Promoting Novel Strategies for the Reliability Assessment of Guided Wave Based SHM Systems

VITTORIO MEMMOLO, JOCHEN MOLL,
OLIVER SCHACKMANN, STEFFEN FREITAG,
ANASTASIIA VOLOVIKOVA, KILIAN TSCHOKE, ENES SAVLI,
YEVGENIYA LUGOVTSOVA, MARIA MOIX-BONET,
AHMED BAYOUMI and INKA MUELLER

ABSTRACT

Detection of defects is of outmost importance for composite airframe due to barely visible damage which may appear at any time during lifetime. Several Structural Health Monitoring (SHM) technologies have been developed so far to continuously monitor the condition of the structure. Guided Ultrasonic Waves (GUW) are well suited for this application, but there is still a significant gap in reliability assessment and quantification of system performances. Although the theory of Probability of Detection (POD) can be used to this end, but the reliability assessment procedure needs to be properly addressed for each type of system. In this work different concepts are proposed, including two artificial intelligence approaches, tested and compared to enable a reliability assessment comparable to the conventional POD framework. On top, a novel approach on how to combine experimental data of the undamaged structure with simulations of the damaged structure is proposed. The proposed ideas are tested on a single carbon fiber composite specimen, and data is analyzed using all the different concepts. The results show that classic POD and AI-based reliability assessment can be compared as well as experimental and synthetic data can be used when the experimental variability is matched properly providing a paradigm shift in the reliability assessment field.

INTRODUCTION

Reliability assessment is a major problem for SHM systems. The community is missing methods on how to show that specific SHM systems meet the requirements within a certification process [1]. Only few approaches exist, mainly trying to adopt the procedure of Probability of Detection (POD) known from Non-Destructive Testing (NDT) to SHM are available [2, 3]. These approaches all fight with drawbacks resulting from the differences between NDT and SHM, like the fixed sensor positions, the dependence of

Vittorio Memmolo, University of Naples “Federico II”, Italy
Jochen Moll, Oliver Schackmann, Goethe University Frankfurt, Germany
Steffen Freitag, Anastasiia Volovikova, Karlsruhe Institute of Technology, Germany Kilian Tschoke, Enes Savli, Fraunhofer IKTS, Dresden, Germany
Yevgeniya Lugovtsova, BAM, Germany
Maria Moix-Bonet, DLR Braunschweig, Germany
Ahmed Bayoumi, Inka Mueller, Bochum University of Applied Sciences, Germany

POD on damage location as well as high costs, especially for advanced SHM systems applied to sophisticated structures. To prevent the necessity of expensive experimental campaigns, approaches based on simulation only have been developed, including the location dependency by using POD maps [4] or including SHM specific variations in the simulation [5].

This paper presents new thoughts and ideas on how to tackle the problem. Instead of presenting these separately, a brief overview on the concepts to be discussed comparatively in the context of the conference is given. Three non conventional concepts will be presented: first, an AI-based concept using Artificial Neural Networks (ANN), second, an AI-based concept using Convolutional Neural Networks (CNN) and third, a Model-Assisted POD (MAPOD) concept, combining simulation and experiment. To show the applicability of these concepts, the same dataset from the experiment is used, namely raw data from the *Open Guided Waves* (OGW) platform [6]. While the AI-based approaches are using experimental data only, the MAPOD approach necessarily needs simulation data, which is added to the existing experimental dataset. The further steps on data processing are specific for each concept.

The following section describes the data set and the three concepts. It is followed by a section presenting the results, which are summarized in the conclusion.

DESCRIPTION OF DATA SET AND DESCRIPTION OF CONCEPTS

Open Guided Wave Dataset

The experimental dataset used for testing the concepts is available at the OGW online platform and has been acquired using a carbon fibre reinforced polymer (CFRP) specimen equipped with a network of piezoelectric transducers, see [7] for detailed information on the manufacturing and inspection of the plate and [6] for additional information regarding the omega stringer and the data acquisition. A Hann-windowed sinusodial signal with a central frequency of 40 kHz is used for excitation. To imitate the effect of damage on the guided wave propagation, artificial reversible damage of multiple sizes is employed.

For the MAPOD concept, the necessary simulation data set is based on Elastodynamic Finite Integration Technique (EFIT) simulations. The theoretical background is described in detail in [8, 9]. The advantages of this simulation strategy are the very fast calculation while being able to accurately describe also more complex structures with curved edges, stringers, etc. The simulated dataset as well as the experimental dataset are shown in Figure 2(a).

To assess the condition of the structure, a damage indicator (DI) is defined by evaluating the energy difference between two signals as:

$$DI_E = \sum_{k=1}^{N} \left(x_C(k) - x_B(k)\right)^2 \tag{1}$$

where $x_B(k)$ represents the discrete signal response of the pristine state and $x_C(k)$ of the current state in Volts at time step k. N represents the maximum number of data points. Here, the pristine state is calculated as the average of five signals measured at

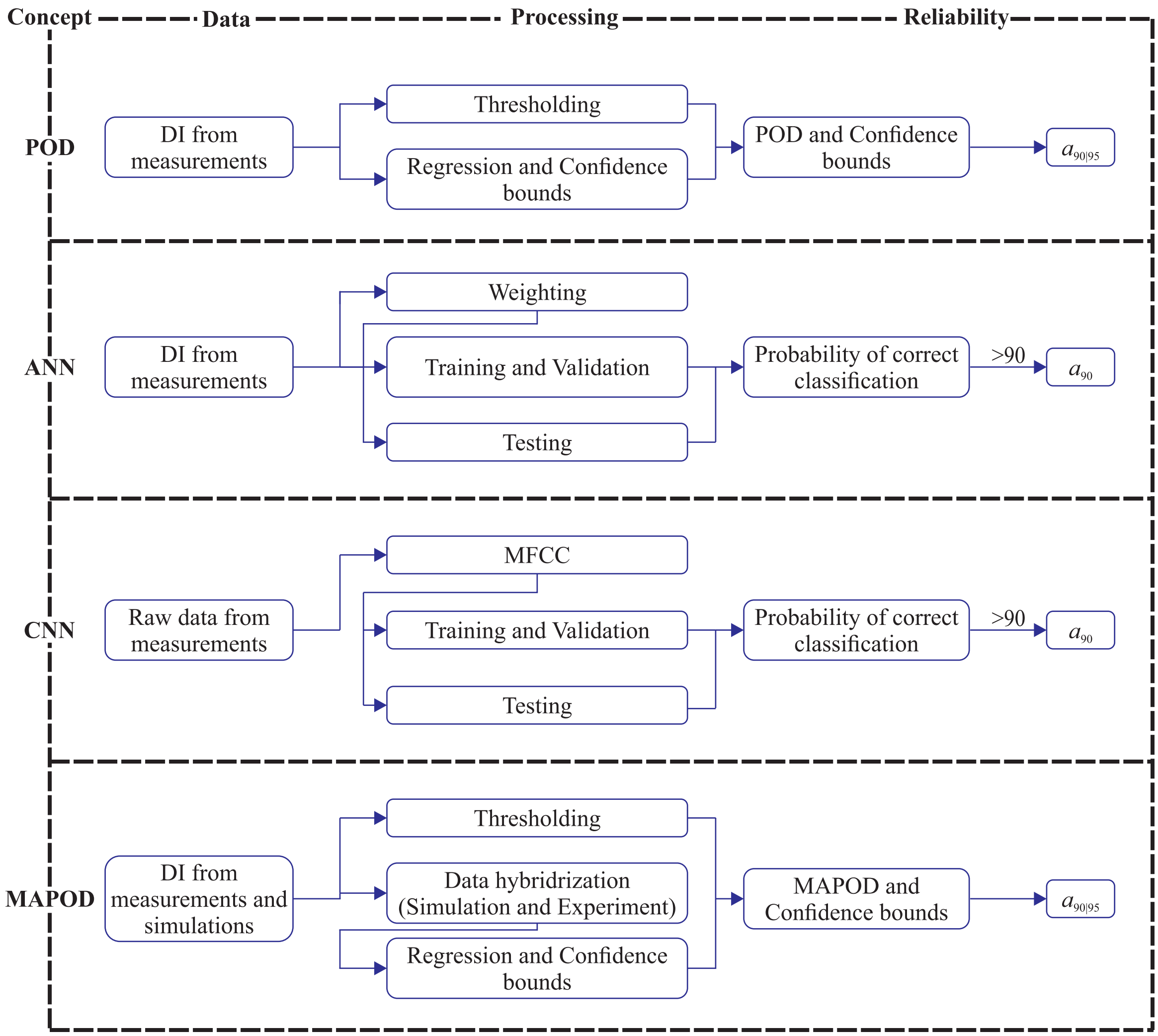

Figure 1. Comparison of reliability assessment steps for the different proposed concepts.

undamaged state. A close to zero DI value indicates a healthy state, while increasing DI values correlate with growing defect sizes.

Comparison of POD, AI and MAPOD Concepts

The reliability assessment of an SHM system aims to evaluate whether a required minimum defect is detected with a statistically meaningful confidence and at an acceptable level of false alarms. Although many approaches are available, POD is still the most accepted metric to return the minimum size of the damage, which is detectable by the NDT/SHM methods. However, new concepts need to be investigated to accomplish the necessity either to assess new diagnostic algorithms (AI) or introduce synthetic data for predictions (MAPOD). Every concept requires a specific procedure for reliability assessment, whose steps are summarized in Figure 1 and described hereinafter.

The ***classical POD concept*** relies on the use of measurement derived data to build up the POD versus the flaw size, more details can be found in [10, 11]. Generally, there are two approaches to estimate the POD curve; the Hit/Miss method and $\hat{a}$ vs a

method. Among different approaches, the $\hat{a}\ vs.\ a$ method is well suited to look into DI-based SHM systems because it can be used as the signal response ($\hat{a}$) to predict damage dimension (a). Employing the $\hat{a}\ vs.\ a$ approach, it is possible to estimate the signal response of an SHM system versus flaw dimension in a statistically meaningful way and the inherent POD is defined as:

$$POD(a) = \int_{\hat{a}_{dec}}^{\infty} f_{\hat{a}|a}\left(\hat{a}\right)\, d\hat{a}\ , \tag{2}$$

where $f(\hat{a})$ is the normal probability density function around the predicted response. Applying $95\ \%$ confidence bounds to account for prediction errors, the flaw dimension $a_{90|95}$ which returns $90\ \%$ POD is generally referred as minimum detectable damage, as it can be detected by the system in a statistically significant way. The $a_{90|95}$ value calculated for each sensor pair individually can be considered as means of comparison with non-classical approaches, together with the false alarm ratio established. It is worth noting that the threshold of the damage indicator is tuned to be twice the noise level.

One of the major efforts of this work consists in the use of AI-based predictors for damage detection and quantification of the reliability with a POD-like concept. The main idea of the AI-based method, used in this publication, is to take into account all paths of the experimental setup from *OGW* to train and test a neural networks in classifying whether a damage is present or not and assess the reliability of such a network through the probability of correct classification.

The ***ANN concept*** is based on the idea introduced in [12] to use the damage indicator (DI) as given by Eq. (1) as input of a feed-forward ANN to get a binary classification of the plate state (damaged/undamaged) as output. To tune the unknown network parameters and test the generalization performance of the ANN to unseen data, the OGW data set is divided into a training and a generalization testing data set. The training data set consists of 20 samples (DIs) for the undamaged (baseline) condition and 5 samples for each of 13 different damage sizes at two damage locations D2 and D3, both along all 65 actuator-receiver paths except the path 3-9, i.e. 9750 samples. After the ANN-model is trained, the generalization performance is evaluated with a generalization testing data set, which consists of 20 undamaged and 5 damaged samples for each damage size and damage location D1 along only path 3-9, which leads to 85 samples for the generalization testing.

The fully connected ANN consists of one input neuron (damage indicator), two hidden layers with 8 neurons per layer and one output neuron, which classifies the input value of the damage indicators to either undamaged or damaged condition. The hidden neurons are activated by the hyperbolic tangent sigmoid function (tansig) and the softmax activation function is used in the output neuron. The Levenberg-Marquardt-Algorithm is applied to train the ANN, where 70 % of the 9750 samples are used for the identification of the unknown network parameters (synaptic weights and bias values), 15 % of the samples are employed for validation and 15 % for testing the training performance, respectively.

To tune the false alarm rate, a weighting of the training errors according to the damage size is introduced according to [12]. The weighting factor for the damaged samples is computed for each damage size as the ratio of the area of damage and the maximal area of damage. This leads to weighting factors between 0.024 for the smallest damage size and 1.0 for the largest damage size. Due to the low amount of undamaged samples (just 13 %), the undamaged samples are weighted by a factor of 1.8, which results in a false alarm rate of 0.15 in the training data set.

The ***CNN concept*** relies likewise on the use of a trained CNN-model to distinguish damaged from undamaged plate. Otherwise, CNNs have the advantage of processing data instead of using scalar information thereof. The idea is again to feed CNN with Guided Ultrasonic Wave (GUW) measurements for the binary classification of damage (hit/miss). However, due to their structure, CNNs are particularly suited to process images. To this end, GUW signals are converted into a suitable (image-like) form by transforming the time history into a spectrogram (SP) using a short-time Fourier transform. This latter has the advantage of preserving most of frequency and time information and representing a suitable input for the machine learning algorithm. Nonetheless, to enable a data-efficient training phase, the use of smaller images is preferable. Therefore, the signals are further compressed into Mel-Frequency Cepstrum Coefficients (MFCC) matrices and then used to discriminate between undamaged and damaged status.

The CNN structure presented in [13] successfully applied to the Fashion-MNIST data set served as a template. This has been designed as a slightly more sophisticated version of the MNIST dataset, which is very well known for testing machine learning algorithms, and consists of 70000 grayscale images of size 28x28 of fashion items from ten categories. Since the size of these images is very similar to those of the MFCC matrices (20x32), the same model architecture should work well for the task at hand. Further details are reported in [14]. The data sets for training and generalization testing of the CNN-model are the same as used for the ANN-model.

To quantify the reliability of the AI-based results, the probability of correct classification is calculated for the training and generalization testing phase. To get the minimal detectable damage size and compare the results with the POD approach, the probability of correct classification is evaluated for each damage size available in the OGW data set and for the probability level of 0.9, the corresponding damage size is back computed by a linear interpolation between the closest available damage sizes.

The ***MAPOD concept*** in general aims to involve a statistically independent data set using simulation, but most approaches are based on simulation data only, thus, being more a simulation-based POD than a model-assisted POD [4,15–18]. The MAPOD concept proposed in this work consists of a novel approach on how to combine experimental data of the undamaged structure with simulations of the damaged structure in order to use the experiment to account for several uncertainties that can not be simulated [19]. Its procedures are similar to the POD approach. From linear regression analysis of the simulation data, a POD curve is generated. But the experimental input is used to adopt the slope, intercept and normal distribution around the regression line. The experimental measurements are recorded from one damage-free structure, while damage scenarios are simulated using EFIT.

The idea is that the influences leading to data scattering are independent in the experiment (variation of DI at undamaged state) and simulation (variation of DI around the regression line at damaged state). The scattering around the linear regression line of the simulated data set and the influential parameters of the experimental campaign are both normally distributed. Therefore, adding randomly selected values within the normal distribution interval of the experiment baseline measurements to the simulation data sets is possible. Then, subtracting these two normally distributed values results in varying the linear regression analysis to:

$$\mathbf{Y} = m_c\mathbf{X} + (\beta_c) \pm \mathbf{N}(0, \tau_c) \tag{3}$$

where m_c, β_c is the slope and intercept of the combined data set, and τ_c is the standard deviation of the scattering around the regression line with a mean value equal to zero. The combined linear regression parameters result from combining experimental baseline measurements with the simulated damage states dataset. This approach is considered a conservative approach [18].

RESULTS

As the traditional POD based on the experimental data is only used as a reference for all other results, the POD and MAPOD results are shown first, followed by the results of the two AI concepts. Figure 2(a) displays the relationship between the damage indicator (DI) and damage size A for the experimental and simulational damage scenario data sets. The results show a good agreement between the two data sets, indicating the accuracy of the EFIT numerical model. Figure 2(b) shows the POD and MAPOD results of the data set analyzed under the same threshold (0.015) for a damage size that varies between [50-1000] mm^2. The blue lines represent the classical POD approach evaluated using the experimental data of the undamaged and damaged state. The fuchsia lines give the

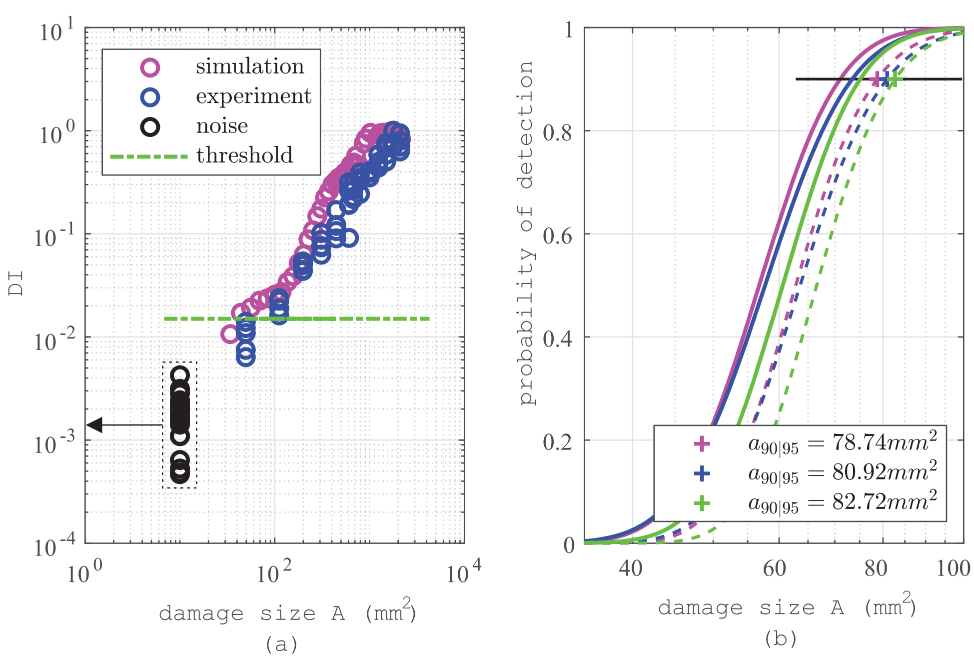

Figure 2. (a) Comparison of DI vs. damage size A for simulation and experimental data sets at a fixed threshold. (b) Comparative study of POD curves evaluated using experimental data set only, simulation data set only, and the combined concept. Data from path 3-9 and damage D1.

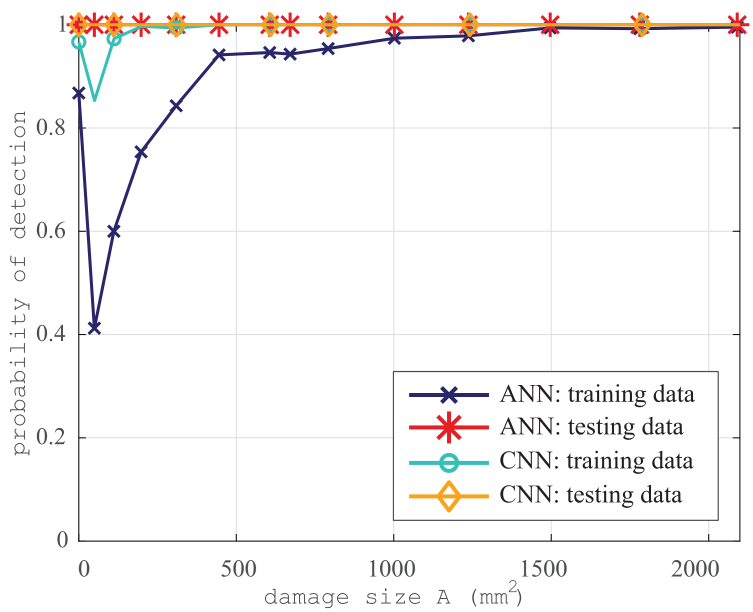

Figure 3. Probability of correct classification for CNN and ANN using data from all paths with damage D2 and D3 (training) and generalizing to damage D1 for path 3-9 only (testing).

results of the POD analysis purely based on the simulation data of the damaged state. The green lines show the results of the new MAPOD concept, combining simulated damage scenarios with experimental baseline measurements. The results show that the POD curve of the experimental data set, along with its confidence bounds, lies between the POD analysis of the simulation and the combined data set. Combining simulated damage scenarios with experimental baseline measurements results in a shift of the POD curve to the right, an increase in the standard deviation of the parameter of POD, and an increase in the confidence bounds. In other words, adding the experimental features to the simulated data set leads to a more conservative way of estimating the POD, with a larger range of uncertainty. By using simulations for the damaged states, it is no longer necessary to damage the real structure for estimating POD without loosing accuracy or being less conservative.

The ANN is evaluated systematically to compute the probability of correct classification for the undamaged samples (to get the false alarm rate) and for each damage size. In Figure 3, the probability of correct classification is presented for the training data set (D2 and D3 along all paths except 3-9) and for the generalization test data set (D1 at path 3-9). The values for a damage size of zero correspond to the probability of correct detection of undamaged samples, i.e. 1.0 minus the false alarm rate, which is 0.13 for the training data set, where the overall probability of correct classification is 87%. Using linear interpolation between the probability of correct classification for the different damage sizes, the damage size of 388 mm^2 with a POD of 0.9 can be evaluated in the training data set. It should be noted, that the results can be improved by using multiple damage indicators as inputs of the ANN and by filtering the training data, e.g., according to its distance to the damage [12]. The trained ANN has a high generalization performance, as the generalization testing data set leads to a $100\ \%$ correct detection for each damage size and a $0\ \%$ false alarm rate. The DI threshold of the trained ANN is back computed to 0.0085.

The same procedure is used to evaluate systematically the CNN and compute the probability of correct classification for undamaged and damaged states. In this case, the overall training and validation accuracy is $97\ \%$ while the false alarm rate is less than $4\ \%$. However, it is again possible to state that the trained CNN has a high generalization performance, as the testing data set also leads to a 100 % correct detection for each damage size and a $0\ \%$ false alarm rate. Having a better look at the probability of correct classification versus damage size for the training dataset, even though the accuracy is slightly lower (when all paths are considered, a few channels have little information content, which reduces the accuracy), the probability of correct classification is always higher than $85\ \%$ and the damage size of 74.3 mm^2 with a POD of 0.9 can be evaluated in the training data set.

Since in both cases, extrapolation for damage size smaller than damage $\#1$ cannot take place, it is worth highlighting that it is possible only to state that the equivalent metrics of $a_{90|95}$ for AI predictions using testing dataset is smaller than 50 mm^2.

CONCLUDING REMARKS

This paper presents different concepts for performance analysis of guided wave-based SHM systems using either real data or a combination of real and synthetic data. The former dataset is from the *Open Guided Waves* platform, while the latter is obtained by injecting EFIT simulation results. First, the comparison of the classic POD approach exploiting the first dataset and the MAPOD concept relying on the second dataset shows that the last one is slightly more conservative in assessing the minimum detectable size, which is generally beneficial for real-world applications, making it an efficient alternative for assessing the reliability assessment of SHM systems, without the need to apply variable damage scenarios on a real application. Second, AI based concepts showed the feasibility of reliability prediction using the probability of correct classification for either pristine condition (probability of false alarm) or damaged case (probability of detection). In particular, testing both ANN and CNN along the same path used for the classic approach returns very high reliability with the possibility of correctly identifying the minimum size of the damage. However, it is not possible to extrapolate the minimum detectable size, as there is not any model for POD like in the classic concept.

The collaboration of the authors was supported by the DFG within the Scientific Network *Towards a holistic quality assessment for guided wave-based SHM* (project number 424954879).

REFERENCES

1. Cawley, P. 2018. “Structural health monitoring: Closing the gap between research and industrial deployment,” *Structural Health Monitoring*, 17(5):1225–1244.
2. Falcetelli, F., N. Yue, R. Di Sante, and D. Zarouchas. 2022. “Probability of detection, localization, and sizing: The evolution of reliability metrics in Structural Health Monitoring,” *Structural Health Monitoring*, 21(6):2990–3017.

3. Meeker, W. Q., D. Roach, and S. S. Kessler. 2019. “Statistical Methods for Probability od Detection in Structural Health Monitoring,” in *Proceedings of 2019 IWSHM*.
4. Tschöke, K., I. Mueller, V. Memmolo, and et al. 2021. “Feasibility of model-assisted probability of detection principles for structural health monitoring systems based on guided waves for fiber-reinforced composites,” *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, 68(10):3156–3173.
5. Calmon, P., O. Mesnil, R. Miorelli, X. Artusi, B. Chapuis, and O. D’almeida. 2019. “Model assisted probability of detection for guided wave imaging structural health monitoring,” in *Structural Health Monitoring 2019*, DEStech Publications, Inc.
6. Moll, J., C. Kexel, J. Kathol, C.-P. Fritzen, M. Moix-Bonet, C. Willberg, M. Rennoch, M. Koerdt, and A. Herrmann. 2020. “Guided Waves for Damage Detection in Complex Composite Structures: The Influence of Omega Stringer and Different Reference Damage Size,” *Applied Sciences*, 10(9):3068.
7. Moll, J., J. Kathol, C.-P. Fritzen, M. Moix-Bonet, M. Rennoch, M. Koerdt, A. S. Herrmann, M. G. Sause, and M. Bach. 2019. “Open Guided Waves: online platform for ultrasonic guided wave measurements,” *Structural Health Monitoring*, 18(5-6):1903–1914.
8. Schubert, F. 2004. “Numerical time-domain modeling of linear and nonlinear ultrasonic wave propagation using finite integration technique - theory and applications,” *Ultrasonics*, 42:221–229.
9. Tschöke, K. and H. Gravenkamp. 2018. “On the numerical convergence and performance of different spatial discretization techniques for transient elastodynamic wave propagation problems,” *Wave Motion*, 82:62–85.
10. Berens, A. P. 1988. “NDE Reliability Data Analysis,” in *ASM Metals Handbook*, ASM Intermational, Materials Park, vol. 17, 9 edn., pp. 689–701, ohio.
11. Annis, C., E. Bray, H. Hardy, and P. Hoppe. 2009. “Nondestructive evaluation system reliability assessment,” *United States Department of Defense, MIL-HDBK-1823A*.
12. Mueller, I., S. Freitag, V. Memmolo, and et al. 2023. “Performance Assessment for Artificial Intelligence-Based Data Analysis in Ultrasonic Guided Wave-Based Inspection: A Comparison to Classic Path-Based Probability of Detection,” in *European Workshop on Structural Health Monitoring*, Springer International Publishing, Cham, pp. 953–961.
13. Géron, A. 2019. *Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow*, O’Reilly Media, ISBN 9781492032649.
14. Schackmann, O., V. Memmolo, and J. Moll. 2023. “Regression based continuous damage size estimation in complex composite structures using guided ultrasonic waves through a unified CNN approach,” *Ultrasonics*, In review.
15. Aldrin, J. C., E. A. Medina, E. A. Lindgren, C. F. Buynak, and J. S. Knopp. 2011. “Protocol for reliability assessment of structural health monitoring systems incorporating model-assisted probability of detection (MAPOD) approach,” Tech. rep.
16. Meyer, R. M., S. L. Crawford, J. P. Lareau, and M. T. Anderson. 2014. “Review of literature for model assisted probability of detection,” .
17. Stepinski, T., T. Uhl, and W. Staszewski. 2013. “Advanced structural damage detection: from theory to engineering applications,” .
18. Chapuis, B., P. Calmon, F. Jenson, et al. 2018. *Best practices for the use of simulation in POD curves estimation*, Springer.
19. Bayoumi, A., T. Vogt, and I. Mueller. 2023. “New approach for reliability assessment of guided wave-based structure health monitoring system on a pipe application,” in *8th International Workshop on Reliability of NDT/NDE*, SPIE, vol. 12491, pp. 41–51.