<u>C</u>ovariance of <u>L</u>imit <u>D</u>efining <u>P</u>airs (CLDP): A Novel Approach to Establishing Detection Sensitivity for Structural Health Monitoring Data

Seth S. Kessler
Metis Design Corporation

Christine M. Schubert Kabban
Air Force Institute of Technology

IWSHM-2023

ABSTRACT

This paper explores a novel approach to establishing detection sensitivity for SHM. The proposed approach uses data associated with the largest flaw such that all smaller flaws were not detected and the smallest flaw such that all larger flaws were detected for that specimen. These points are termed "limit defining" because they directly contribute to defining the most valuable portion of the probability of detection (POD) curve. A POD curve can be efficiently built using a standard hit/miss analysis, and an upper confidence bound can be generated by calculating the Covariance of the Limit Defining Pair (CLDP) to take the dependence into account. A fatigue experiment was conducted on 60 aluminum bars using a carbon nanotube (CNT) based fatigue crack gauge. 100 data points were collected for each specimen as the crack grew to ~2mm. 1000 classic POD curves were generated by randomly selecting 1 point from each specimen to observe variability, and subsequently the same data was analyzed using the CLDP. Finally, random subsets of specimens were selected to simulate POD statistics associated with testing fewer specimens. Overall, CPLD has proven to be a robust and reliable alternative to traditional detection sensitivity evaluation for SHM sensors.

INTRODUCTION

MIL-HDBK-1823A has been used for the past two decades to assess, compare and specify detection sensitivity for non-destructive inspection (NDI) equipment. The key metric is $a_{90/95}$—the flaw that can be detected 90% of the time with 95% confidence—however the entire probability of detection (POD) versus flaw size curve is typically used when calculating overall acceptable system risk. MH1823 provides guidance and example procedures for producing this POD curve, with references to statistical software that can facilitate the calculations. A key distinction for the methods described in MH1823 is that complete measurement independence is explicitly presumed, i.e. it is not permissible to grow a flaw on the same test article that is measured more than once, or take multiple measurements of the same flaw while changing other variables. As currently described in the handbook, only one measurement can be taken per test specimen, as additional measurements would then have some degree of dependence.

Consistent with NDI techniques, POD must be considered when determining the detection capability and monitoring intervals for Structural Health Monitoring (SHM) methods using processes aligned with the statistical methods described in MH1823A. Current detection sensitivity assessment approaches laid out in MH1823A are challenging to implement for SHM however. The complication arises from the presumption of measurement independence. These methods were developed for legacy NDI equipment where inspectors brought reusable hardware and sensor probes to a set of pre-damaged specimens and collected the necessary data. However, by definition SHM sensors need to be permanently attached to the specimens (regardless of whether or not they are installed pre-damage to collect a baseline measurement or post-damage), so while it is possible to formulate SHM-POD in such a fashion with 60-100 specimens as suggested in the handbook, it is impractical to the extent that it would become prohibitively expensive. The present effort seeks to remedy this, by establishing a statistically equivalent probability of detection (SEPOD) approach for SHM-POD.

APPROACH

The data generated for SHM-POD analysis was collected using WISP Fatigue Crack Gauges (FCG). The WISP FCG is fabricated by sealing polymer nanocomposite (PNC) deposited onto patterned metallic electrodes between two thin flexible substrates. The entire FCG assembly is ~200 µm thick and weighs ~10 mg/cm^2. Rather than measuring the finite effects of braking traces in traditional crack gauges, there are literally trillions of carbon nanotubes (CNT) forming the electrical path, each contributing to either a parallel or series resistance. As a crack grows anywhere under the PNC, the measured resistance will change as a function of the crack length and orientation with respect to the electrodes. Technically, the name for this approach is Direct Current Potential Difference (DCPD). Resistance of the PNC network can be analytically solved as a function of crack length (Eq 1), and for small cracks (a << gauge length w) the equation can be simplified such that the crack length is approximately proportional to the square root of the normalized resistance change times a geometric constant related to the gauge form factor (Eq 2).

$$R = R_0 - R_S \frac{4}{\pi} \ln\left(\cos\left(\frac{\pi\, a}{2\ \mathrm{w}}\right)\right) \tag{1}$$

$$a = \mathrm{G}_f \sqrt{\Delta \bar{R}} \qquad \text{small } a/w \tag{2}$$

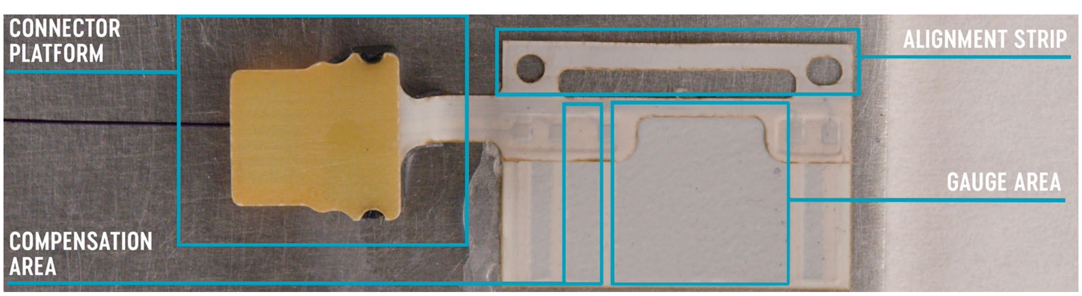

Figure 1: Photographs of the WISP Fatigue Crack Gauge (FCG)

EXPERIMENTAL SETUP

To generate flaws for developing an SHM-POD approach, 12” long by 1” wide 6061-T6 Aluminum bars were waterjet cut from cold rolled 1/8” sheet stock. Following ASTM E647 “Standard Test Method for Measurement of Fatigue Crack Growth Rates” a ~0.2” notch was EDM cut in the center of one edge of the bar to provide a reliable crack initiation point. WISP FCG sensors were bonded centered on the EDM notch. To grow natural cracks, specimens were fixtured in 4-point bending, and a fixed displacement was applied using a programmable stepper motor. First 75 cycles would be applied between 100με and 1000με followed by 5 cycles between 100με and 1500με. This pattern was repeated 9 times, and on the 10th repetition the first set of cycles was increased from 75 to 200 cycles. The effect of this loading routine was the creation of marker-band striations on the cracked surface that could be used for truth data after testing. WISP data was collected after each high-strain block. A total of 60 independent specimens were instrumented and tested. Before specimens were cut for marker-band analysis, the distance between the bonded gauge and the EDM notch was recorded using a micrometer. Overall, the average distance was 0.25mm, however there were specimens that were as far as 0.7mm away. Since the WISP FCG has no way of detecting a crack outside of the bond area, the data analysis was conducted using an “offset crack length”, i.e., just the length of the crack beneath the gauge.

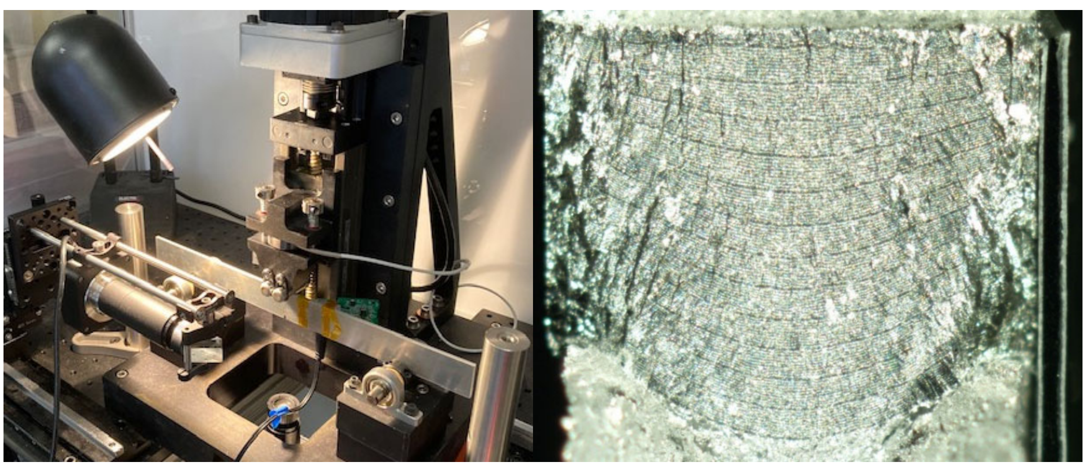

Figure 2: Photographs of 4-point fatigue bending fixture (left) marker-band surface (right)

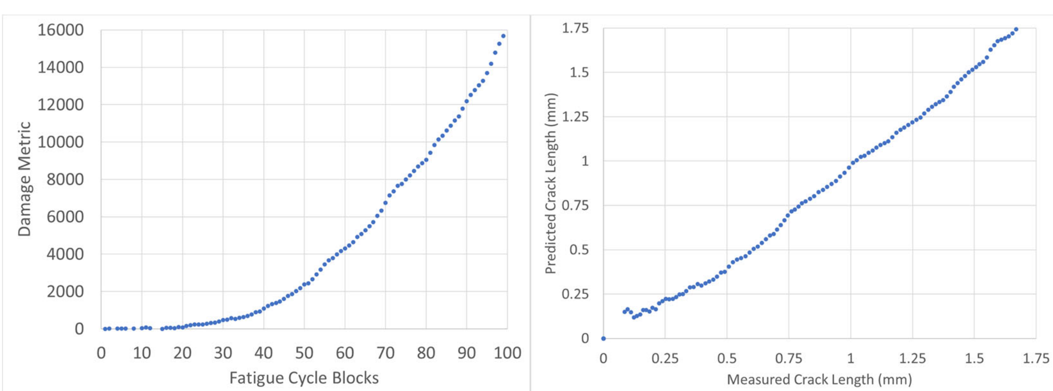

Figure 3: Example of raw WISP FCG data (left) and data converted to crack length (right)

STATISTICAL ANALYSES

Traditional Hit/Miss Analysis

For the hit/miss analysis, a threshold value of 125 was used to indicate a crack being detected (signal > 125 denotes crack detection) or no crack detected (signal < 125 denotes no crack detected). A "hit" was recorded if a physical crack was present and the SHM system recorded a signal above the threshold value, and a "miss" was recorded if a physical crack was present and the SHM system recorded a signal below the threshold value. A logistic regression model was fit to this data using an indicator for the "hit" and the "miss" as the response variable and the crack length as the independent predictor. From this curve the crack length that resulted in a 0.90 probability of "hit" detection was estimated (a_{90}). This fitted curve is called the POD curve. Confidence bounds for the POD curve were then computed to determine the 95% upper confidence bound for the detectable crack length with 0.90 probability ($a_{90|95}$). It is important to note in a traditional hit/miss analysis, observations are assumed to be independent. In data samples where multiple observations are observed on the same specimen, independence cannot be maintained if more than one observation from each specimen is included into the data used to generate the POD curve. Therefore, in order to conduct a traditional hit/miss analysis, code was written to randomly select a single observation from each specimen to create an independent dataset from which to build the POD curve. Next, a sensitivity analysis was conducted to investigate how much these POD statistics would vary depending on the "luck of the draw". The same random selection code was executed 1000 times in conjunction with the hit/miss analysis. Finally, it was investigated how much these POD statistics would vary depending on the number of specimens tested. The previous procedure was followed again, but this time for each of the 1000 iterations, the single datapoint was chosen from a subset of the 60 specimens selected at random. The subsets were generated repeatedly from 59 specimens down to a subset of only 10 specimens.

For each set of 1000 iterations the mean, standard deviation, minimum, and maximum value for the a_{90} and $a_{90/95}$ were reported as a function of specimen subset size. Two additional "diagnostic" values were also reported:

- % Convergent reports on the percentage of the 1000 iterations that resulted in convergent $a_{90/95}$ values. There are isolated cases where $a_{90/95}$ cannot be determined because the upper confidence bound for the a_{90} curve never intersects with a 90% probability or the bound itself is not able to be estimated, particularly when there the specimen subset size gets small.
- % BLED (Below Largest Experimental Datapoint) reports on the percentage of the 1000 iterations that resulted in $a_{90/95}$ values that were below the largest Offset Crack Length measured within the subset of specimens selected for that iteration. Beyond the non-convergent cases, some isolated cases converged but resulted in irrational $a_{90/95}$ values that in some cases were even larger than the specimen width. It was determined that a reasonable cut-off value for a rational $a_{90/95}$ value was no larger than an experimentally observed crack length, as otherwise the value was purely a mathematical extrapolation.

All non-convergent cases and those with values above the largest experimental datapoint were removed before calculating the $a_{90/95}$ mean and standard deviation.

Covariance of Limit Defining Pairs (CLDP)

For SHM however, the sensors are single-use, permanently installed elements that effectively become disposable once attached to the specimen, so a large test matrix can become prohibitively expensive. Thus, the motivation was for an alternative approach to SHM-POD to reduce the required quantity of test specimens, while taking advantage of the fact that once instrumented, continuous SHM data is essentially "free". Covariance of Limit Defining Pairs (CLDP) is a novel variation of the traditional hit/miss analysis, making use of accepted statistical practices already acknowledged in MH1823. The main principle of CLDP is making more efficient use of each specimen so that every datapoint is equally "valuable", through sparing use of repeated measures while accounting for its dependence. The CLDP analysis is conducted as follows:

a. Conduct a POD test matrix collecting as much data as practically possible during each test, assuring that at least 5 points are collected before and after the damage threshold has been crossed, whether through direct real-time observation, past experience or excessive data volume.
b. For each specimen, identify the "limit defining pair".
    i. Largest "miss" such that all prior data points are also misses.
    ii. Smallest "hit" such that all subsequent data points are also hits.
c. Perform a hit/miss analysis on the "limit defining pairs", using covariance to account for the dependence between these repeated measures.

Note that the "limit defining pairs" could be adjacent to each other if the damage metric always trends upward, or could be separated by points that oscillate around the threshold due to noise. Two version of CLDP have been implemented, 1B1A and 2B2A, in reference to how many points (1 or 2) are used before and after the detection threshold.

For the CLDP analysis presented here, the same procedure as the traditional hit/miss analysis was followed, where POD statistics were calculated for reduced specimen subset, with 1000 random combinations chosen for each subset size. The main difference in this case is that for each combination, the limit defining pairs remained constant as opposed to selecting a random datapoint from each specimen in the subset. Additionally, the same diagnostic % Convergent and % BLED values were presented.

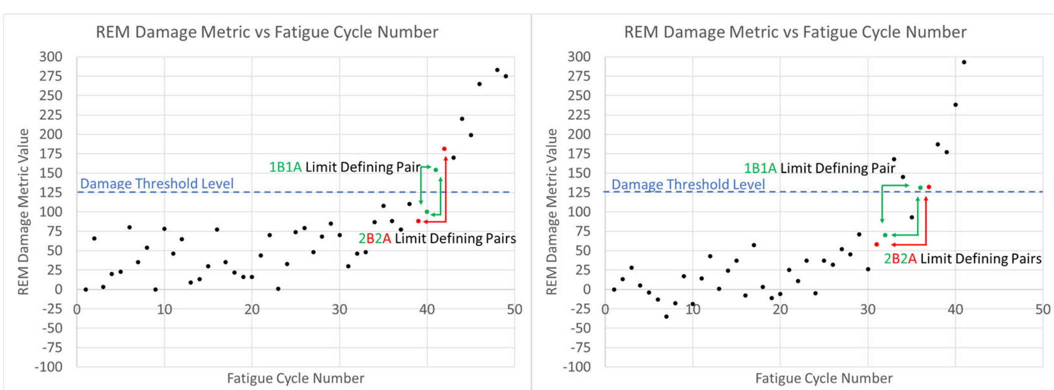

Figure 4: Illustrations of cases of Limit Defining Pairs selected for 1B1A and 2B2A analyses

Comparison of Results

Looking at the traditional hit/miss analysis, it is interesting to note that the a_{90} standard deviation is relatively high compared to the mean value (Figure 6). One could reasonably expect a ~20% variability in the a_{90} value just based on random draws of the data point used from each specimen. Under 40 specimens, the mean and standard deviation for $a_{90/95}$ starts to grow exponentially, and ultimately there are many non-convergent cases. Overall, it can be seen that even with 60 specimens, a simple hit/miss analysis may not yield repeatable results for this SHM-POD, and that it is quite unlikely that less than 40 specimens could be used to generate a meaningful POD curve. Looking at the CLDP hit/miss results, great improvements can be seen immediately. The a_{90} mean value is very consistent throughout the entire specimen subset range simulated. Similarly, the $a_{90/95}$ values are stable for the CLDP cases, the standard deviations are reasonable for the random draws of specimen subsets and there are many fewer non-convergent cases. Using 2 limit defining pairs (2B2A) yields meaningful results under 20 specimen subsets. Perhaps most importantly, the CLDP POD statistics are consistent with the mean traditional hit/miss analysis values for higher specimen quantities.

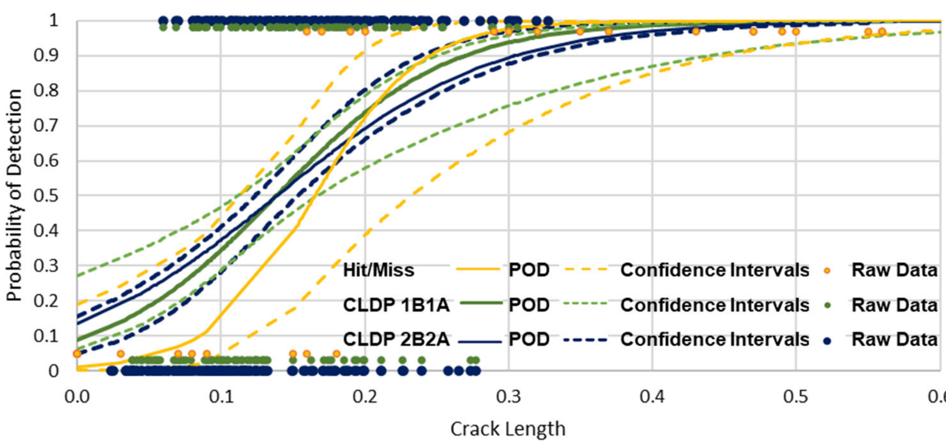

Figure 5: Overlayed example POD curves for traditional hit/miss, CLDP 1B1A and 2B2A

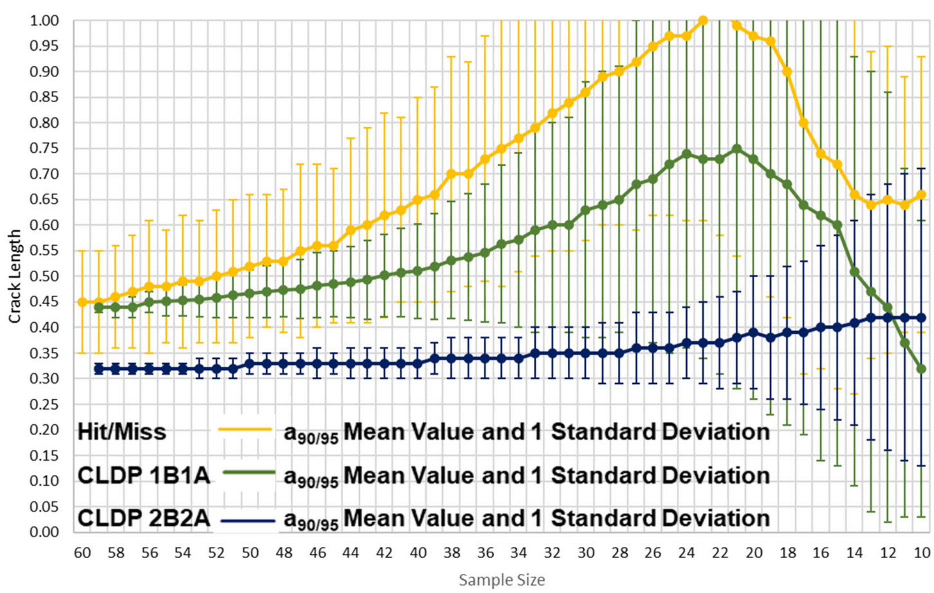

Figure 6: $a_{90/95}$ means w/standard deviation as a function of specimen subset size (all methods)

Recommendations for SHM Detection Sensitivity Approach

The main goal of this investigation was to develop a POD analysis approach that could allow for reducing the number of physical experiments performed while maintaining the integrity of the POD statistics for the purposes of calculating risk of structural failure. Now that it has been demonstrated that a CLDP hit/miss analysis has been validated to provide reliable POD statistics compared to a traditional hit/miss approach, a methodology must be presented for CLDP to achieve the goal of reducing test specimens. Just demonstrating that the mean $a_{90/95}$ value is stable down to fewer specimen subsets is not sufficient, because while it is more likely to produce a conservative $a_{90/95}$ value (because while the standard deviation is growing the mean $a_{90/95}$ is also increasing), it is also very possible to produce a non-conservative $a_{90/95}$ value. This would defeat the purpose of having 95% confidence intervals for calculating risk. The proposed CLDP methodology is as follows for evaluating a new SHM method:

a. Test a large set of specimens (60), collecting SHM and truth data as frequently as practical, at least 5 datapoints before and after the detection threshold.
b. Calculate the POD statistics according to the CLDP 2B2A hit/miss approach, using 1000 random draws to generate $a_{90/95}$ mean and standard deviation values for specimen subsets down to 10 specimens (minimum)
c. Generate a table of "penalty factors" to assure with high confidence the $a_{90/95}$ derived from fewer specimens is conservative, as follows:

(2σ $a_{90/95}$ of reduced dataset) + ($a_{90/95}$ of full dataset) - (mean $a_{90/95}$ of reduced dataset)

Future SHM-POD analysis would be conducted as follows:

a. Execute the selected number of experiments, perform CLDP 2B2A analysis.
b. If the $a_{90/95}$ value does not converge below the largest experimentally observed value, conduct additional experiments until $a_{90/95}$ converges.
c. Penalized upper confidence bound can be generated by calculating penalty factors for $a_{10/95}$, $a_{30/95}$, $a_{50/95}$, $a_{70/95}$, $a_{90/95}$ and fitting the POD curve.

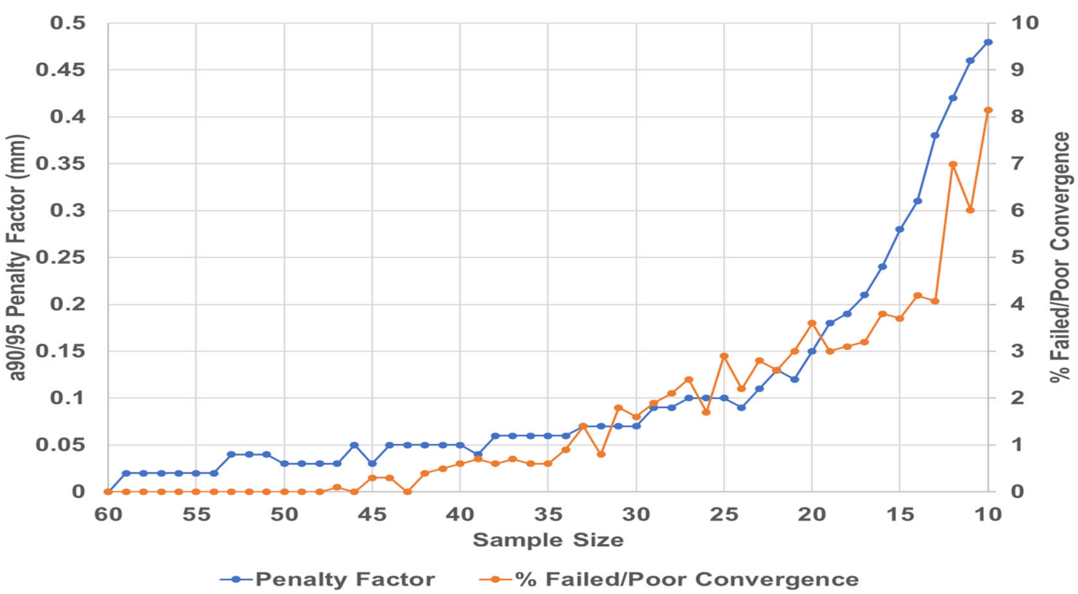

Figure 7: $a_{90/95}$ penalty factor & % convergence as a function of specimen subset

CONCLUSIONS

This present work investigates a novel approach to deriving SHM-POD, leveraging existing tools from MH1823. CLDP is a variation on a traditional hit/miss analysis with two related nuances. First, rather than collecting a single datapoint from a recommended distribution of flaw sizes, data is collected continuously for each specimen as the flaw is grown, and subsequently selecting a set of two limit defining pairs above and below the detection threshold. Second, when performing the hit/miss analysis covariance must be used to account for the dependence of those pairs of data used for each specimen.

Experiments were conducted using the WISP FCG to validate the CLDP approach. POD statistics were generated following a traditional hit/miss analysis, and the resulting data illustrated that the traditional analysis yields a relatively large variation depending on which points had been selected, and that the POD statistics were not very reliable for fewer numbers of tested specimens. Subsequently the same dataset was evaluated using CLDP. By steering the analysis to use the most “valuable” datapoints for each specimen, the POD statistics yielded better convergence and more reliable results that were consistent for many fewer tested specimens. Finally, a methodology was presented for using the CLDP approach to determine SHM-POD for smaller quantities of experiments by using a “penalty factor” to assure conservatism in the results.

Overall, this investigation was successful in its goal of developing and validating an approach for generating SHM-POD, leveraging the strengths of SHM methods to reduce the burden of testing more physical specimens. SHM has the potential to greatly improve asset availability while reducing sustainment costs through CBM, and the CLDP methodology provides a practical and reliable approach for evaluating the detection sensitivity of SHM sensors so that these statistics can be applied to risk assessments calculations consistent with current NDI practices.

ACKNOWLEDGEMENTS

This research was conducted at the Metis Design Corporation, Boston MA. It was funded by the U.S. Government under Air Force SBIR Agreement FA8649-20-9-9068. The views and conclusions contained in this document are those of the authors and should not be interpreted as representing the official policies of the U.S. Government.

REFERENCES

1. MIL-HDBK-1823A, DoD Handbook: Nondestructive Evaluation System, Reliability Assessment (07 Apr 2009)
2. Raghavan A., Kessler S.S., Dunn C.T., Barber D., Wicks S. and B.L. Wardle. "Structural Health Monitoring using Carbon Nanotube (CNT) Enhanced Composites." Proceedings of the 7th International Workshop on Structural Health Monitoring, September 2009, Stanford University.
3. Kessler S.S., Dunn C.T., Wicks S., Guzman deVilloria R.and B.L. Wardle. "Carbon Nanotube (CNT) Enhancements for Aerosurface State Awareness." Proceedings of the 8th International Workshop on Structural Health Monitoring, September 2011, Stanford University.
4. Kessler S.S., Thomas G., Borgen M. and C.T. Dunn. "Performance Analysis for CNT-based SHM Composite Structures." Proceedings of the 9th International Workshop on SHM, September 2013, Stanford University.
5. Kessler S.S., Thomas G., Borgen M. and C.T. Dunn." Carbon Nanotube Appliques for Fatigue Crack Diagnostics." Proceedings of the 10th International Workshop on SHM, September 2015, Stanford University.
6. Kessler S.S and C.T. Dunn "Reliability Assessment for NanoEngineered Fatigue Crack Sensor." Proceedings of the 11th International Workshop on Structural Health Monitoring, 12-14 September 2017, Stanford University.
7. Kessler S. S., Swindell P. and W. Meeker. “Detection Sensitivity Analysis for a Potential Drop SHM System,” Proceedings of the 12th International Workshop on SHM, September 2019, Stanford University.