AVIATION-BASED NONDESTRUCTIVE EVALUATION DATA ANALYTICS

Eric A. Lindgren[1]

[1] Air Force Research Laboratory, Materials State Awareness Branch,
Eric.Lindgren@us.af.mil

Abstract: Recent connectivity and digitization of aviation maintenance equipment has increased the potential of developing an Internet of Things 4.0 approach to enhance aircraft availability. Typically, these systems generate more data which nucleates interest in using analytical methods, such as enhanced data analytics (EDA), to increase the effectiveness of current aviation maintenance practices. It is important to recall that EDA methods are based on statistical regression and classification techniques. However, before such algorithms can be applied, considerations must be given to the quantity and quality (precision, accuracy, and noise) of the data to enable EDA. Several case studies are presented that explore these questions and indicate a careful assessment of the data is required to understand the accuracy and the distribution of the results from such analysis. The potential for the use of EDA is explored further using nondestructive evaluation (NDE) data. A significant challenge for these analytical methods is the limited amount of data captured for the features of interest, such as fatigue cracks and corrosion. Recall that trends in fleets lead to replacement or modification initiatives before an extensive number of flaws are present. To mitigate this limitation, the Air Force Research Laboratory (AFRL) and collaborators have explored and implemented alternative methods to assist in the analysis of NDE data that integrates at least two of the following: heuristics, model-based, and data-derived analysis techniques. These algorithms are called Assisted Defect Analysis (ADA). In addition, success has occurred when retaining the expertise of inspectors, i.e. humans-in-the-loop, to ensure the quality of the decision-making process. AFRL calls this approach Intelligence Augmentation (IA). The USAF has a rich history of using IA to analyze large NDE data sets, typically acquired from inspections that use automated scanning to acquire data. Several representative examples that include at least two of the three analysis methods are discussed, including the implementation process. These examples illustrate the benefit of integrating all resources to enable accelerated decisions with data limitations and the value of retaining humans-in-the-loop. Future opportunities include improved integration of models, especially as a function of their maturity through validation.

Keywords: Nondestructive evaluation, data analytics, assisted defect analysis

INTRODUCTION

There is a growing increase in interest and attention in EDA which are statistical methods for data analysis to self-extract attributes in the data, such as relationships and/or trends in in data that are not easily seen through typical human observation. With the potential to secure more Nondestructive Testing (NDT) data through the transformation to fully digital instruments connected in as envisioned by the Internet of Things (IoT) 4.0, there is an increased interest to use EDA methods as the diagnostic tool to determine if a flaw is present in NDT data. Justification for the use of EDA includes improved accuracy, improved reliability, and faster disposition time by decreasing or eliminating dependence on human interpretation and analysis of NDT data. The initial focus for the use of EDA addresses the detection of flaws although there is exploration in the use of EDA to provide additional information on characterizing the size and location of flaws.

When considering the applicability of EDA for flaw detection, it is important to recall that these technical approaches are based on statistical methods, namely regression or classification of data. The concept includes the use of multiple statistical methods in parallel combined with multiple layers of analysis to extract statistical trends in the data to enable decisions that are not readily detectable through more classical methods. This multi-dimensional data analysis methods frequently are called neural networks. These approaches can either be "supervised" using data with known ground truths for training, or be unsupervised and allowed to form the statistical relationships without training data. As these methods rely on data, critical attributes of the data must be considered for their use. This includes the amount of available data, the accuracy of the data, and noise present in the data.

CONSIDERATIONS WHEN USING EDA FOR ENGINEERING DECISIONS

The detection of flaws using NDT capabilities is an engineering decision that requires a statistical metric of capability to ensure safety of systems. In damage tolerance methods of integrity management [1], the capability is frequently validated by a probability of detection (POD) study that follows the guidance provide in MIL HDBK 1823 A [2]. To make these types of assessments possible, it is necessary to have metrics on the data that include such factors as quantity, quality, and fidelity, which includes such relatively simple factors as signal-to-noise (SNR) ratios. The outcome of a POD study that follows the guidelines of MIL HDBK 1823A will have appropriate statistical metrics to enable their use of these values in risk calculation that ensure the safety of systems. In the DAF, this is part of the Aircraft Structural Integrity Program (ASIP) [1] and the Propulsion Systems Integrity Program (PSIP) [3].

Aligned to POD studies, data metrics affect the use of EDA. These factors become increasingly more critical as a function of the risk to a system if a flaw is not detected. Therefore, detailed understanding of the data being used is important to enable proper use of the EDA algorithms. Recent work illustrated the impact of data quantity and SNR on the ability of a supervised neural network-based diagnostic [4]. The study used a synthetic data set and introduced Gaussian noise at different percent levels at different number of data points used to train the EDA algorithm. The neural network used for this study was a multi-layered perceptron with four layers and 50 layers in each hidden layer.

The results of this evaluation are show in Figure 1. The plot illustrates the log of the mean square error of the neural network as a function of SNR for differing number of data points in each data set. The SNR varies from an infinite value to one that is very poor of only 10 to 1. The number of data points in each data set varies from 50 up to 14,000. The outcomes are presented in standard box plots with the outliers indicated by red indices for each set of numbered data points.

It is clear from this data set is that the improved SNR and larger data sets results in a lower value for the mean squared error. This outcome is intuitively anticipated as it is expected that more data with higher fidelity will result in improved outcomes. However, this example highlights some of the challenges of using EDA for NDT data analysis. Even with the highest level of SNR, the smaller data sets have outliers that are considerably deviant for the mean values. When considering the impact of safety of systems, these outliers are the equivalent of a large, missed flaw that could impact the safety of a system. It is important to recall that it is not the smallest flaw that can be detected, but the largest flaw that could be missed that impacts the safety of a system. This is especially true in aviation where single load path structures are expected to have an extraordinarily low risk of failure [1].

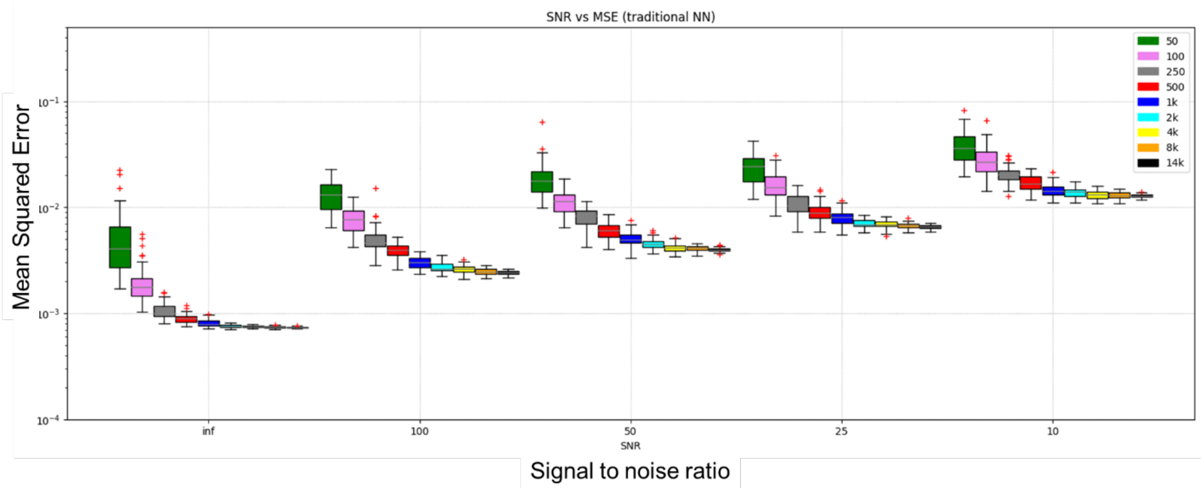

Figure 1. Multi-layer perceptron results illustrating mean square error as a function of data quantity and SNR

This data sensitivity study demonstrates two critical issues that need to be considered when applying EDA algorithms to NDT data. The first is the number of data points required to enable improved performance of EDA methods. Large training sets of actual flaws are hard to generate due to the time and cost of preparing such samples. A common complaint of POD studies that follow the guidance of MIL HDBK 1823A is the cost to prepare samples with characterized flaws. For a vs a-hat assessments this is 40 flaws and for hit-miss assessments it is 60 flaws. Large data sets of flaw responses in NDT data are difficult to find and/or generate as the engineering response to the detection of a growing number of flaws is either to modify or replace the structural element of concern before a large population of flaws is present. An option that has been pursued includes the use of simulation to generate the required data sets for training. However, a challenge is to make these simulations representative of the flaws found in actual structures. This approach requires a validation process that would require a similar sized data set of empirical data to fully validate the simulation from an engineering perspective.

The second issue is the ability to address outliers and nuances in data that can be indicators of flaws. The concern is the tendency of statistical methods to ignore such features when using large data sets. Unless the attributes of the outlier and nuance change in data is included in sufficient large quantities in training, the approach would tend to dismiss such features in the data which could result in missed flaws. Conversely, if the EDA is sensitive to outliers, then the concern becomes that large number of false calls that could decrease the value of implementing the EDA algorithm.

Thus, the lessons learned from the analysis of representative data include the need to have the right data for training, including multiple flaws that are independent from each other. It is extremely important to recall that resampling the same data is not acceptable unless proper statistical methods to address correlated data are included in the analysis. Similarly, it is not acceptable to test EDA methods using the same data that was used for training. Another aspect is to ensure factors, such as SNR, that can affect the statistical analysis of data are included in the training data sets. In addition, if simulation data is used in training, it must be from validated models that capture all the anticipated variances found in the NDT data for the anticipated inspection. Lastly, the desired precision and accuracy of the diagnostics to be performed by the EDA must be defined to ensure the amount of available data is sufficient to meet these objectives. This last consideration is especially true for unsupervised methods.

CHALLENGES FOR EDA FOR NDT

As indicted by the sensitivity studies in the previous section, a significant challenge for the use of EDA in NDT data is to capture the effect of all the factors that can influence the capability to detect the flaws of interest. Figure 2 is a representation of these factors that the author has used extensively to illustrate the additional challenges when migrating from a laboratory to an operational environment. The three general classes of challenges can be summarized as equipment variability, structural complexity and variability, plus flaw complexity and variability. In addition, these parameters can change as a function of the life of a system which increases the capability validation difficulty of the NDT system when integrated into system life management.

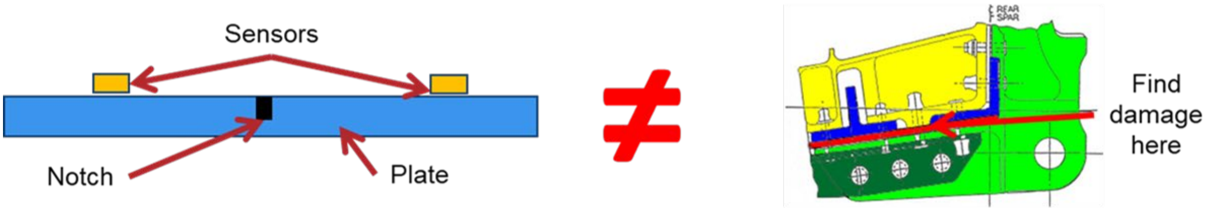

Figure 2. Representative increase in challenges when migrating from a laboratory environment (left) to an operational environment (right)

Equipment variability is the easiest of the three sources of variability to address from a research and development perspective. The variability in equipment settings can be defined and managed, but the unknown that frequently needs to be quantified is sensor variability and its impact on the diagnostics of flaws. Common NDT procedures address this with calibration processes which alleviate many of these concerns. However, small changes in sensor configuration, such as coil tilt within eddy current sensors or slight depolarization of well used ultrasonic transducers, can have an effect if the flaw detection response is used in training or testing. Experience has shown this can have an effect, but it is the lowest impact factor of the three items.

Flaw to flaw variability can have a much greater impact on the NDT response. Previous studies have illustrated that the same size flaw can vary the amplitude response from an eddy current response by over 20 percent of a full screen height reading [5]. Similar results can occur in ultrasonic testing as well as other NDT techniques. For ultrasound, fatigue crack morphology and tortuosity can affect a response. Local stress considerations from a fit-up of assemblies and changes due to use can vary crack closure which, in turn, affects the magnitude of the ultrasonic signal. The variability can be addressed in simulation provide all the attributes of the flaw that affect detection are included in the simulation studies. This includes their interaction, which can become a very large study, especially when considering engineering level validation of the simulation.

While flaw to flaw variability can be broadly categorized as a function of the type of flaw, structural variability can become much more challenging in the analysis of NDT data. This is largely due to the extensive range of aviation structures evaluated by NDT. In addition, other considerations include the materials being used, including metals, polymers, and composites, the manufacturing process being used, which are too many to list and can include automation, partial automation, or hand assembly, plus the assembly process used to joint components, such as fastening and bonding.

A significant challenge is how to evaluate how all these parameters can affect the NDT response, both individually and in multiple combinations. Consider the simple fastened joint between two metal surfaces. Factors that need to be included in a sensitivity study include up to 22 variables addressing equipment, flaws, and structure [6]. Structural considerations include such things as composition of each layer, the possibility of shims and their composition, assembly quality, such as fastener hole tilt or skew, and fit-up stresses as a function of what type of fastener is used and how it is installed. In addition, how these factors change as a function of time due to maintenance, repairs, modifications, and even use need to be included.

Using EDA techniques for these applications can become very daunting when considering all the parameters that need to be addressed to make diagnostic decisions using automated processes. This includes how the statistical processes adjust to account for changes that occur as a function of time. In addition, how these affect the diagnostic capability of the NDT data must be validated to enable their use in system risk and life management. Therefore, the proper capturing of these factors in statistically representative methods presents itself as a significant challenge, but also a significant research and development opportunity.

DAF APPROACH TO ASSISTED ANALYSIS OF NDT DATA

The Air Force Research Laboratory (AFRL) has been leading the development of algorithms to assist in the diagnostics of NDT data, including one of the first implementation for an aviation application [7]. Attributes that have made this approach successful include the use of multiple approaches to develop algorithms for the diagnostic capability combined with the approach that the algorithms will not replace all human interpretation of NDT data. The algorithms are used as a capability to facilitate and guide the interpretation to make the workload on an inspector easier and focused on the critical elements of the diagnostic process that do not easily rend themselves for automation. AFRL has called this approach Intelligence Augmentation (IA), but an alternative term being sued in the scientific community is Collaborative Intelligence (CI) [8]. This reflects how software tools and capabilities can be used to assist in the analysis of NDT data, which AFRL has named Assisted Defect Analysis (ADA).

ADA algorithms combine multiple approaches to provide the optimized method to facilitate NDT diagnostics. These can be placed into three general categories. The first uses heuristic-based methods that incorporate “rules of the road” that closely mimic the methods by which inspectors interpret data. The second is a model-based algorithms that use simulation to capture the impact of variability and can compensate for this variability in the diagnostics. The third uses EDA methods to tease out as a much diagnostics information as possible from data sets that are frequently much smaller than what would be required for robust EDA analysis.

Successful application of ADA has frequently included at least two of these approaches into an integrated diagnostic algorithm for the specific NDT application being addressed. This include the use of test data to ensure the intent of the application is being met and that the available data meets the needs of the application before a comprehensive validation study is accomplished. The output of the ADA diagnostic is not the final disposition of an indication. Depending on the application, the output can be used to screen data that has no criteria to contain a defect to enable inspectors to focus their attention to suspect portions of the inspection data. Alternatively, the output can be used to provide guidance on the nature of an indication so the proper disposition process can be rapidly identified and implemented, minimizing the time a system is in the inspection stage of a maintenance process. The key attribute of this approach is the human inspector is kept in the loop. The human functions to ensure data quality, data fidelity, and can review any ADA outputs to make the final determination regarding an indication.

REPRESENTATIVE DAF SUCCESSES

The following represents several examples developed by AFRL and transitioned to the DAF. The ADA capabilities are presented as a function of increasing complexity from the perspective of combining the three technical approaches outlined in the previous section. However, this order should not be considered a listing of increasing complexity as each application had its unique degrees of complexity and used different approaches tailored to the desired outcome.

A representative application that emphasizes the use of heuristics occurs in the manufacturing of aerospace composite structures, especially primary load carrying structures such as wing and fuselage skins. These parts require 100 percent ultrasonic inspection to detect delaminations and porosity where common rejection criteria are for delaminations greater than 0.25 inches in diameter or porosity that exceeds 2 percent. When considering the large areas to be inspected at manufacturing (note: this is not a requirement once a system is fielded), a bottleneck in the production flow can occur with the large volume of data to be assessed by inspectors. To minimize this bottleneck, a heuristic-based algorithm was developed to closely mimic the steps taken by an inspector to review data collected from these inspections [9].

The ADA algorithm leverages the available A-scan and B-scan data that accompanies the C-scan data. Multiple steps are taken in each of the three data representations to determine if an indication has features associated with delaminations that exceed the reject criteria. The representative result is shown in Figure 3 where C-scan features are identified as suspected defects and others are identified as benign. Though both may appear similar in the C-scan, attributes of the front wall, back wall, and volumetric gating can be used to distinguish between acceptable and rejectable features. The rejectable features are highlighted to the trained inspector who makes the final determination regarding the indication. With this approach, inspection processes have been greatly accelerated, though exact metrics are not available for publication.

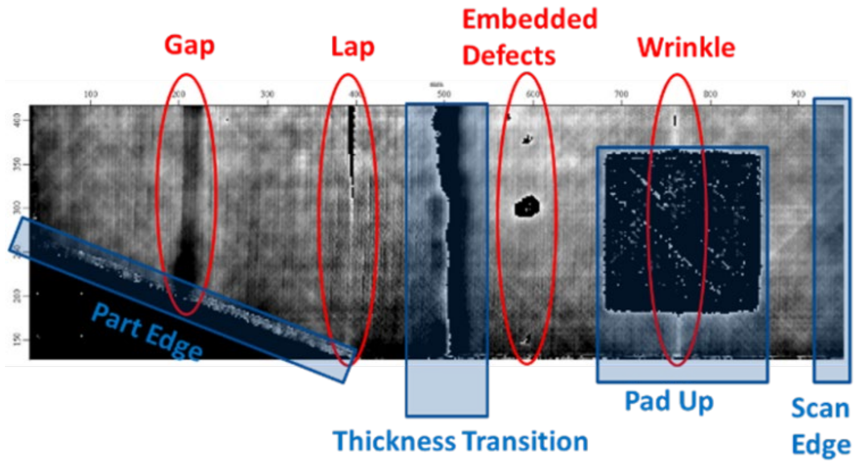

Figure 3. Ultrasonic c-scan of a composite test article indicating regions identified by the ADA algorithms as potential flaws

Another representative case study includes the use of both simulation and heuristics to identified flaws and discriminate between types of flaws. The specific application is for rotating turbine engine components evaluated by an automated inspection system that can provide highly registered data. Using a combination of model-based assessments and heuristic analysis methods, the response from data with varying probe conditions can be evaluated and provide guidance on what features are from suspected indications and what are due to the probe variability [10]. A representative illustration of the experimental and model data being compared in impedance planes is show in Figure 4 and highlights the impact of small probe variations on the attributes of the impedance plane not due to a flaw. Additional steps in the develop process resulted in the ADA algorithms providing guidance to the inspectors when features in the data indicated non-metallic inclusions were present as opposed to fatigue cracks in the components being assessed. The ADA being developed for this application is in its final stages of refinement before it will be evaluated by a formal validation process.

The third example combines elements of heuristics, simulations, and large data set analysis to realize a successful outcome on a very complex inspection. The application addresses the lower forward spar cap on C-130 aircraft [7] as shown in Figure 5. The approach leverages development at the academic level for both the generation and detection of ultrasonic creeping waves [11], plus the use of algorithms to discern the presence of cracks in a less complex, but still challenging, application [12]. As described in [7], the solution included the use of analytical methods to fully represent the propagation paths within the structure, simulation tools to explore various attributes of the inspection data as it propagates in the structure, plus the use of advanced processing methods, namely echo dynamics and local correlation functions, to discriminate between responses from potential flaws to those from other geometric reflectors found intermittently in the structure. In addition, over 2000 representative inspection opportunities from both harvested and mock-up test articles were used to refine the decision-making process for the ADA algorithms.

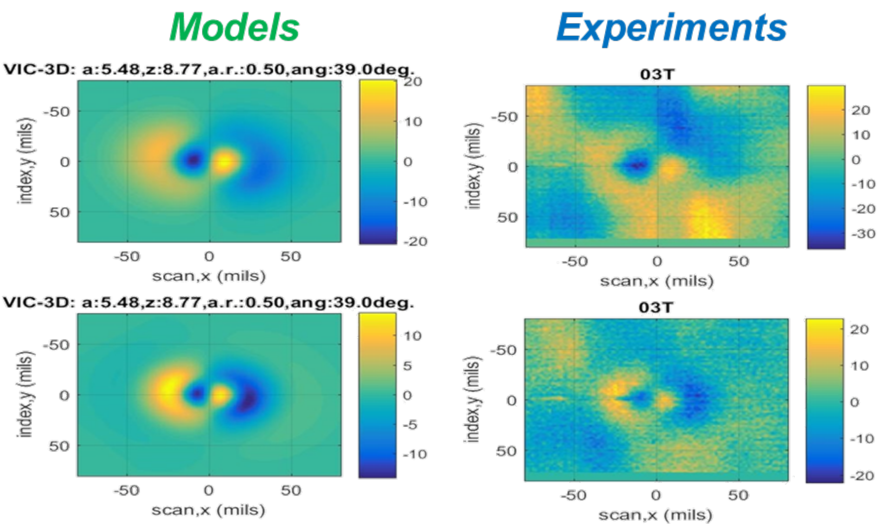

Figure 4. Simulation and experimental data from representations of eddy current scans indication differences due to probe variability.

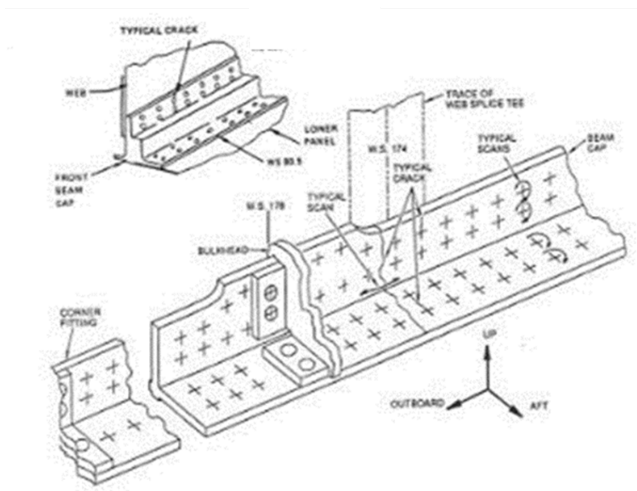

Figure 5. Lower forward spar cap of the C-130 illustrating complexity of the regions to be inspected.

The inspection process was fully validated by a comprehensive POD study before being deployed. The inspections were accomplished by contractor field teams that would collect the data and ensure it had sufficient quality to be evaluated by the ADA. Suspected indications identified by the ADA was sent to an NDT engineer to make a final determination if the indication was confirmed and needed to be sent to engineering for disposition.

The next generation of ADA will expand the capability of algorithms to facilitate the identification of flaws to the capability to characterize the flaws in ways tat are not possible today. While inspectors can use methods to approximate flaw size, attributes like fatigue crack depth are especially challenging. However, using a combination of heuristics, simulations, and data-driven analytical methods, the use of ADA to determine the depth of a fatigue crack from a bolt hole eddy current inspection was shown to have an accuracy of 8.5 percent for fastener holes with minimal variability [13]. The nest steps in the development process use this integrated approach to address fastener hole variability, such as skew and out-of-round attributes, and provide a crack depth with a statistical measure of accuracy to enable rapid disposition of these flaws in aerospace structures.

SUMMARY

There is a continued potential for EDA methods to enhance data analysis and diagnostics for NDT data. However, there needs to be a realistic approach that includes evaluation of the data quantity, quality, and fidelity. This ensures it has the desired attributes that enable the EDA techniques to provide outcomes with sufficient statistical metrics for the results to be used in engineering decisions. This is especially significant for applications where NDT is used to ensure safety of critical systems. A representative example illustrates the challenges in using EDA techniques for smaller and noisy data sets. The outcome of the results, as quantified by the mean square error, has a broader range in value and can have outliers that would imply potentially missed flaws if this approach was used for NDT data sets. The data for flaws can be augmented by simulated results, but these must contain all the anticipated variability and complexity of the NDT evaluation technique.

Representative variability includes attributes of the evaluation equipment, the flaws, and the structure being assessed, especially when it has been maintained, repaired, modified, or even changed due to use as is common for aging aviation assets. Variability in NDT equipment can be frequently addressed by calibration and other processes. However, it is important to note that nuances in the data that could be within calibration could make the use of EDA challenging due to the statistical nature of these methods. Flaw variability is another challenge due to the high degree of differences in flaws. As an example, no two fatigue cracks have the exact same morphology or tortuosity. This causes variability in the response even from flaws with the same overall size. Structural variability increases as the use or age of the system increases. Representative examples, such as local fit up stresses from fastened assemblies, indicate how local changes can affect the NDT interrogation method and lead to variability of an EDA diagnostic even if the flaw remains the same.

To address these challenges, AFRL is pursuing hybrid approaches that integrate heuristic, model-based, and data-driven diagnostic algorithms to facilitate and reduce the workload of inspectors while not taking them completely out of the loop. Representative examples for several DAF related applications have demonstrated the power of combining at least two of these methods to enable complex inspections and diagnostics of NDT data. The ADA algorithms are combined with human analysis to maximize the value of the algorithms by reducing the workload of inspectors so they can focus on the critical data that could be indications of flaws being present. Future work includes plans to expand the capabilities of ADA algorithms to characterized flaws with statistical metrics of accuracy. Initial development efforts have shown the potential of this capability, which would decrease the disposition time of indications and increase availability of the system to the end-user.

ACKNOWLEDGMENTS

The author expresses his deep appreciation for the pioneering contributions and collaboration with Dr. John Aldrin of Computational Tools. Mr. David Forsyth of Texas Research Institute-Austin is recognized for his contribution to implementing ADA algorithms for multiple NDT applications. The EDA analysis would not be possible without the work performed by Mr. Tushar Gautam and Drs. Kirby, Hochhalter, and Zhe of the University of Utah.

REFERENCES

1. MIL-STD-1530D, DoD Standard Practice for Aircraft Structural Integrity Program, (31 August 2016)

2. U.S. Department of Defense, Handbook, Nondestructive Evaluation System Reliability Assessment, MIL-HDBK-1823A, (7 April 2009).

3. MIL-STD-3024, Department of Defense Standard Practice, Propulsion System Integrity Program (15 APR 2008)

4. Lindgren, E.A., “Intelligence Augmentation for Aviation-based NDE Data,” 2022 Aircraft Structural Integrity Program Conference, available at: http://www.arctosmeetings.com/agenda/asip/2022/proceedings/presentations/P23251.pdf

5. Forsyth, D.S., Ocampo, J., Millwater, H., and Montez, J., “Structural Health Monitoring, Risk, and, Reliability, 2015 Aircraft Structural Integrity Program Conference, available at: https://www.researchgate.net/publication/344886205_Structural_Health_Monitoring_Risk_and_Reliability

6. Lindgren, E.A.,Knopp, J.S., Aldrin, J.C., Steffes, G.J., and Buynak, C.F., (2007) "Aging Aircraft NDE: Capabilities, Challenges, and Opportunities," *Review of Progress in QNDE*, V26, p1731.

7. Lindgren, E. A., Mandeville, J. R., Concordia, M. J., MacInnis, T. J., Abel, J. J., Aldrin, J. C., Spencer, F., Fritz, D. B., Christiansen, P., Mullis, R. T., and Waldbusser, R., (2005) "Probability of Detection Results and Deployment of the Inspection of the Vertical Leg of the C-130 Center Wing Beam/Spar Cap", 8th Joint DoD/FAA/NASA Conference on Aging Aircraft.

8. Epstein, S. L. (2015), “Wanted: Collaborative Intelligence,” *Artificial Intelligence*, 221, 36-45, https://doi.org/10.1016/j.artint.2014.12.006

9. Aldrin, J.C., Forsyth, D.S., and Welter, J.T., (2016) “Design and Demonstration of Automated Defect Analysis Algorithms for Ultrasonic Inspection of Complex Composite Panels with Bonds,” *Review of Progress in QNDE*, 1705, P. 020002,

10. Aldrin, J.C., Oneida, E.K., Shell, E.B., Sinha, V., Keller, K., Na, J.K., Hutson, A.L., Sabbagh, H.A., Sabbagh, E., Murphy, R.K. and Mazdiyasni, S., (2019) “Model-based Inversion of Eddy Current Data for Classification and Sizing of Planar and Volumetric Discontinuities,” Electromagnetic Nondestructive Evaluation XXII, 44, p.80

11. Nagy, P. B., Blodgett, M., and Golis, M. (1994). “Weep Hole Inspection by Circumferential Creeping Waves,” NDT & E International, 27(3), 131-142.

12. Aldrin, J.C., Achenbach, J.D., Andrew, G., P’an, C., Grills, R., Mullis, R.T., Spencer, F.W. and Golis, M., (2001) “Case Study for the Implementation of an Automated Ultrasonic Technique to Detect Fatigue Cracks in Aircraft Weep Holes,” Materials Evaluation, 59, n 11, p.1313-1319.

13. Aldrin, J. C., Lindgren, E. A., and Forsyth, D. S., (2019) "Intelligence Augmentation in Nondestructive Evaluation," *Review of Progress in QNDE*, AIP Publishing, vol. 2102, no. 1, p. 020028.