal{DI}$ belong to a population with the same means (null hypothesis $\mathcal{H}_0$), i.e., the systems is classified as healthy state, against the alternative hypothesis $\mathcal{H}_1$ that the population means are not all the same, i. e., damaged state [8, 4]. A Tukey's multiple comparison test to decide whether the results of ANOVA are statistically significant is also performed to enable us visible to distinguish the clusters correlated with different damages.

3 EXTRAPOLATION OF AR COEFFICIENTS

After clustering using ANOVA[5], a new set of AR models for each initial damage recognized is estimated since the reference model given by Eq. (2) is not anymore 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 state and do 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, i.e., $A(q) \approx \mathscr{A}_j(q)$, once $|a_i - \mathcal{A}_{ji}| < \delta$ for all i and δ is a small value. It is essential to observe that the damage index given by eq. 4 is sensitive to the changes to detect damage, but the specific changes in the coefficients are in general smooths when this structural variation is in the initial states.

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

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

[4] i.e., e_{unk} and e_{ref} have a normal distribution

[5] Different classifiers can provide adequate results; the requirement here is to have the initial damage well classified to estimate a model for extrapolating.

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. To enable to perform adequate extrapolations of the AR coefficients in the reference and damaged conditions, the number of states needs to be $n_d \geq 3$; otherwise, only linear extrapolation is plausible. So, it was assumed $n_d = 3$ structural states in damage clusters to permit to use cubic splines. This piecewise polynomials can be used to extend each polynomial coefficient λ_i associated to a future state through [9]:

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

where $z_0 < \cdots < z_{n_a}$ are defined intervals associated with the order n_a, $f_i(z)$ a smoothing spline estimate by some minimizer of a penalized criterion, and ξ_i is an independent random error. More details to find the spline f_i can be seen in [10].

The important issue here is to recognize that now it is possible to have a future model to predict the data $x_j(k)$, when $j = n_d + 1$ represents a future state. This 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 lines of matrix $\mathbf{A}$. Thus, a new index and hypothesis tests can also be estimated to evaluate the progression and the changes of distribution of the extrapolated prediction error $\varepsilon_j(k)$ and in the damage index $\mathcal{DI}$.

4 EXPERIMENTAL EXAMPLE

Figure 1 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 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. 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.

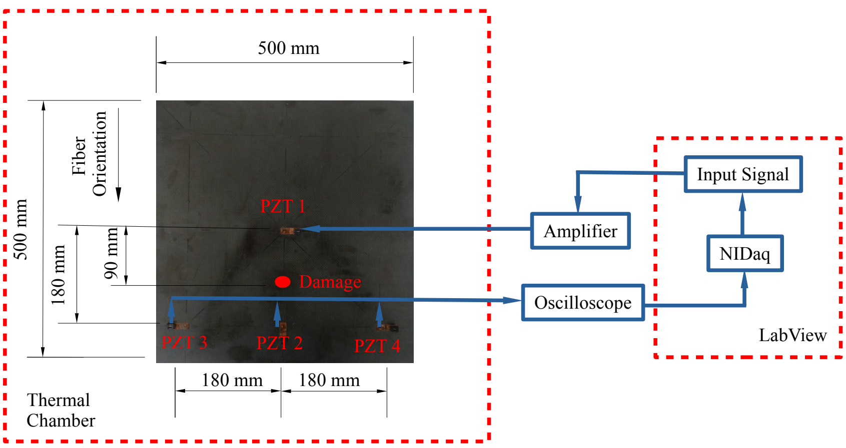


Figure 1: Illustration of the experimental setup with details about the geometry and instrumentation utilized.

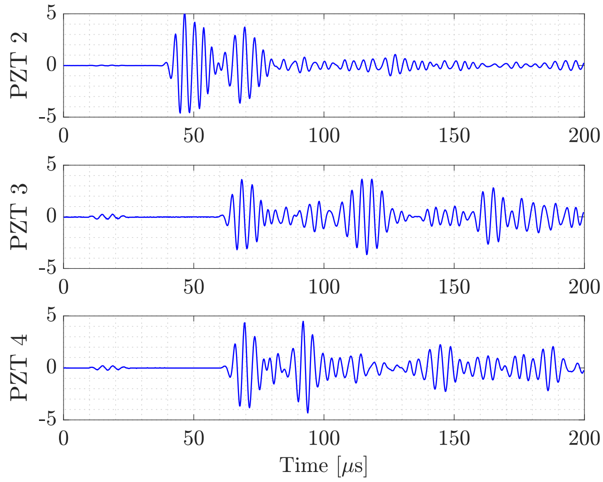


Figure 2: Healthy output time-series when PZT 1 acts as an actuator with a central frequency of 250 kHz.

Figure 2 illustrates the output time-series measured by PZTs in a healthy state normalized by Eq. 1. An industrial adhesive putty was glued on the plate surface to simulate gradual damage by an additional mass increasing progressively the coverage area in the path between PZT 1 and PZT 2. This change modifies local material properties with a similar effect to the real damages in composites structures as performed by Lee et al. [11]. In each structural state, 100 tests were repeated for an adequate statistical characterization of the proposed damage detection.

4.1 AR model identification

The AIC order selection with a focus of prediction indicates that order of $n_a = 40$ is sufficient to give an adequate validation for all paths of propagation, as seen in fig. 3. 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 executed to reduce computational processing time. Other realizations were filtered using this reference model by Eq. 2 to estimate the prediction error in the healthy state.

Figure 4(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, i. e., the model has identified adequately, as observed in the fig. 4(b).

4.2 Damage detection

Once a reference model, named by H, is correctly identified, it is used to detect some possible structural change. Only half of the 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 progressive damage severity associated with area covered given by 490 mm^2 (0.19%), 707 mm^2 (0.28%) and, 962 mm^2 (0.38%), respectively. Each structural condition was also measured by 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 Eq. 3 and then the damage index $\mathcal{DI}$ is estimated by each test using the Eq. 4. The Lilliefors test is used to warranty that the variance of the prediction errors $\sigma^2(\cdot)$ in healthy states have normal

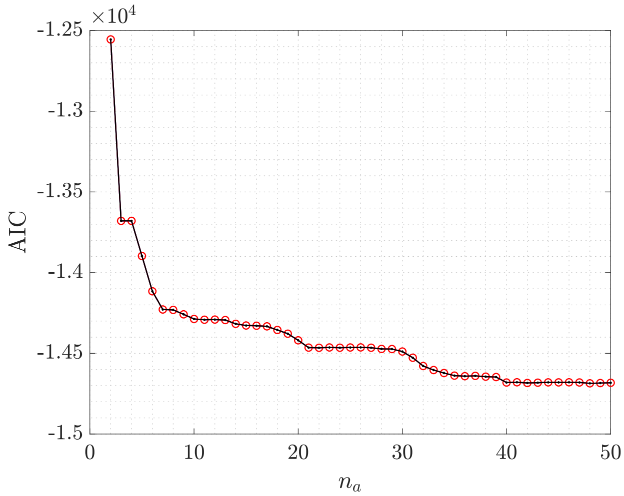
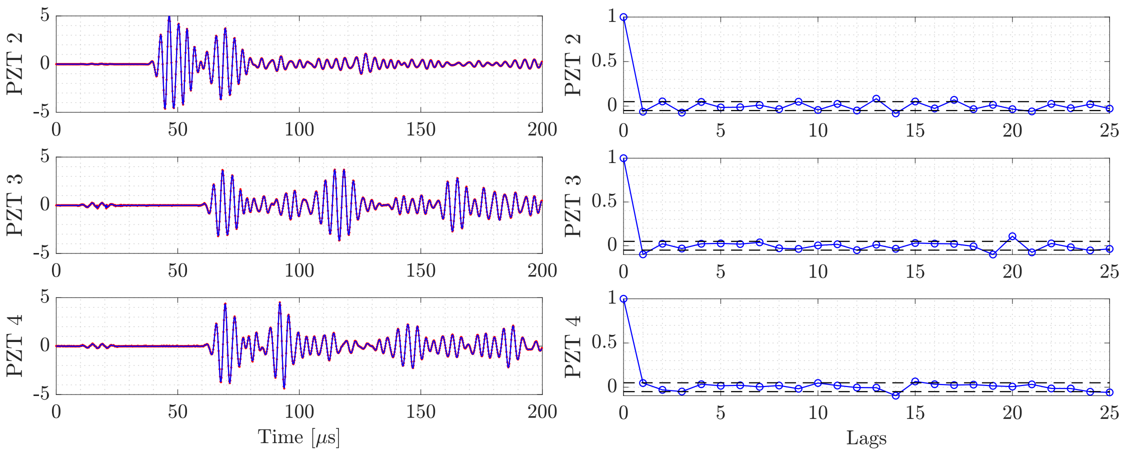


Figure 3: AIC order selection with focus in prediction.



(a) – Measured vs ··· Predicted.

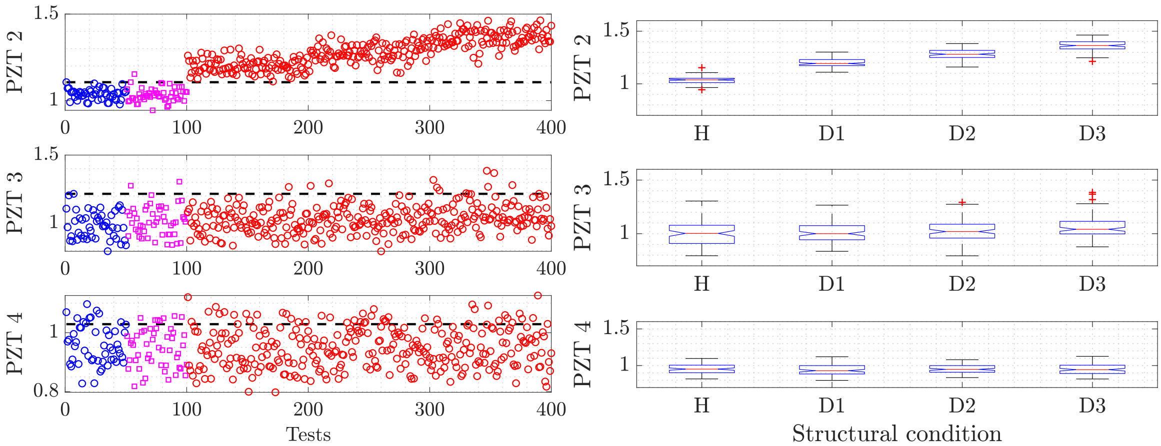
(b) Whiteness test criteria within of 95% confidence bound patch.

Figure 4: Healthy output predicted by healthy model AR(40).

distributions, i. e., the $\mathcal{DI}$ trends to have a $\mathcal{F}-$ distribution in the reference condition [6].

Figure 5(a) illustrates the index $\mathcal{DI}$, where 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 5(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 6 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 that is achieved by PZT 2, that 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 a Tukey multiple comparisons to see whereby statistically significant the clusters are.



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

Figure 5: 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.

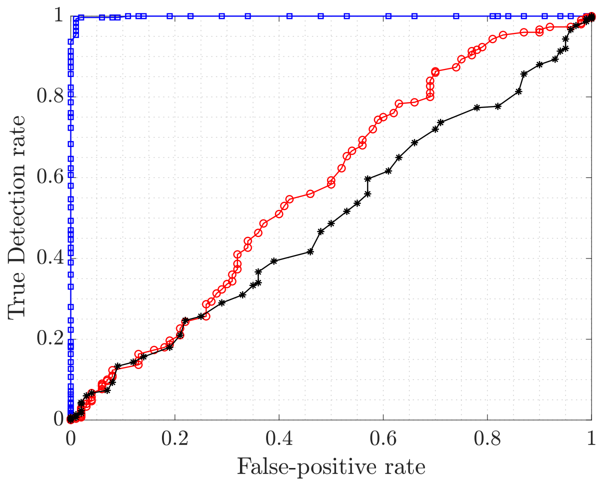


Figure 6: Receiver Operating Characteristics (ROC) curve: □ PZT 2, o PZT 3 , and $*$ is PZT 4.

Figure 7 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 to implement the extrapolation explained in the next section. Other PZTs are not used here to identify new models because none variations are observed permitting the identification of a new polynomial $\mathscr{A}_j(q)$ in Eq. 5.

4.3 Prevision of future AR damaged model

Three damaged models are estimated using one of the conditions presents in the clusters in fig. 7 after detecting and clustering. Combined with the reference model, a matrix given by Eq. 6 is formed with $n_d = 3$ and $n_a = 20$ to help the extrapolation. A cubic spline polynomial created with the data in matrix

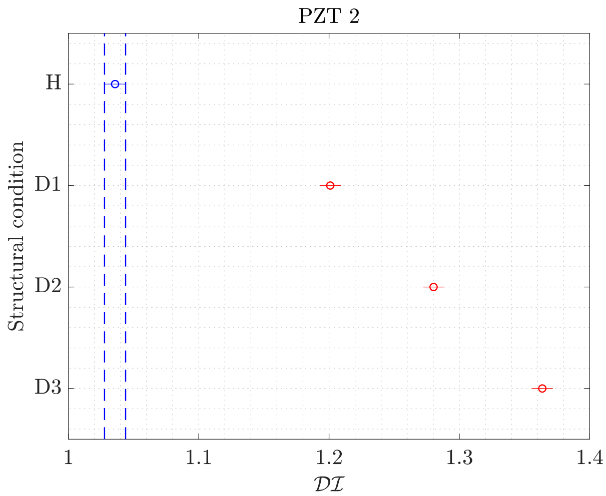
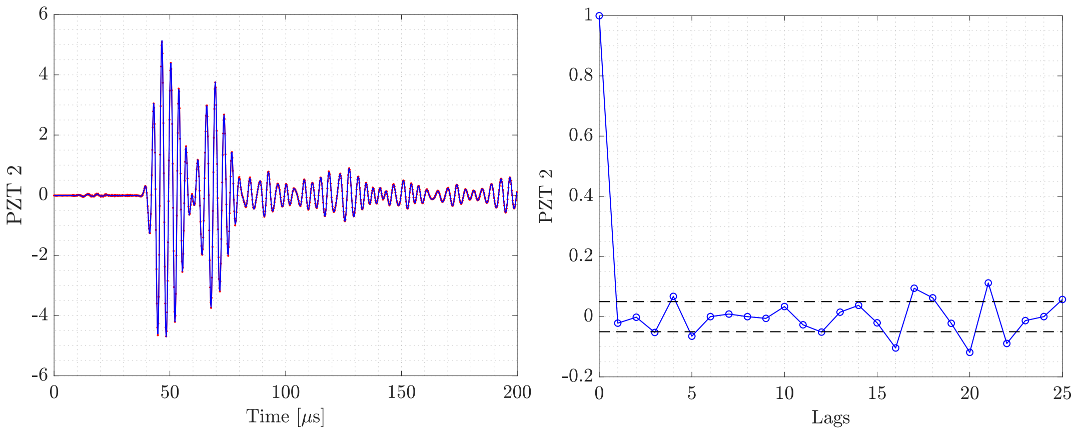


Figure 7: Classification of the clusters of different damage' states using ANOVA of $\mathcal{DI}$. 3 groups (in red) have means significantly different from healthy H.

A permits to estimate an extrapolated model $\Lambda(q)$ to use Eq. 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. First, a damage D4 with a surface area covered of 0.5% is used to predict the future state. Figure 8(a) presents the comparison between the measured and predicted by the extrapolated model in damage condition D4. Figure 8(b) shows the residual analysis by the autocorrelation function to confirm an adequate prediction.

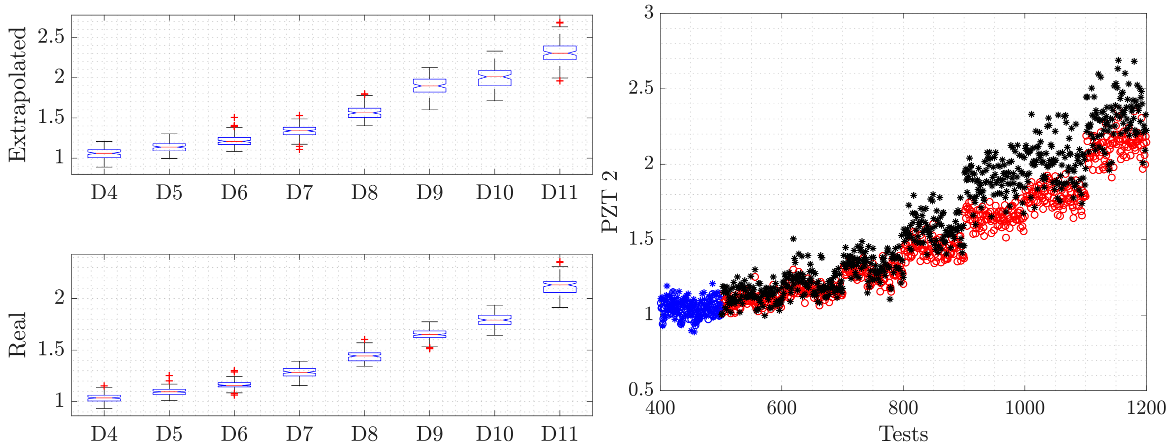


(a) – Measured vs ··· Predicted.

(b) Whiteness test criteria.

Figure 8: Output predicted by the extrapolated model AR(40) for damage condition D4.

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 when the damage happens, now assuming the new reference as D4. The box-plot is also presented in fig. 9(a) for the other damage conditions to observe its similarity comparing the real 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. Figure 9(b) illustrates the predictions obtained by the filtered model using the real data (measured) and extrapolated model. It is found a more significant dispersion 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 level of severity of the damage, once the extrapolation is done considering does not modify the framework and order of the AR model. 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 a real model identified when the damage increases and is helpful to make decisions.



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

Figure 9: 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 real damage D4 represents the new reference, $*$ extrapolated damaged D4 (extrapolated new reference), $*$ extrapolated damage, and real Damaged (o) is the future damaged condition.

Figure 10 shows the ROC curve comparing the similar performance between the extrapolated and the real model to detect all future damage conditions. The extrapolated model in structural state D4 also allows dividing the clusters associated with the other damages using ANOVA, as seen in fig.11.

5 CONCLUSIONS

This paper demonstrated that multiple AR identified models in the healthy and in the initial damaged conditions could be helpful to perform an extrapolation to a future state if the damages progress in a similar way. This kind of technique combined with a Bayesian approach for extrapolating, seeking to reach high levels of SHM's hierarchy, may provide a significant physical insight if we compare with the trend curves computed by damage-sensitive index with some machine learning approach. First, a model is identified associated with parameters that can predict future states or conditions before the occurrence. Of course, the basic premise is that the evolution of the initial damage is not abrupt and does not happen in different points. Investigations regarding these issues are necessary because in a real-world scenario the behavior of a damage evolution is complex.

To identify with other input signals, for example, random input, and with different temperatures and operational conditions is required to be able to conduct tests with a focus in a simulation and not the only prediction. Other models can also be employed, in particular, Gaussian Process (GP) combined with AutoRegressive models because of the ability of these models to predict with band confidence and admit

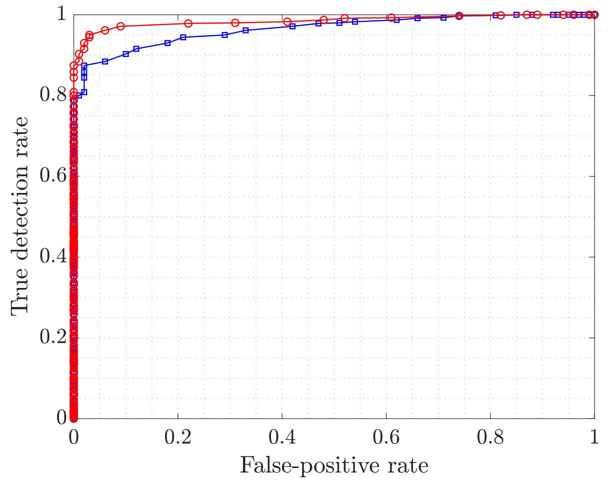


Figure 10: Receiver Operating Characteristics (ROC) curve: □ Extrapolated, and o Real.

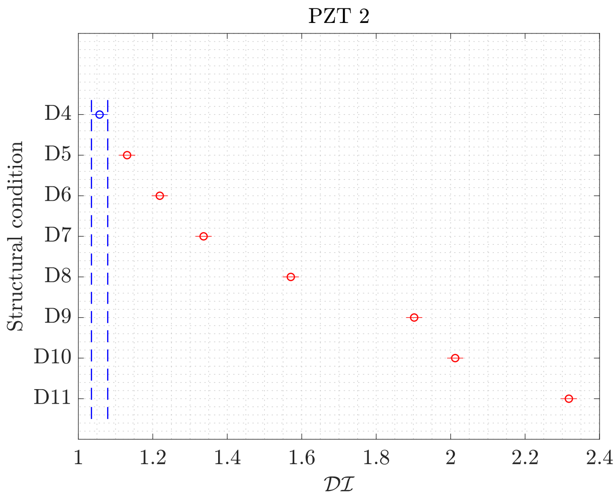


Figure 11: Classification of the clusters of different damage' states using ANOVA of $\mathcal{DI}$ assuimg extrapolated model. 3 groups (in red) have means significantly different from healthy H.

a priori probability identified to extrapolate a future state. Numerical simulation using finite element models are further beneficial to correlate more intense and complicated damage progression with an extrapolated data-driven model, and some studies are started in this direction.

ACKNOWLEDGEMENTS

The first and second 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 307520/2016-1.

REFERENCES

[1] Nardi, D., Lampani, L., Pasquali, M., Gaudenzi, P. Detection of low-velocity impact-induced delaminations in composite laminates using Auto-Regressive models. *Comp. Struct.* (2016) **151**: 108-

113. https://doi.org/10.1016/j.compstruct.2016.02.005

[2] Kim J., Kim, K. and Sohn, H. Subspace model identification of guided wave propagation in metallic plates. *Smart Mater. Struct.* (2014) **23**(3):035. https://doi.org/10.1088/0964-1726/23/3/035006

[3] da Silva, S. Data-driven model identification of guided wave propagation in composite structures. *J. Braz. Soc. Mech. Sci. Eng.* (2018) **40**:543. https://doi.org/10.1007/s40430-018-1462-4

[4] Gonsalez, C. G., da Silva, S., Brennan, M. J., Lopes Jr, V. Structural damage detection in an aeronautical panel using analysis of variance. *Mech. Syst. Signal Process.* (2015) **52-53**:206-216. https://doi.org/10.1016/j.ymssp.2014.04.015

[5] Ljung, L., *System Identification*, 2nd edn. Prentice-Hall PTR, Upper Saddle River, (1998).

[6] Shiki, S. B., da Silva, S., Todd, M. D. On the application of discrete-time Volterra series for the damage detection problem in initially nonlinear systems. *Struct. Health Monitor.* (2017) **16**(1): 62-78. https://doi.org/10.1177/1475921716662142

[7] Kopsaftopoulos, F., Fassois, S. Vibration based health monitoring for a lightweight truss structure: Experimental assessment of several statistical time series methods. *Mech. Syst. Signal Process.* (2010) **24**:1977-1997. https://doi.org/10.1016/j.ymssp.2010.05.013

[8] Hogg, R. V., Ledolter, J. *Engineering Statistics*, Mac Millan Publishing Company, University of Michigan, (1987), ISBN:9780023557903.

[9] Wolberg, G., Alfy, I. An energy-minimization framework for monotonic cubic spline interpolation. *Journ. of Comp. and Applied Math.*. (2002) **143**:145-188. https://www.sciencedirect.com/science/article/pii/S0377042701005064?via%3Dihub.

[10] Brien, C. M. Smoothing splines: Methods and applications by Yuedong Wang. *Intern. Statistical Rev.* (2012) **80**(3):475-476.

[11] Lee, J. S., Park, G., Kim, C. G., Farrar, C. R. Use of relative baseline features of guided waves for in situ structural health monitoring. *Journal of Int. Mater. Syst. and Struct.* (2011) **22**(2):175-189. https://doi.org/10.1177/1045389X10395643.