Probabilistic Damage Quantification via the Integration of Non-parametric Time-Series and Gaussian Process Regression Models

AHMAD AMER and FOTIS KOPSAFTOPOULOS

ABSTRACT

One of the major challenges the structural health monitoring (SHM) community is facing today is related to enabling accurate and robust damage detection and quantification under varying operating and environmental conditions with a limited number of sensing data sets. In particular, state-of-the-art active-sensing SHM technologies face significant difficulties in accurately detecting and quantifying damage within uncertain environments due to the lack of appropriate statistical learning and inference methods. As such, it is critical for the community to have available tools that accurately detect and quantify damage, in a probabilistic sense, using the –oftentimes limited– available information. In this study, a novel probabilistic damage detection and quantification approach is proposed based on the integration of non-parametric statistical time series representations and Gaussian Process Regression Models (GPRMs). Initially, non-parametric models based on Short-Time Fourier Transform (STFT) power spectral density (PSD) estimates are used in order to detect damage and statistically determine signal (wave) paths that carry the most information about damage size. Next, GPRMs are trained using Damage Index (DI) values from selected wave propagation paths and used to estimate the damage size based on the current DI sets. The experimental assessment is presented for data recorded from a notched aluminum plate. It is shown that, using wave propagation paths selected by non-parametric statistical models, estimation accuracy increases and damage size estimation confidence levels become narrower, thus providing a robust and efficient damage detection and quantification approach.

INTRODUCTION

With the advent of new sensing and data analysis methods, active-sensing Structural Health Monitoring (SHM) systems have shown potential in detecting, localizing and quantifying damage. However, due to their increased sensitivity to operating (boundary conditions, loads, etc.) and environmental (temperature, humidity, etc.) conditions, active-sensing SHM methods still face significant challenges when it comes to

Ahmad Amer, PhD Student, Email: amera2@rpi.edu. Fotis Kopsaftopoulos, Assistant Professor, Email: kopsaf@rpi.edu. Intelligent Structural Systems Lab (ISSL), Department of Mechanical, Aerospace and Nuclear Engineering, Rensselaer Polytechnic Institute, Troy, NY, USA

in the Proceedings of the 12th International Workshop on Structural Health Monitoring
September 2017, Stanford, CA, USA

their operation in dynamic environments and multiple operating conditions under uncertainty [1–3]. In addition, even in the case of constant or controlled environments and "hotspot" monitoring, such methods face difficulties in quantifying damage due to the uncertainty in the damage propagation patterns and the variation in the locations of the sensors within nominally identical structural components [1]. When it comes to damage quantification, the currently-employed approaches are of deterministic nature, i.e. they do not allow for the extraction of appropriate confidence intervals for damage size estimation, and require the collection and analysis of an immense amount of baseline data from a population of identical structures to account for changing conditions. Thus, active-sensing SHM damage detection and quantification methods are still lacking the required confidence for large-scale applicability and approaches that are both accurate and robust need to be developed, all without increasing data footprint.

Statistical time series methods have been widely used for vibration-based damage detection, localization, and quantification within a probabilistic framework [4–6]. However, their use in active-sensing guided-wave SHM has been limited. In a recent study [2], the authors have proposed the use of non-parametric time series models and corresponding statistical hypotheses tests for probabilistic damage detection via the use of Power Spectral Density (PSD) estimators of wave propagation signals. Although such non-parametric models have the advantage of simplicity, their potential to effectively tackle the more elaborate task of damage quantification is limited. Therefore, advanced statistical learning and inference methods are necessary to address the probabilistic quantification of damage in the face of uncertainty. In this context, stochastic modeling via Gaussian Process Regression Models (GPRMs) has been used for regression and classification applications within the machine learning community [7, 8], as well as for the identification of structural dynamics under varying operating conditions [9] and surrogate modeling of computationally-expensive high-fidelity models [10, 11].

In this study, a novel active-sensing probabilistic damage detection and quantification method is proposed based on the integration of non-parametric time series models with stochastic GPRMs. The fundamental premise of the method is that the incorporation of optimized actuator-sensor path information, in the form of properly-selected damage indices (DIs), in the GPRM-based statistical learning for damage quantification can provide accurate and robust damage size estimation with confidence intervals. First, non-parametric Short-Time Fourier Transform (STFT) models are applied in order to identify the actuator-sensor wave propagation paths that intersect damage, thus leading to the selection of the most damage-sensitive signal paths in order to optimize the amount of information being fed into the GPRM-based statistical learning phase. Next, after the GPRM training phase has been completed via the use of DIs obtained from the paths selected in the first step under different damage sizes, GPRMs are used with fresh DI values in order to estimate their corresponding damage size along with statistical confidence intervals, thus fulfilling the damage quantification part of this study. The use of GPRMs allows for the accurate prediction of damage size even when a *single* sensor-actuator path is used for training. Furthermore, multivariate GPRMs, that utilize the appropriate number of paths as determined in the first step, are shown to exhibit narrower damage confidence intervals when compared to their univariate and multivariate with randomly selected paths, counterparts. To the authors best of knowledge, this is the first time GPRMs are utilized for the quantification of damage in active-sensing SHM.

GAUSSIAN PROCESS REGRESSION MODELS

The problem of data-based statistical learning for inference and prediction can be tackled *via* stochastic GPRMs that can be used for both regression and classification and take the form of a full predictive distribution [7, 9]. In the present work, GPRMs are employed to "learn" the relationship between damage indices and damage size within a Bayesian probabilistic framework. In particular, the model output $g(\boldsymbol{x})$, i.e. damage size, is represented via a functional relationship of latent variables $f(\boldsymbol{x})$ from a Gaussian process of the input(s) (scalar or vector; also referred to as covariates or predictors) $\boldsymbol{x}$, i.e. the DI values of actuator-sensor signal propagation paths, based on a Bayesian linear regression model [7]:

$$g(\boldsymbol{x}) = f(\boldsymbol{x}) + \boldsymbol{h}(\boldsymbol{x})^T \boldsymbol{\beta}, \quad \text{where} \quad f(\boldsymbol{x}) \sim \mathcal{GP}(0, k(\boldsymbol{x}, \boldsymbol{x}')). \tag{1}$$

In these equations, $k(\boldsymbol{x}, \boldsymbol{x}')$ designates the covariance function (kernel) of the zero-mean Gaussian process $f(\boldsymbol{x})$. The D-dimensional input vector $\boldsymbol{x}$ is transformed into a p-dimensional feature space using the set of basis functions $\boldsymbol{h}(\boldsymbol{x})$, that along with the coefficients (parameters) vector $\boldsymbol{\beta}$, to be inferred from the data, represent the mean of the model. Incorporating explicit basis functions in the GPRM allows for the complete specification of the Gaussian process mean, thus enabling interpretability of the model and convenience of expressing prior information.

Fitting such a GPRM includes the joint optimization over the parameters $\boldsymbol{\beta}$ with the kernel hyperparameters, which are free parameters that the covariance function depends upon, and are specific to the type of function being used [7]. Thus, given a training data set $\{(\boldsymbol{x}_i, y_i),\ i = 1, 2, \ldots, n\}$, the output (herein the damage size) can be modeled via the vector form of the probabilistic GPRM:

$$P(\boldsymbol{y}|\boldsymbol{f}, \boldsymbol{X}) \sim N(\boldsymbol{y}|\boldsymbol{f} + \boldsymbol{H}\boldsymbol{\beta}, \sigma^2 \boldsymbol{I}) \tag{2}$$

such that

$$\boldsymbol{y} = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix}, \quad \boldsymbol{X} = \begin{pmatrix} \boldsymbol{x}_1^T \\ \boldsymbol{x}_2^T \\ \vdots \\ \boldsymbol{x}_n^T \end{pmatrix}, \quad \boldsymbol{f} = \begin{pmatrix} f(\boldsymbol{x}_1) \\ f(\boldsymbol{x}_2) \\ \vdots \\ f(\boldsymbol{x}_n) \end{pmatrix}, \quad \boldsymbol{H} = \begin{pmatrix} \boldsymbol{h}(\boldsymbol{x}_1^T) \\ \boldsymbol{h}(\boldsymbol{x}_2^T) \\ \vdots \\ \boldsymbol{h}(\boldsymbol{x}_n^T) \end{pmatrix}. \tag{3}$$

Here, it is assumed that the observed values $\boldsymbol{y}$ differ from the function values $g(\boldsymbol{x})$ by additive zero-mean Gaussian noise with variance σ^2. Essentially, the use of GPRMs instead of regular linear models, entails a joint Gaussian distribution upon the regression parameters, thus providing confidence bounds for the prediction/estimation process [7, 9]. GPRMs also involve the transformation of a nonlinear problem in the original input space into a linear one in a high-dimensional feature space using the "kernel trick" [7, 12]. This allows for capturing uncertainties that may stem from operational and environmental variability [9].

In this study, the input (predictor) matrix ($\boldsymbol{X}$) comprises of DI values obtained from both single and multiple actuator-sensor wave propagation paths (n is the total number of observations and D is the number of actuator-sensor paths included in the model) and is used to train univariate and multivariate, respectively, GPRMs for an aluminum notched plate, which are then utilized to estimate the notch size y.

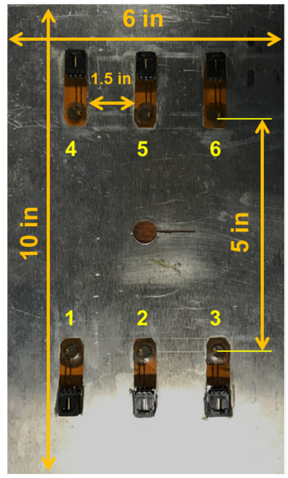

Figure 1. The notched Al plate experimental coupon used in this study.

THE EXPERIMENTAL SETUP

A 152.4×254 mm (6×10 in) 6061 Aluminum coupon (2.36 mm/0.093 in thick) with a 12.7 mm (0.5 in)-diameter hole in the middle was used in this study. Using Hysol EA 9394 adhesive, 6 lead zirconate titanate (PZT) piezoelectric sensors (type PZT-5A from Acellent Technologies, Inc), having a diameter of 3.175 mm ($1/8$ in) and a thickness of 0.2 mm (0.0079 in), were attached to the Al plate as shown in Figure 1. A 2 mm (0.079 in) notch was cut extending from the hole of the Al plate using an end mill, then elongated in 2-mm increments up to 20 mm (0.79 in), for a total of 10 damage cases, using a 0.81 mm (0.032 in) handsaw.

A series of 5-peak tone-burst actuation signals with 90 V amplitude peak-to-peak and various center frequencies were generated in a pitch-catch configuration. A total of 20 data sets per structural case (11 cases in total; one healthy and 10 damage cases) were collected at a sampling rate of 24 MHz using a ScanGenie III data acquisition system (Acellent Technologies, Inc). The data were then exported to MATLAB for analysis[1].

RESULTS AND DISCUSSION

In the context of active-sensing using guided waves, there are often multiple sensors installed at the area being monitored, and every actuator-sensor path in the network has to be examined in order to assess the integrity of the component, which may compromise the accuracy and robustness of the analysis process. Figure 2 panels a and b show the signal received at sensor 6 when sensor 2 was actuated (see Figure 1). As shown, because this is a damage-intersecting path, a significant change can be observed in the signals as notch size increases. Exploring actuator-sensor paths that do not intersect damage, such as path 1-5 (Figure 2 panels c and d) one can observe that the received signals at sensor 5 exhibit a significantly smaller change in amplitude with notch size. Thus, information from damage-non-intersecting actuator-sensor paths, such as path 1-5, naturally carry

[1]Matlab version R2018a.

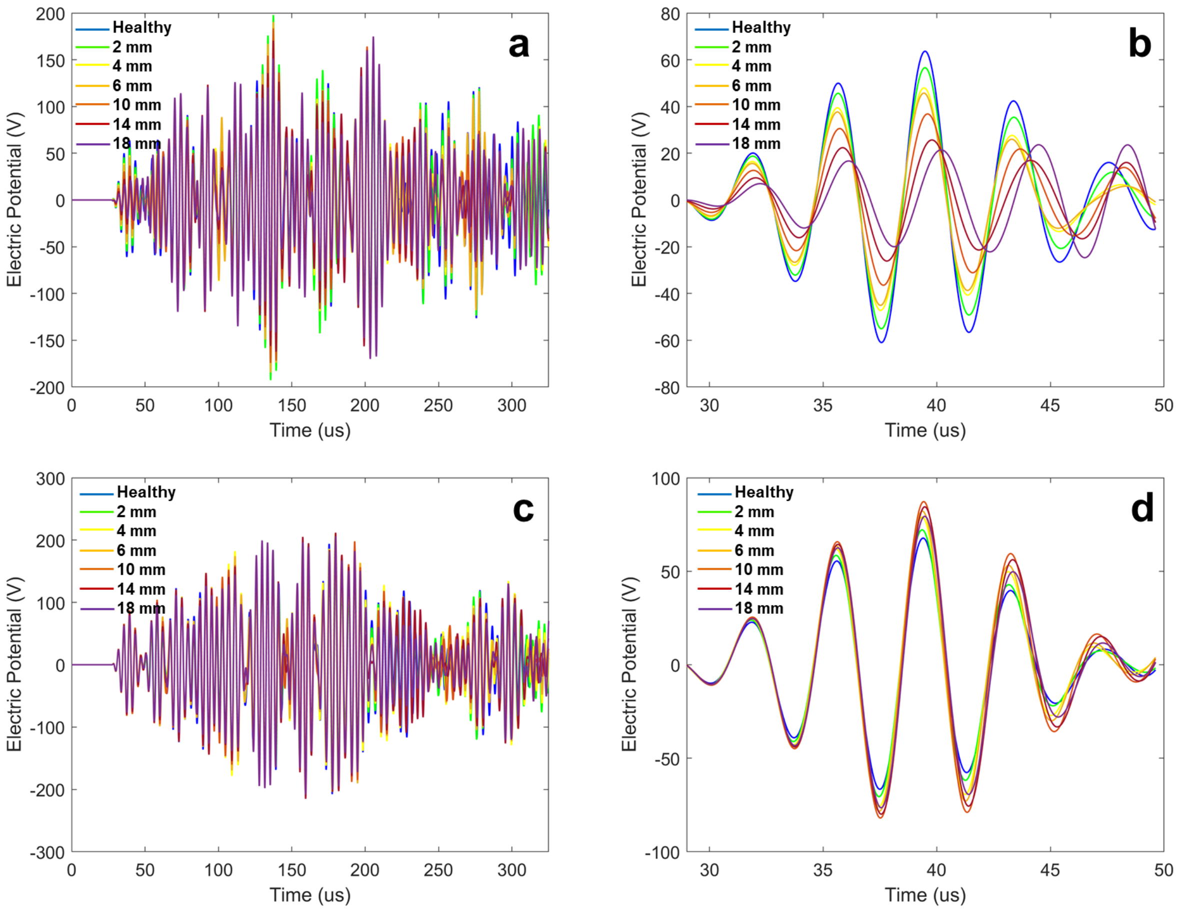

Figure 2. Indicative signals for the notched Al plate. (a) Full signal received at sensor 6 when sensor 2 is actuated, and (b) its first-arrival wave packet. (c) Full signal received at sensor 5 when sensor 1 is actuated, and (d) its first-arrival wave packet.

less information when it comes to damage quantification compared to paths that intersect damage. In order to further support this point, one state-of-the-art DI from the literature [1], was explored to see how damage intersection affects damage quantification using the DI approach. Figure 3 panels a and b show the evolution of the selected DI with notch size for indicative damage-intersecting (path 2-6) and damage-non-intersecting (path 1-5) paths (see Figure 2), respectively. From Figure 3, it becomes apparent that the performance of damage quantification models may be significantly affected by the type of information and corresponding data employed to train them.

In this sense, a statistical method was recently developed to assess which actuator-sensor paths intersect damage and which do not, as a preliminary filter for damage quantification models, where only data sets carrying the most important information are processed [2]. In this method, an STFT-based non-parametric model is used and damage-intersecting paths are selected according to the PSD change of the path due to damage. Briefly, damaged areas that provide a reduction in guided-wave impedance (reduce wave energy attenuation that would otherwise occur in a healthy component) would allow more signal energy to pass, while less energy would pass if the damaged area causes the increase in wave impedance. Based on these observations, the developed

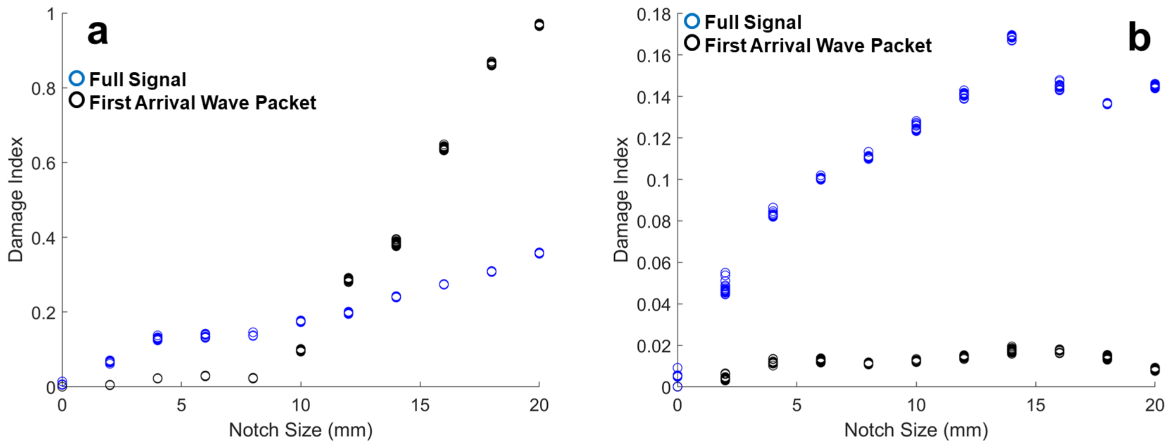

Figure 3. The evolution of the damage index [1] as applied to indicative actuator-sensor paths: (a) damage-intersecting path 2-6 and (b) damage-non-intersecting path 1-5.

method can also identify damage-non-intersecting paths. Figure 4 shows indicative results of applying such a model[2] onto the first arrival wave packets of the signals shown in Figure 2. As can be seen in Figure 4a, because the notch is a damage of the second type mentioned above, a decrease in the median/mean STFT-based PSD value with increasing notch size can be observed for path 2-6. On the other hand, the increase in the PSD with notch size for the damage-non-intersecting path (see Figure 4b) can be explained in terms of the energy being scattered off the notch in damage-intersecting paths when sensor 1 is actuated; those scattered waves of energy go to sensors that are not directly on the path of the notch. Thus, using this approach, it is possible to identify damage-intersecting actuator-sensor signal propagation paths that are expected to contain the most valuable information in terms of damage sensitivity. This feature can be utilized directly in increasing the accuracy and robustness of damage localization and quantification algorithms. In this study, it is applied to improve the performance of GPRMs for damage quantification.

Damage Quantification via Univariate GPRMs

For the probabilistic damage quantification task, GPRMs[3] for indicative damage-intersecting and non-intersecting paths were trained using 15 DI values for every structural case (healthy case and notch sizes from 0 mm to 20 mm), thus 165 observations were used to train the model (11 structural cases $\times$ 15 data sets). Next, 5 DI values per case were used to estimate the notch size (55 observations) and assess the method. All models presented in this study utilized linear basis functions and exponential kernel functions in the estimation process. GPRMs were first implemented for one actuator-sensor path at a time, i.e. GPRMs based on a single input variable which is the DI set of that specific path, in order to assess their ability to quantify damage using a single path. Figure 5 shows the results of applying univariate GPRMs onto the indicative

[2]Matlab function *spectrogram.m* (window size: 400-900; nfft: 960; noverlap: 95%)

[3]Matlab functions *fitrgp.m* (n: 11 $\times$ 15 data sets for training; basis function: linear; kernel function: exponential) and *predict.m* (n: 11 $\times$ 5 data sets for validation/estimation)

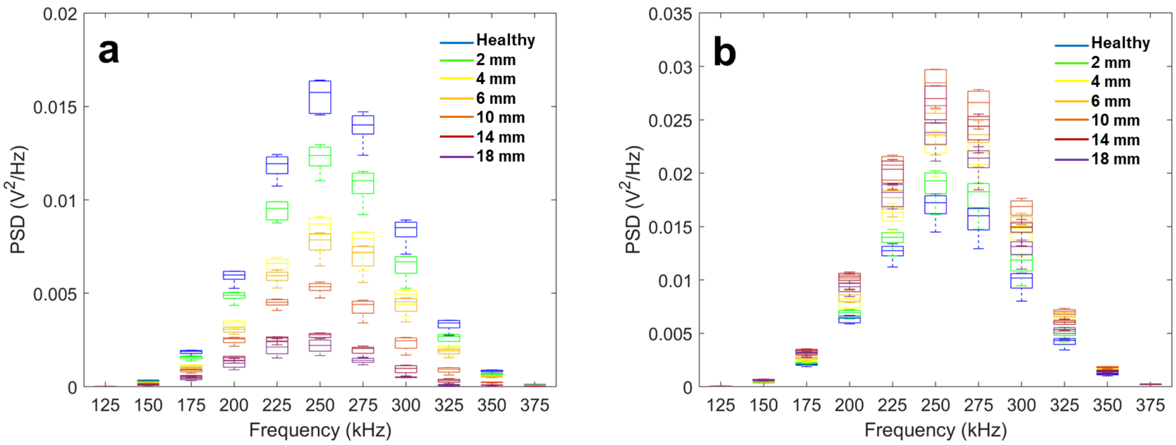

Figure 4. Indicative results of applying the non-parametric model onto actuator-sensor paths. Short Time Fourier Transform (STFT) plots for the signals of (a) damage-intersecting path 2-6 and (b) damage-non-intersecting path 1-5.

damage-intersecting (panel a) and non-intersecting (panel b) paths shown in Figure 2. The estimated damage sizes, using 5 calculated DI sets that were not used in model training, are presented as red stars; the actual damage sizes are depicted in blue circles. An important conclusion from this figure is that the use of a damage-intersecting path in GPRM significantly improves the damage size estimation accuracy compared to using signals from actuator-sensor paths that do not directly intersect damage. Thus, using the non-parametric model described in the previous section, damage-intersecting paths can be selected and fed into both univariate and multivariate GPRMs in order to get an accurate estimation of damage size.

Damage Quantification via Multivariate GPRMs

In this section we investigate the performance of multivariate GPRMs, i.e. using DIs obtained from several wave propagation paths as GPRM inputs, in terms of probabilistic damage quantification. Indicative GPRMs were trained using 15 data sets, but this time incorporating DIs from multiple paths, hence multivariate models, where each DI represents an input variable. The aforementioned non-parametric model used for identifying damage-intersecting paths was coupled with multivariate GPRMs in order to assess the performance of quantification when using paths that contain the most accurate information about damage size. Figure 6 panels a and b show indicative damage estimation results based on GPRMs trained *via* two sets of three randomly-selected paths, where the real notch size, all 5 estimated notch sizes (from 5 DI sets per case), and 95% confidence intervals for each individual structural case are shown. Figure 6c shows the same plot for three paths that were marked as damage-intersecting by the STFT model. An apparent difference between the latter panel and former two is the significantly improved accuracy of the notch size estimation. Hence, prior selection of damage-intersecting paths allows for both improved accuracy and narrower confidence intervals.

In order to explore the quantification process using multiple paths, selected both blindly and based on the STFT model, the standard deviation of the estimated notch

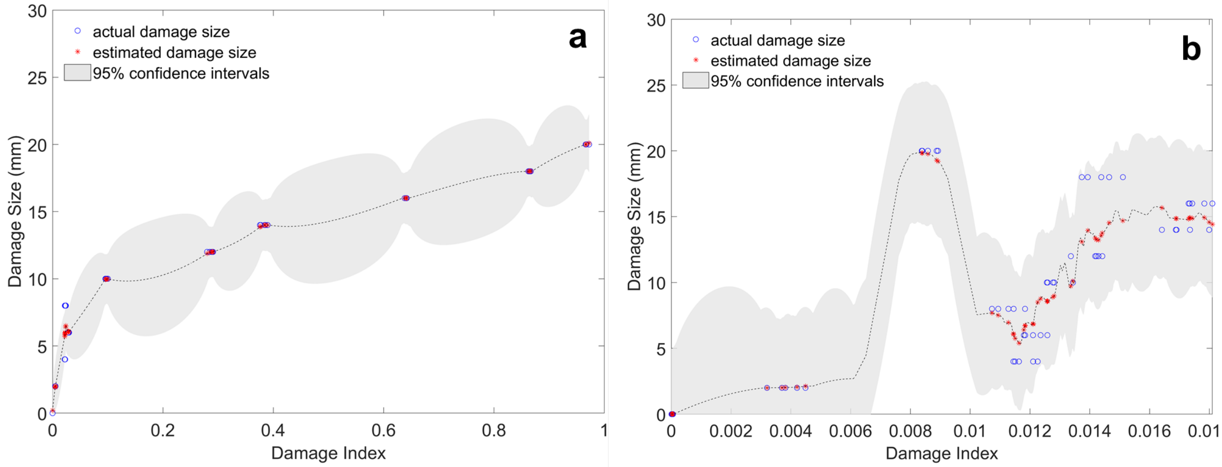

Figure 5. Indicative univariate GPRM estimation plots trained using 5 DI sets of (a) damage-intersecting path 2-6, and (b) damage-non-intersecting path 1-5.

and the mean estimation errors were used as metrics for assessing the performance of multivariate GPRMs. Figure 7a presents the mean of the standard deviation of the estimated damage sizes as more paths are included in the training phase. The blue bars are for wave paths chosen blindly, i.e. with no regard to the nature of the path in terms of damage intersection. Although there are 30 paths available, the standard deviation mean levels off after 14 paths are used to train the model, so up to 16 paths are shown. Red bars indicate paths marked as damage-intersecting by the non-parametric model, which were 12 in this case. As shown, using as few as 2 selected paths for training the model, the standard deviation mean decreases by more than 35% compared to the blind (random) path selection. As the number of training signal paths (DI data sets) increases, the models trained via paths selected by the non-parametric method converge to a lower mean standard deviation, even at 12 paths. Using more than 12 paths, the mean standard deviation decreases until it levels off at 14 paths. Thus, it can be concluded that the damage size estimation confidence levels become narrower when only selected paths are utilized in training the GPRM, thus leading to more accurate damage quantification.

Figure 7b shows the mean estimation error for all notch sizes versus the number of incorporated paths when paths, and therefore DI data sets, are chosen blindly or via the non-parametric model. Again, the estimation error when the model is trained based on selected paths is significantly smaller compared to the blind selection of paths with no regard to their nature. As expected, the blue and red estimation error percentages converge to similar values as the number of training paths increases.

CONCLUDING REMARKS

In this work, an active-sensing probabilistic damage quantification method based on the integration of non-parametric time-series and GPRMs was introduced. Non-parametric STFT models were used in a notched Al plate to pinpoint damage-intersecting paths that carry damage-size information. Both uni- and multivariate GPRMs were trained using single/multiple actuator-sensor paths from 15 data sets, and then used to estimate damage size for 5 data sets. It was shown that the use of damage-intersecting paths

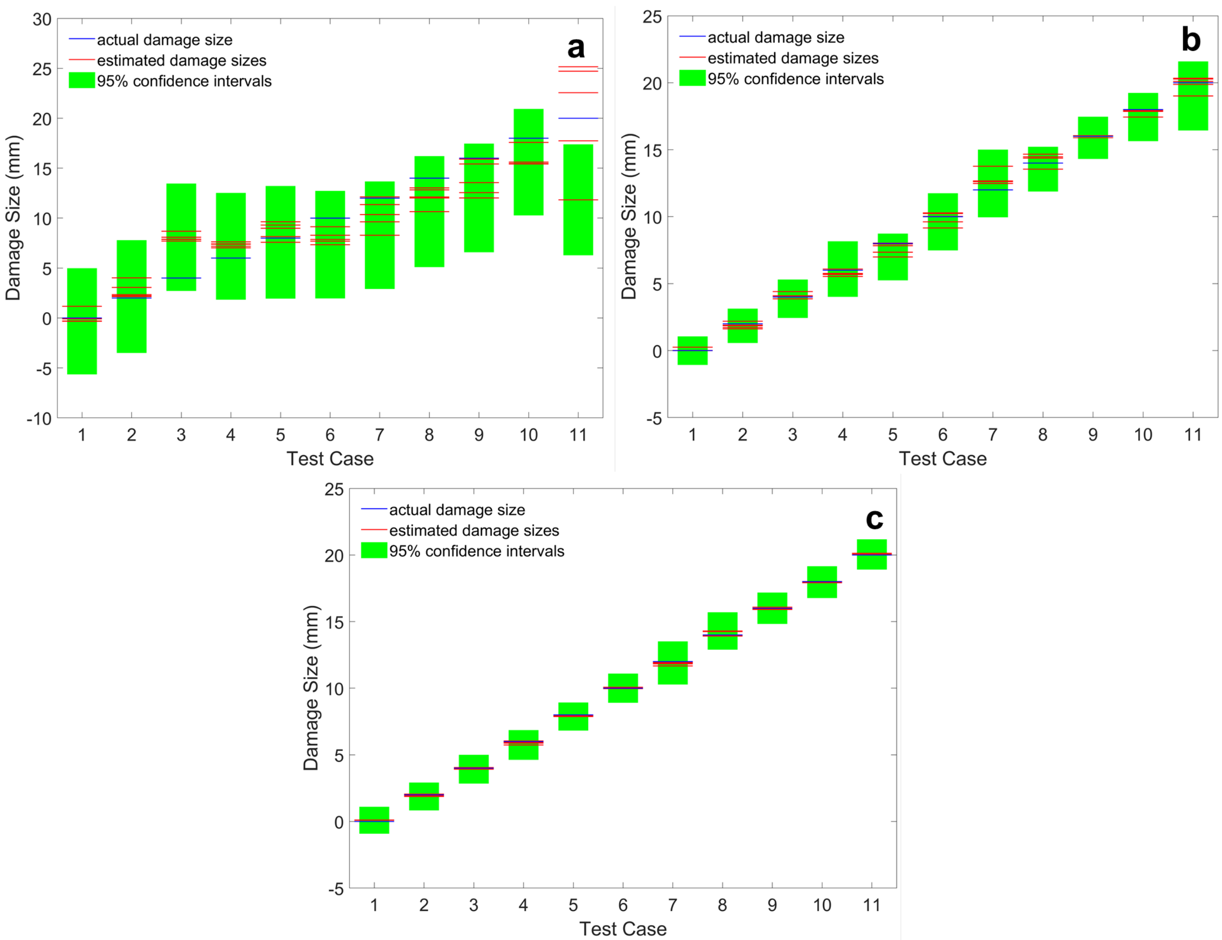

Figure 6. Indicative damage size estimation boxplots based on multivariate GPRMs: (a and b) 5 DIs used for two sets of three randomly selected paths; (c) 5 DIs used from three paths indicated by the non-parametric model.

for univariate GPRM training results in increased damage estimation accuracy. Furthermore, using paths marked by the non-parametric model as damage-intersecting to train multivariate GPRMs, it was observed that damage size estimation confidence intervals become significantly narrower compared to using paths blindly- (randomly-) selected. The reliability of the proposed damage detection and quantification method was assessed using the mean of the estimation standard deviation, as well as the mean estimation error. It was shown that using as few as two properly-selected actuator-sensor paths for training multivariate GPRMs, the estimation standard deviation mean was much lower than that for multivariate GPRMs trained using randomly-selected paths. The results of this study indicate the potential of coupling non-parametric time-series models with stochastic GPRMs for improved learning towards probabilistic damage quantification.

ACKNOWLEDGEMENTS

This work is carried out at the Rensselaer Polytechnic Institute under the Vertical Lift Research Center of Excellence (VLRCOE) Program, grant number W911W61120012, with Dr. Mahendra Bhagwat and Dr. William Lewis as Technical Monitors.

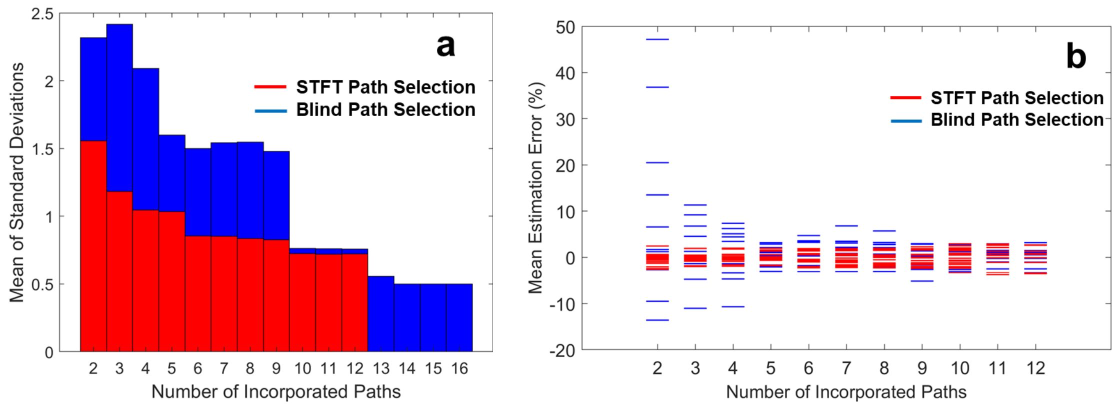

Figure 7. Reliability quantification for the multivariate GPRMs trained using selected and blind (random) paths. (a) Mean of estimation standard deviations for 2-16 paths. (b) Estimation error percentage for 2-12 paths.

REFERENCES

1. Janapati, V., F. Kopsaftopoulos, F. Li, S. Lee, and F.-K. Chang. 2016. "Damage detection sensitivity characterization of acousto-ultrasound-based structural health monitoring techniques," *Structural Health Monitoring*, 15(2):143–161.
2. Amer, A. and F. P. Kopsaftopoulos. 2019. "Probabilistic active sensing acousto-ultrasound SHM based on non-parametric stochastic representations," in *Proceedings of the Vertical Flight Society 75th Annual Forum & Technology Display*, Philadelphia, PA, USA.
3. Ahmed, S. and F. P. Kopsaftopoulos. 2019. "Uncertainty quantification of guided waves propagation for active sensing structural health monitoring," in *Proceedings of the Vertical Flight Society 75th Annual Forum & Technology Display*, Philadelphia, PA, USA.
4. Kopsaftopoulos, F. P. and S. D. Fassois. 2010. "Vibration based health monitoring for a lightweight truss structure: experimental assessment of several statistical time series methods," *Mechanical Systems and Signal Processing*, 24:1977–1997.
5. Kopsaftopoulos, F. P. and S. D. Fassois. 2013. "A Stochastic Functional Model Based Method for Vibration Based Damage Detection, Localization, and Magnitude Estimation," *Mechanical Systems and Signal Processing*, 39(1–2):143–161.
6. Das, S., P. Saha, and S. K. Patro. 2016. "Vibration-based damage detection techniques used for health monitoring of structures: a review," *Journal of Civil Structural Health Monitoring*, 6:477–507.
7. Rasmussen, C. E. and C. K. I. Williams, eds. 2006. *Gaussian Processes for Machine Learning*, MIT Press.
8. Seeger, M. 2004. "Gaussian processes for machine learning," *International Journal of Neural Systems*, 14:69–106.
9. Avendaño-Valencia, L. D., E. N. Chatzi, K. Y. Koo, and J. M. Brownjohn. 2017. "Gaussian process time-series models for structures under operational variability," *Frontiers in Built Environment*, 3:69.
10. Erdogan, Y. S., M. Gul, F. N. Catbas, and P. G. Bakir. 2014. "Investigation of uncertainty changes in model outputs for finite-element model updating using structural health monitoring data," *Journal of Structural Engineering*, 140:04014078.1–14.
11. Ling, Y. and S. Mahadevan. 2012. "Integration of structural health monitoring and fatigue damage prognosis," *Mechanical Systems and Signal Processing*, 28:89–104.
12. Rosipal, R. and L. J. Trejo. 2001. "Kernel Partial Least Squares Regression in Reproducing Kernel Hilbert Space," *Journal of Machine Learning Research*, 2:97–123.