QUANTITATIVE MULTI-INSPECTION-SITE COMPARISON OF PROBABILITY OF DETECTION FOR VIBROTHERMOGRAPHY NONDESTRUCTIVE EVALUATION DATA

Ming Li[1], Stephen D. Holland[2], William Q. Meeker[1]

[1]Center for Nondestructive Evaluation and Department of Statistics, Iowa State University, Ames, Iowa, USA

[2]Center for Nondestructive Evaluation and Department of Aerospace Engineering, Iowa State University, Ames, Iowa, USA

This paper describes the estimation of probability of detection (POD) for a vibrothermography inspection procedure. The results are based on a large scale experiment on specimens with two different kinds of metal containing fatigue cracks. The specimens were tested independently at three inspection sites: Iowa State University (ISU), Pratt and Whitney (PW) and General Electric (GE). Despite the substantially different vibrothermography configurations and experimental measurement responses, the estimated PODs as function of crack length and dynamic stress were similar for all three inspection sites, which make quantitative POD comparisons possible across different inspection sites.

Keywords: maximum likelihood, noise interference model, POD, sonic infrared inspection, thermosonics

1. INTRODUCTION

Vibrothermographic inspection, also known as sonic infrared or thermosonics, is a nondestructive evaluation (NDE) method that can be used to detect delaminations in composite materials or cracks in metals [1-5]. There has been, however, little quantitative research to study the transduction from vibration into heat that underlies vibrothermography and the effect that these mechanisms will have on probability of detection (POD). There is also concern about the large amount of experimental setup variability across and within vibrothermography systems [1-5]. To address these concerns, a large experimental study involving vibrothermography inspection was conducted on two specimen sets at three different inspection sites. This study is described in Holland et al. [6]. A collection of 63 Titanium Ti-6Al-4V specimens and 63 Inconel-718 specimens containing fatigue cracks were independently tested at ISU, PW and GE. Each inspection site has a different vibrothermography system. In this paper we also describe statistical models and methods that can be used to estimate vibrothermography POD from the experimental data.

Holland et al. [6,7] have developed an algorithm, based on a physical model, to reduce the vibrothermography sequence-of-image data in each experimental measurement into a scalar measure of temperature increase. The algorithm performs a surface-fit of the heat from the crack to an elliptical Gaussian envelope. The heating temperature is estimated by integrating the observed heat over the peak of the Gaussian envelop and dividing by an enlarged area. The temperature calculated in this way is

approximately 95% of the peak value of the surface fit. As illustrated later in this paper, for small cracks, the amount of heat generated is close to the noise level of the inspection system and there are many such observations for both the Titanium and the Inconel specimens. In applications like this where there is a large amount of noise in the data, the traditional statistical methods for estimating POD can lead to unconservative bias in POD estimates as the noise can lift the regression line which in turn lifts POD [8,9]. In this paper, we apply a noise interference model [9] to the vibrothermography data, providing POD estimates as function of crack length and dynamic stress. Our results show that estimates of POD for the different vibrothermography experimental configurations are similar. These results support the viability of using vibrothermography to detect and evaluate cracks inside metals.

This paper is organized as follows: In Section 2 we briefly describe the experimental configuration of the vibrothermography systems and the reduction of the complete experimental sequence-of-image data to a scalar temperature increase value. In Section 3 we present graphical displays of the scalar temperature increase as function of crack length and dynamic stress for the whole data set. Section 4 describes the noise interference model (NIM). Section 5 presents the detailed statistical analysis of the vibrothermography data sets for both materials. In Section 6, we present some conclusions.

2. EXPERIMENTAL SETUP

The particular vibrothermography inspection system used in our experiments is illustrated conceptually in FIGURE 1. This system involves an excitation source to excite

the sample, an infrared camera to record heating of the specimen, and a laser vibrometer to monitor vibration in the specimen. The excitation source (a piezo stack at ISU, an ultrasonic welder at PW and GE) is pneumatically pressed to the sample, and the sample itself is gripped with a rigid or compliant clamp. A coupling medium, such as paper, plastic, or cardstock is usually used to separate the tip of the vibration source from the sample. The specimen is typically excited from 1 to 2 seconds duration. The goal is to cause the crack surfaces to rub and generate heat. The sample surface temperature profile is captured by a sequence of images recorded by an infrared camera and the sample surface velocity is measured by a laser vibrometer. Both the temperature profile and the surface velocity are typically recorded at short time intervals for each measurement. The vibrometer sampling rates in the experiment were 1 MHz for ISU and 500 kHz for GE and PW, and the infrared camera sampling rates were 90 Hz for ISU and 189 Hz for GE and PW.

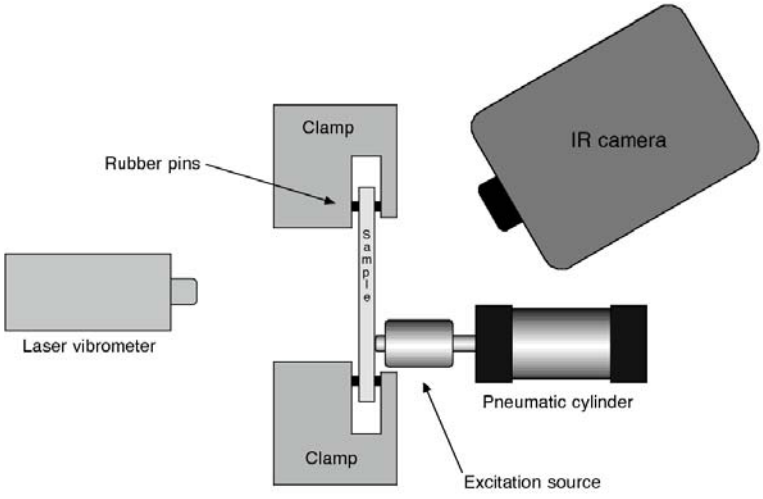

FIGURE 1. Common configuration for a vibrothermography inspection system

The ultrasonic welder and piezo stack that are used as excitation sources typically generate 1 to 2 kW of vibrational power at a fixed frequency such as 20 kHz. For this study the specimens were tuned to a nominal natural resonance near 20 kHz. During an inspection, the vibrational excitation power is coupled into the specimen near the natural resonance and frictional rubbing between crack surfaces generates heat. The known mode shape of the natural resonance allows calculation of the dynamic vibrational stress on the crack from the transverse velocities measured with the vibrometer. For each vibrothermographic inspection, the scalar heat-increase response and dynamic stress were obtained from the sequence of infrared images and the vibrometer records, as described in [6].

3. HEAT-INCREASE RESPONSE DATA

The scalar heat-increase response was modeled as a function of the dynamic stress and crack length. The inspection data sets for all three inspection sites are shown in FIGURE 2 where different symbols represent various dynamic stress ranges as indicated by the legend. The inspection system noise level is around 0.03K which is indicated by the horizontal dashed lines in FIGURE 2. This noise level was determined by the infrared camera specification and it is not sensitive to dynamic stress levels. All three inspection systems had infrared cameras of equivalent technology. Detection error is dominated by the Johnson noise of the infrared cameras, and all three systems had the same noise equivalent temperature difference. We chose 0.03K as the detection threshold so that we could have a fair comparison of PODs across inspection sites and a reasonably low probability of false alarms. It is clear that a large portion of the data is below the noise

level and the traditional statistical method for POD analysis is no longer valid. To better estimate the signal response from the noisy data, we used the noise interference model.

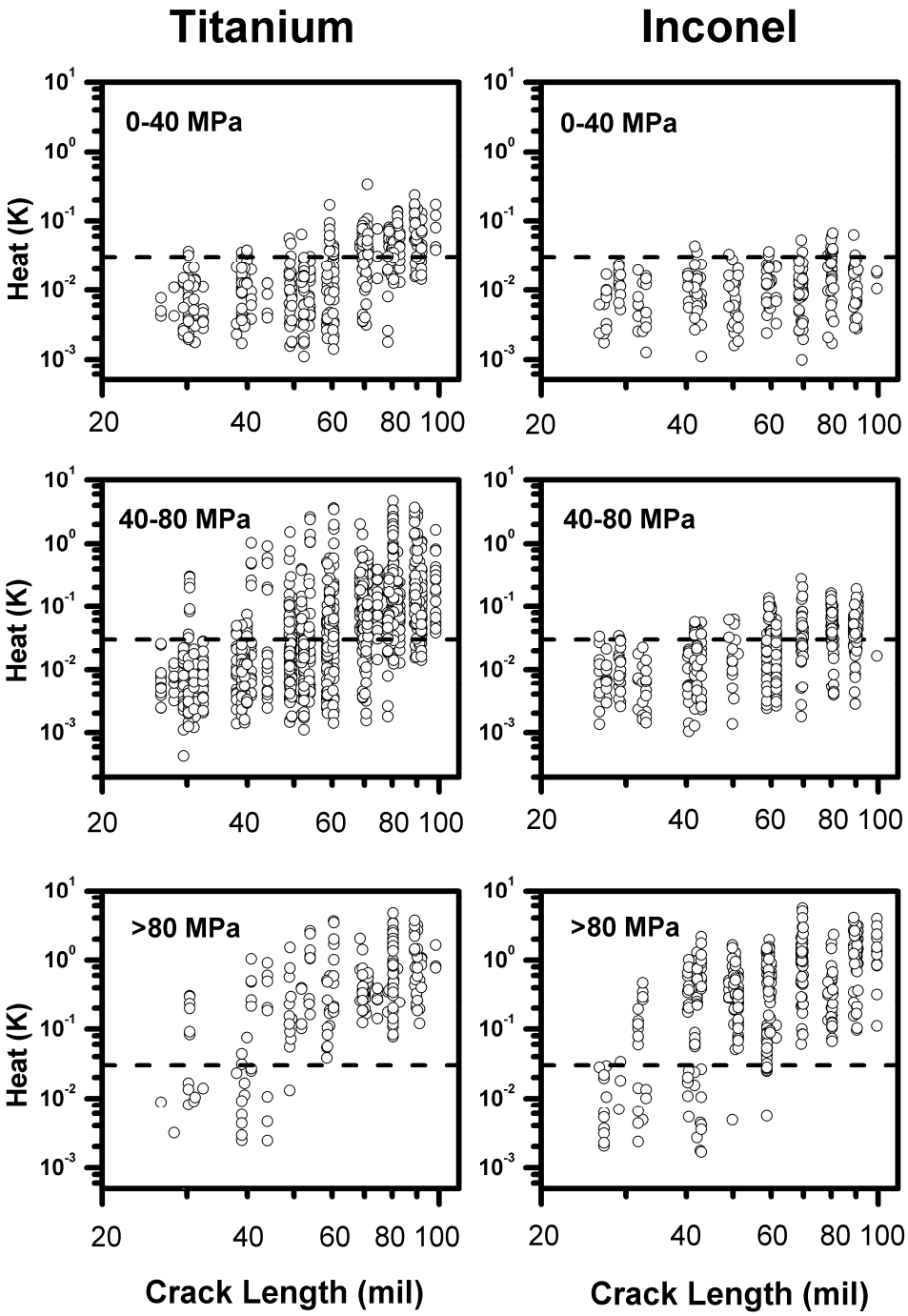

FIGURE 2. The measured vibrothermography heat-increases as a function of crack length and dynamic stress for Titanium specimens (left) and Inconel specimens (right). The horizontal dashed lines indicate the approximate noise level of the inspection system.

4. NOISE INTERFERENCE MODEL

The traditional statistical method for estimating POD from an NDE study with a quantitative response is the $\hat{a}$-versus-a method described, for example, in [8]. The traditional $\hat{a}$-versus-a method has, for small targets, an asymptotic limit for POD that approaches zero as the flaw size approaches zero. This characteristic is in contradiction to the fact that for zero crack size (i.e. specimens without cracks) the POD should be approximately equal to the probability of a false alarm (i.e., the probability that a noise artifact will exceed the detection threshold). When NDE measurements are taken in locations where there are no target flaws, the reading can still be of some value to quantify measurement and background noise. Such noise data are usually used to estimate the probability of a false alarm. In locations where there are very small flaws, the observed response could be the result of a noise-causing artifact rather than the small flaw. Based only on the measurements, we cannot be sure whether the measurement came from a flaw or a noise-causing artifact.

To account for a possible mixture of flaw and noise responses, we extend the $\hat{a}$-versus-a POD analysis to our noise interference model (NIM). A more detailed description of the NIM can be found in [9]. Before fitting a model, transformation of the original physical quantities are often needed. For example, one might use a logarithm or square root transformation, depending on the data itself and its variance structure. We define the observed measurement response or its transformation as y, the signal response or its transformation as y_{signal}, and the noise response or its transformation as y_{noise}. The NIM components are as follows:

- The signal response is modeled as $y_{signal} = f(\boldsymbol{\beta}, \mathbf{x}) + \varepsilon_s$ where $\boldsymbol{\beta}$ is a vector of regression parameters, $\mathbf{x}$ is a vector of explanatory variables such as crack length and dynamic stress or their transformations, and ε_s is the residual term, assumed to be normally distributed with mean zero and variance σ_s^2, i.e. $\varepsilon_s \sim N(0, \sigma_s^2)$.
- The noise response y_{noise} is assumed to be normally distributed with mean μ_n and variance σ_n^2.
- The observed measurement response (i.e. the experimental measurement) is the maximum of the signal and noise: $y = \max(y_{noise}, y_{signal})$.

With the measurement data and specified $f(\boldsymbol{\beta}, \mathbf{x})$, estimates of the parameter vector $(\hat{\boldsymbol{\beta}}, \hat{\mu}_n, \hat{\sigma}_n^2, \hat{\sigma}_s^2)$ and the estimated variance covariance matrix of these estimates can be obtained through standard maximum likelihood methods described, for example, in [10]. The POD estimate, as a function of the explanatory variable $\mathbf{x}$, can be calculated through $\mathrm{POD}(\mathbf{x}) = \Pr(y > y_{\mathrm{th}}) = 1 - \Phi\left(\frac{y_{\mathrm{th}} - f(\hat{\boldsymbol{\beta}}, \mathbf{x})}{\hat{\sigma}_{\mathrm{s}}}\right)\Phi\left(\frac{y_{\mathrm{th}} - \hat{\mu}_{\mathrm{n}}}{\hat{\sigma}_{\mathrm{n}}}\right)$ where y_{th} is the detection threshold and $\Phi(x)$ is the standard normal cumulative distribution function [9].

We have shown theoretically and by simulation that the NIM asymptotic POD for zero crack size is very close the probability of false alarm. The standard error of the

estimated NIM POD is smaller than the traditional model, which indicates that the NIM model provides a better fit to the data and will provide better statistical inferences [9].

5. NIM APPLIED TO VIBROTHERMOGRAPHY DATA

5.1 NIM Model Application Overview

In this section, we first apply the NIM to the vibrothermography Titanium and Inconel data sets separately for each of the inspection sites. Then we compare the POD results across inspection sites.

In this paper we have two explanatory variables: crack length and dynamic stress, and the relationship between the response variable and explanatory variables is $y_{\text{signal}} = f\left(\beta_0, \beta_1, \beta_2, x_1, x_2\right) + \varepsilon_s$ where $y_{\text{signal}} = \log_{10}(heat)$; $x_1 = \log_{10}(crack\ length)$; and $x_2 = \log_{10}\left(stress\right)$. Different choices of $f\left(\beta_0, \beta_1, \beta_2, x_1, x_2\right)$ are used for the Titanium specimens and the Inconel specimens because material differences affect the underlying generation of heating

Although for the same material the estimated parameters are different across inspection sites, due to the significant variation of inspection system setup, the final estimated POD functions for the three inspection sites are similar.

5.2 Inspections on the Titanium Specimens

As mentioned previously, the inspection system designs at the three inspection sites are substantially different, with different excitation sources and amplitude settings that were used to obtain a range of excitation amplitudes. Each time a unit is energized,

we obtain both the signal response (amount of temperature increase) and the amount of dynamic stress in the specimen. The differences among the inspection sites are partly reflected in the differences in the distributions of the dynamic stress values at each inspection site. FIGURE 3 shows the distribution of dynamic stress for Titanium specimens at the three inspection sites: the ISU data have dynamic stress range from 0 to around 120MPa; the PW data have dynamic stress up to 200MPa, while the GE data has a bi-modal distribution with many dynamic stress values larger than 200MPa. The bi-modal distribution for the GE inspection site is attributed to the resonance frequencies of the GE ultrasonic welders matching the resonance frequencies of those Titanium specimens precisely, yielding anomalously high motion amplitudes and stress levels.

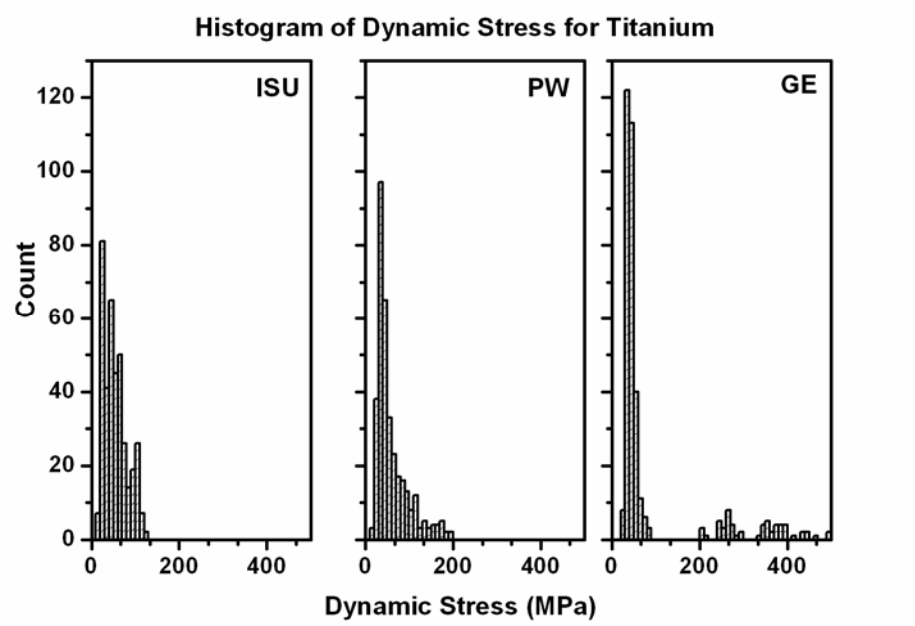

FIGURE 3. Distributions of dynamic stress for Titanium specimens at ISU, PW and GE.

Both the physical model relating heat-increase to vibration [6,7] and our statistical analysis suggest that the logarithm of heat-increase is a linear function of the logarithm of crack length for a fixed level of dynamic stress. Based on knowledge of the heat

generation mechanism and comparisons among a number of different models we found that the interaction model $y_{\text{signal}} = \beta_0 + \beta_1\left(x_1 - x_1^0\right) + \beta_2\left(x_1 - x_1^0\right)\times\left(x_2 - \overline{x}_2\right) + \varepsilon_s$ provides a good description of the data. Here $x_1^0 = \log_{10}(25)$ is a fixed intersection position at crack length 25 mils; and $\overline{x}_2$ is the average of all measured x_2 values. This model implies a linear relationship between logarithm transformation of heat and crack length for each level of dynamic stress. For a given dynamic stress the slope of the regression line is $\beta_1 + \beta_2\left(x_2 - \overline{x}_2\right)$. All of these lines are constrained to intersect at a common point of 25 mils. We found it necessary to use this fixed point of intersection to avoid having the lines intersect within the range of the data. An intersection within the range of the data will result in a higher POD for smaller stress level for cracks smaller than the intersection point, which physically inconsistent with the heating-generating mechanism in vibrothermography inspection procedures. The signal response parameters $\beta_0, \beta_1, \beta_2,$ and σ_s, as well as the noise parameters μ_n and σ_n, can be estimated from the inspection data by using the maximum likelihood method.

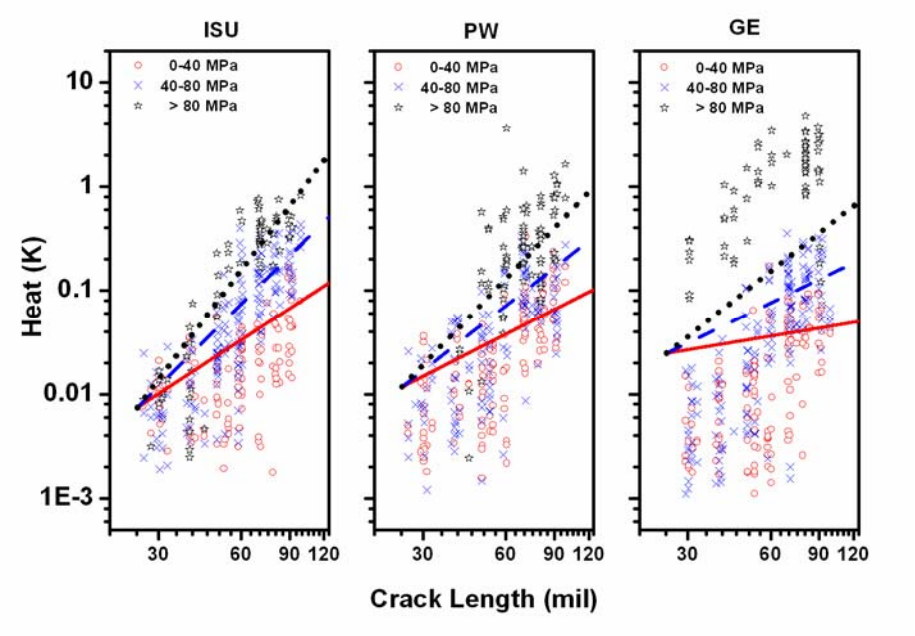

FIGURE 4. Separate NIM analyses for the Titanium data at ISU, PW and GE. The different symbols represent different dynamic stress ranges and the lines are the fitted results for fixed dynamic stress levels of 35MPa (solid lines), 60MPa (dashed lines), and 100MPa (dotted lines).

We fit this same model to the data from each inspection site. FIGURE 4 shows the data using different symbols to represent different dynamic stress ranges and fitted lines for dynamic stress levels 35MPa (solid lines), 60MPa (dashed lines) and 100MPa (dotted lines). At inspection site GE, the regression line with dynamic stress 100MPa (dotted line) is much lower than the data points of dynamic stress larger than 80MPa. This is because at GE many of the inspections in this range had dynamic stress values that were much larger than 80MPa, as indicated at the dynamic stress histogram (FIGURE 3). To keep comparisons consistent, we used the same dynamic stress levels for the regression lines for each inspection site. The parameter estimates from statistical analysis such as regression line intercept and slopes differ across inspection sites. These differences are a reflection of inspection system variations.

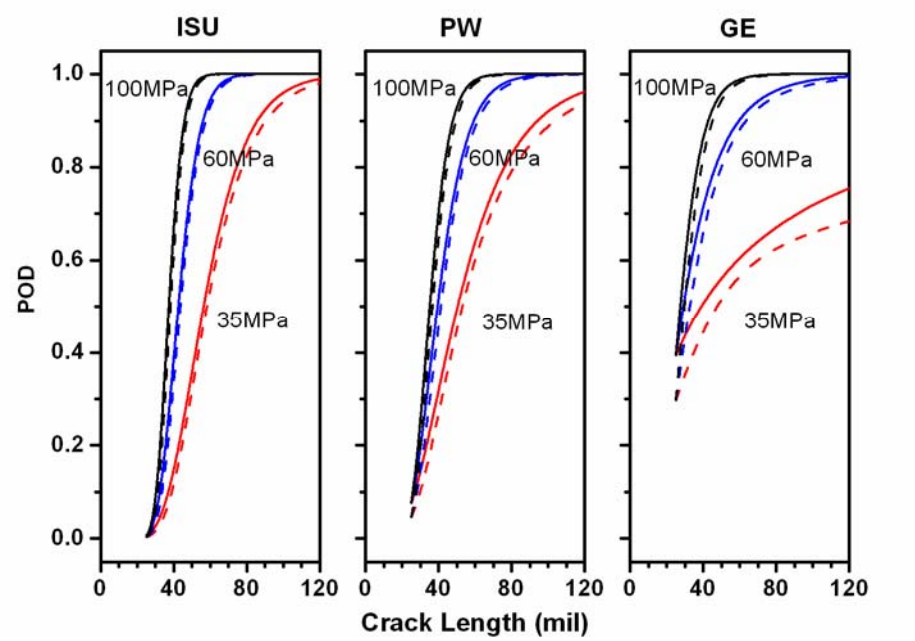

FIGURE 5. Separate POD curves for Titanium specimens at three levels of dynamic stress (35MPa, 60MPa, 100MPa) at ISU, PW and GE: The solid lines represent POD estimates and the dashed lines represent the corresponding 90% POD lower confidence bounds.

With the parameter estimates and the estimated variance covariance matrix of the estimates, the POD for fixed dynamic stress and its corresponding 90% lower bound can be calculated easily as described in Section 4. FIGURE 5 shows the POD estimates for inspections at 35MPa, 60MPa and 100MPa for all three inspection sites (solid lines) along with the corresponding 90% lower confidence bounds (dashed lines). The POD estimates and their lower bounds have similar patterns across all inspection sites except for the GE POD estimate at dynamic stress 35MPa. The relatively small number of data points above the detection threshold at 35MPa and the bi-modal distribution of dynamic stress levels are the main contribution the difference.

5.3 Inconel Data Sets

The Inconel specimens, when compared with Titanium specimens, present different dynamic stress distributions, as shown at FIGURE 6. The different material property between Inconel and Titanium is the main factor for the dynamic stress distribution difference. There are some inspections with dynamic stress larger than 200MPa in the PW data and many more inspections with very high dynamic stress in the GE data. The bi-modal distribution for the PW and GE inspection sites is again attributed to the resonance frequencies of their ultrasonic welders matching the resonance frequencies of those Inconel specimens precisely, yielding anomalously high motion amplitudes and stress levels. Note that such anomalously high amplitudes that are due to resonance matching between the ultrasonic welder and specimen were observed at PW using Inconel but not Titanium.

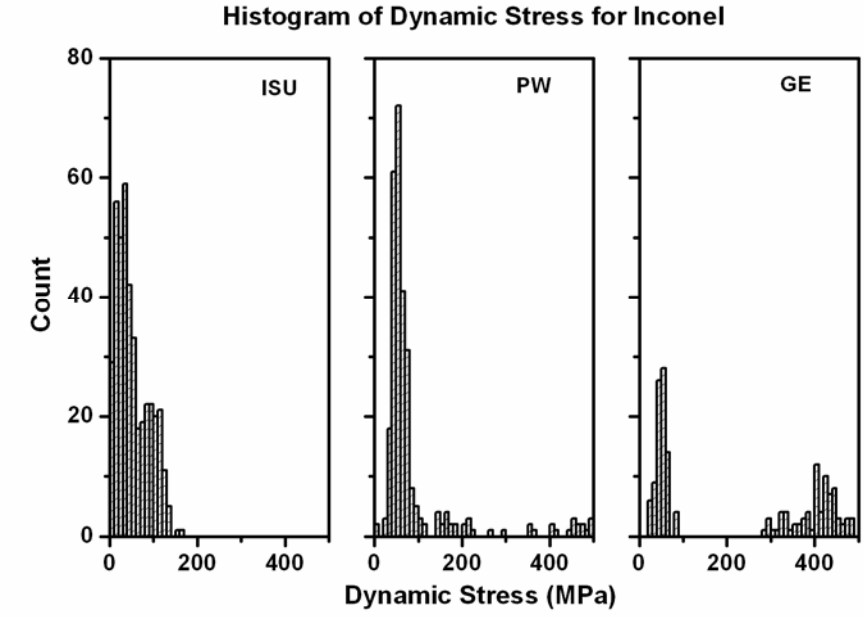

FIGURE 6. Distribution of dynamic stress levels for Inconel specimen across ISU, PW and GE.

For Inconel specimens, the statistical model $y_{\text{signal}} = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \varepsilon_s$ provides a better description of the data. The signal response parameters $\beta_0, \beta_1, \beta_2,$ and σ_s, as well as the noise parameters μ_n and σ_n, can be estimated from the inspection data by using the maximum likelihood method. For this model, the regression lines for different fixed levels of dynamic stress will be parallel as illustrated at FIGURE 7. The estimated POD curves and their corresponding 90% lower confidence bounds for fixed dynamic stress levels at each inspection site are shown at FIGURE 8.

As with the Titanium data, the heat increases in the Inconel specimens as function of crack length, behave differently across the three inspection sites as shown in FIGURE 7. The ISU and PW data have many observations with low heat-increase (below 0.01K), while GE data have many observations with large heat-increase (above 1.0K). The inspection system variations lead to the different distributions of dynamic stress which in turn causes the discrepancy of heat-increases. Again, to keep comparisons consistent, we continue to use the same dynamic stress levels (35 MPa, 60MPa, and 100 MPa) for the regression line and POD estimates at each inspection site.

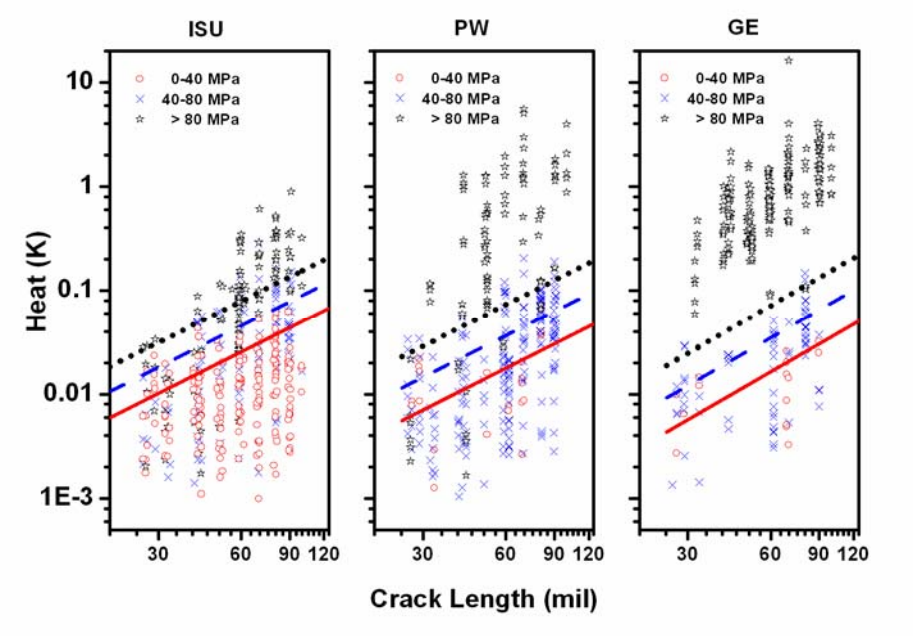

FIGURE 7. Separate NIM analyses for the Inconel data at ISU, PW and GE. The different symbols represent different dynamic stress ranges and the lines are the fitted results for fixed dynamic stress levels of 35MPa (solid lines), 60MPa (dashed lines), and 100MPa (dotted lines).

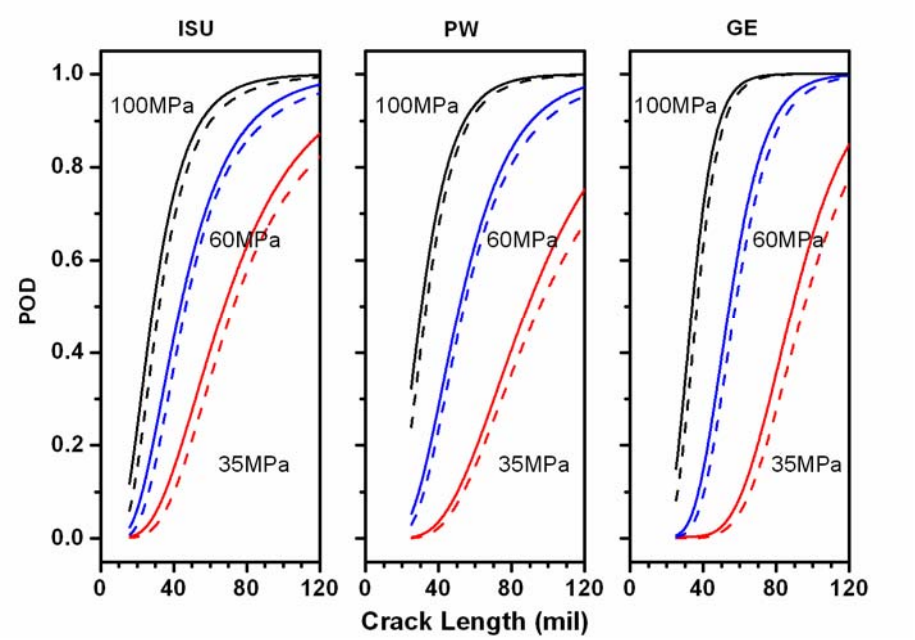

FIGURE 8. Separate POD curves for the Inconel specimens at three levels of dynamic stress (35MPa, 60MPa, 100MPa) at ISU, PW and GE: The solid lines represent PODs and the dashed lines represent the corresponding 90% POD lower bounds.

5.4 POD Comparisons

After analyzing both the Titanium and the Inconel vibrothermographic experimental data for each inspection site, we now compare the POD curves for a fixed dynamic stress level of 100MPa across the three inspection sites for both materials at FIGURE 9. The ISU and PW POD results are almost the same while the GE results are close but with some degree of offset which may due to one or both of the following reasons: (1) GE inspections have fewer data points around the dynamic stress level 100MPa, (2) GE inspections have a bi-modal distribution of dynamic stress and there are many data points at very high dynamic stress range as shown in FIGURE 3 and FIGURE 6. We believe that if the GE system were modified to limit amplitudes, eliminating the

very high dynamic stress inspections, then their POD results would likely be closer to those of ISU and PW

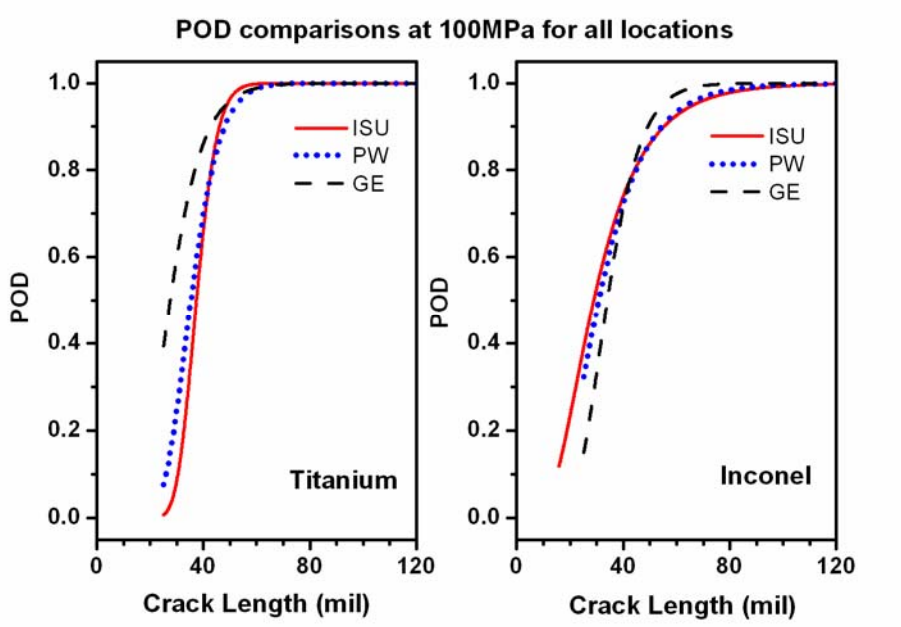

FIGURE 9. POD comparisons across all three inspection sites with fixed dynamic stress (100MPa) for Titanium (left) and Inconel (right)

6. CONCLUSIONS

In this paper, we have applied the noise interference model to a large set of vibrothermography inspection data of metal specimens with two different materials at three different inspection sites. Despite the large difference in the experimental configurations at three inspection sites, similar estimates of POD as a function of crack length for fixed values of dynamic stress were obtained for all inspection sites. This is the first quantitative, multi-inspection-site demonstration of the reliability for vibrothermography method for fatigue crack detection. The estimated POD obtained at this paper only applies to these particular cracks on these 126 specimens. Further

investigation of cracks and materials variability is required to extend the estimated POD to field applications. Another important extension is to develop a statistical method to calculate PODs for cases where dynamic stress is not measurable. In such cases POD could be expressed as a function of excitation amplitude and crack size. We expect that these quantitative results will be useful for the future development of model-assisted POD analysis.

ACKNOWLEDGMENTS

This material is based upon work supported by the Federal Aviation Administration under Contract #DTFA03-98-D-00008, Delivery Order #0058 and performed at Iowa State University's Center for NDE as part of the Center for Aviation Systems Reliability program. We would like to acknowledge Thadd Patton (GE), Tom Bantel (formerly of GE), Zhong Ouyang (Pratt and Whitney), and Christopher Uhl (GE, formerly of ISU) for their participation in the multi-inspection-site comparative experiment.

REFERENCES

1. X. Maldague, *Theory and Practice of Infrared Technology for Nondestructive Testing*, John Wiley and Sons, New York (2001).
2. J. Rantala, D. Wu, and G. Busse, “Amplitude-modulated lock-in vibrothermography for NDE of polymers and composites”, *Research in Nondestructive Evaluation*, **7(4)** 215-228 (1996).

3. L. D. Favro, X. Han, Z. Ouyang, G. Sun, H. Sui, and P. L. Thomas, “Infrared imaging of defects heated by a sonic pulse”, *Review of Scientific Instruments,* **71(6)** 2418-2421 (2000).
4. M. Morbidini, P. Cawley, T. Barden, D. Almond, and P. Duffour, “Prediction of the thermosonic signal from fatigue cracks in metals using vibration damping measurements”, *J. Appl. Phys,* **100** 104905 (2006).
5. C. Ibarra-Castanedo, J. M. Piau, S. Guilbert, N. P. Avdelidis, M. Genest, A. Bendada, and X. P. V. Maldague, “Comparative study of active thermography techniques for the nondestructive evaluation of honeycomb structures”, *Research in Nondestructive Evaluation*, **20:1**, 1-31 (2009).
6. S. D. Holland, C. Uhl, Z. Ouyang, T. Patton, M. Li, W. Q. Meeker, J. Lively, L. Brasche, and D. Eisenmann. “Relating vibrothermographic crack heating to crack size and crack motion” to be submitted (2010).
7. S. D. Holland and J. Renshaw, “Physics-based infrared image enhancement for thermography”, to be submitted (2010).
8. C. Annis, MIL-HDBK-1823A, *Nondestructive Evaluation System Reliability Assessment*, Standardization Order Desk, Philadelphia, PA (2009).
9. M. Li and W. Q. Meeker, “A noise interference model for estimating probability of detection for nondestructive evaluations”, Review of Progress in Quantitative Nondestructive Evaluation Vol. **28**, AIP Conference Proceedings, Vol. **1096**, 1769-1776 (2009).
10. Y. Pawitan, *In All Likelihood: Statistical Modelling and Inference Using Likelihood*, Oxford University Press, New York (2001).