7-2010

Physical Model Assisted Probability of Detection in Nondestructive Evaluation

M. Li
Iowa State University

William Q. Meeker
Iowa State University, wqmeeker@iastate.edu

R. Bruce Thompson
Iowa State University

Physical Model Assisted Probability of Detection in Nondestructive Evaluation

Abstract
Nondestructive evaluation is used widely in many engineering and industrial areas to detect defects or flaws such as cracks inside parts or structures during manufacturing or for products in service. The standard statistical model is a simple empirical linear regression between the (possibly transformed) signal response variables and the (possibly transformed) explanatory variables. For some applications, such a simple empirical approach is inadequate. An important alternative approach is to use knowledge of the physics of the inspection process to provide information about the underlying relationship between the response and explanatory variables. Use of such knowledge can greatly increase the power and accuracy of the statistical analysis and enable, when needed, proper extrapolation outside the range of the observed explanatory variables. This paper describes a set of physical model-assisted analyses to study the capability of two different ultrasonic testing inspection methods to detect synthetic hard alpha inclusion and flat-bottom hole defects in a titanium forging disk.

Keywords
cracks, inclusions, ultrasonic applications, titanium alloys, nondestructive evaluation, QNDE, statistics

Disciplines
Materials Science and Engineering | Statistics and Probability

This article appeared in *AIP Conference Proceedings* 1335 (2011): 1541–1548 and may be found at http://dx.doi.org/10.1063/1.3592113.

PHYSICAL MODEL ASSISTED PROBABILITY OF DETECTION IN NONDESTRUCTIVE EVALUATION

M. Li, W. Q. Meeker, and R. B. Thompson

Citation: AIP Conf. Proc. **1335**, 1541 (2011); doi: 10.1063/1.3592113
View online: http://dx.doi.org/10.1063/1.3592113
View Table of Contents: http://proceedings.aip.org/dbt/dbt.jsp?KEY=APCPCS&Volume=1335&Issue=1
Published by the American Institute of Physics.

PHYSICAL MODEL ASSISTED PROBABILITY OF DETECTION IN NONDESTRUCTIVE EVALUATION

M. Li[1], W. Q. Meeker[1], and R. B. Thompson[2]

[1]Center for Nondestructive Evaluation and Department of Statistics, Iowa State University, Ames, IA 50011
[2]Center for Nondestructive Evaluation, Iowa State University, Ames, IA 50011

ABSTRACT. Nondestructive evaluation is used widely in many engineering and industrial areas to detect defects or flaws such as cracks inside parts or structures during manufacturing or for products in service. The standard statistical model is a simple empirical linear regression between the (possibly transformed) signal response variables and the (possibly transformed) explanatory variables. For some applications, such a simple empirical approach is inadequate. An important alternative approach is to use knowledge of the physics of the inspection process to provide information about the underlying relationship between the response and explanatory variables. Use of such knowledge can greatly increase the power and accuracy of the statistical analysis and enable, when needed, proper extrapolation outside the range of the observed explanatory variables. This paper describes a set of physical model-assisted analyses to study the capability of two different ultrasonic testing inspection methods to detect synthetic hard alpha inclusion and flat-bottom hole defects in a titanium forging disk.

Keywords: POD, Bayesian, Extrapolation, Hard alpha inclusion, Censored Data
PACS: 43.60.UV, 43.60.Cg, 81.70.Cv, 02.50.Sk

INTRODUCTION

Background

Nondestructive evaluation (NDE) is used to characterize the status or properties of components or structures without causing any permanent physical damage. The aerospace industry is one important NDE application area where failing to detect defects inside airplane components can lead to disasters [1]. In virtually all NDE applications, there are random effects and errors involved in the measurements and statistical models are needed to analyze the NDE data sets. MIL-HDBK-1823A [2] describes the standard statistical approaches used in NDE studies. Given a sufficient amount of data over an appropriate region of interest for the explanatory variables (e.g. flaw size and depth), simple empirical statistical models are often adequate to describe the relationship between the response and the explanatory variables. In many applications, however, including the one that motivated this research, the available data are not sufficient to address the questions that need to be

Review of Progress in Quantitative Nondestructive Evaluation, Volume 30
AIP Conf. Proc. 1335, 1541-1548 (2011); doi: 10.1063/1.3592113

answered. Under such circumstances, a physics-based statistical model can sometimes be used to extract the needed information from the limited data. In addition, the physics-based model provides a basis for extrapolation outside the range of the available data.

Motivation and Overview

Hard alpha inclusions in titanium alloy aircraft engine disks can lead to serious accidents. A hard alpha inclusion is a brittle nitrogen-based contamination that could cause fatigue cracks to grow more rapidly than would be otherwise expected in the usually ductile titanium alloy. To develop better NDE tools for detection of hard alpha inclusions, a synthetic inclusion forging disk (known as the SID) was fabricated (details are given in [3]). The SID contains numerous types of synthetic hard alpha (SHA) inclusions and flat bottom holes (FBHs) of different known sizes. For each inclusion type, there are multiple copies which we refer to as "targets." These targets are under different surfaces and at different depths in the SID.

This paper describes a round-robin experiment in which the SID was inspected by two different ultrasonic testing (UT) methods, with different operators at different locations. We describe the modeling and statistical analyses that were used to estimate the probability of detection (POD) for the synthetic hard alpha inclusions and provide the needed extensions to standard methods that have been used traditionally in the analysis of NDE data. Our modeling and analysis include the use of a physics-based model to describe the relationship between NDE signals and flaw characteristics and the use of a mixed effect model to describe random effects in the inspection process. We also introduce the important concept of making inferences on a quantile of the POD distribution.

DATA DESCRIPTION

The titanium SID that was used in the experiments described in this paper contained a large number of cylindrical FBH and SHA targets. For the FBH targets, there were three sizes: #1, #3 and #5 (corresponding to 1/64, 3/64 and 5/64 inches in diameter, respectively). For the SHA targets, there were only two different sizes: #3 and #5. The SHA targets had two different weight percent nitrogen concentrations $\left(N_w\right)$ for each size: 3% and 17%. Thus there were seven different target types. We denote these by #1FBH, #3FBH, #5FBH, #3SHA3, #3SHA17, #5SHA3, and #5SHA17.

The SID was inspected with two different UT inspection methods which are commonly known as the Conventional method and the Multizone method. Both methods have software depth compensation such that the measurement response has little or no dependency on the depth of a target. The UT response from each measurement within an inspection was a voltage that was, for purposes of statistical analysis, converted, through a scale change, to an Effective Flat Bottom Hole (EFBH) response. This kind of standardized response is often used when it is necessary to combine data with differences in calibration level. For the Multizone method, which uses a signal-to-noise ratio detection criterion in addition to the amplitude criterion, there were additional noise measurements also converted to EFBH units. The measurement data are shown at Fig. 1.

PHYSICAL MODEL

Physical models are discussed in detail to describe the principles behind the UT responses for defects with different composition and various sizes in the paper by Thompson, Meeker and Brasche [4]. In this paper we will summarize the key results used

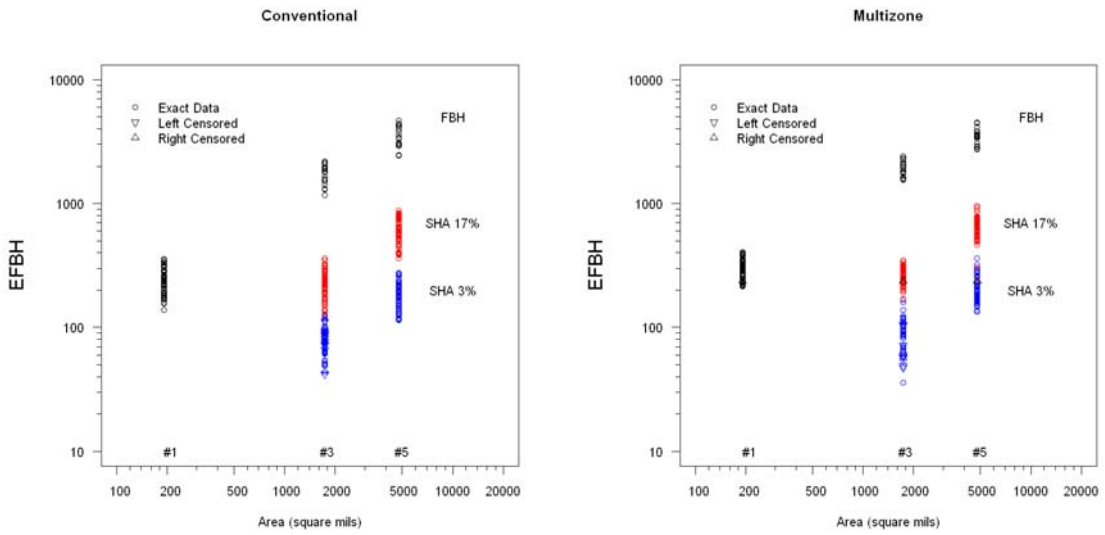

FIGURE 1. A splot of the amplitude data from the Conventional (left) and Multizone (right) inspections.

in the statistical analysis. There are several regimes of scattering determined by the relative values of the target radius b, the ultrasonic wavelength λ, and the beam radius w. As the flaw size grows from very small to very large, one will respectively pass through the following regimes:

- Rayleigh limit: if $b << \lambda$, the signal is proportional to b^3.
- Modified Born approximation: if $b < \lambda$, transition from the Rayleigh limit to the Kirchhoff regime, the signal has a complex pattern depending on the spectrum of the ultrasonic pulse.
- Kirchhoff regime without beam limiting: if $\lambda < b < w$, the signal is proportional to b^2.
- Kirchhoff regime with beam limiting: if $\lambda < w \ll b$, the signal is independent of b.

In this work, experimental measurement and the sizes of the SHA and FBH targets in the SID fall within the Kirchhoff regime. Thus in the following statistical modeling, only the Kirchhoff approximation is used:

$$\text{EFBH} = \left|R(\beta)\right| \frac{\pi w^2}{2}\left(1 - e^{-2(b/w)^2}\right) \qquad (1)$$

where $R(\beta)$ is the reflectance factor. The reflectance factor is one for FBH target and, as described in [4], is a function of Nitrogen concentration for SHA target with β being a tuning parameter to account for diffusion of some nitrogen into the titanium alloy matrix during the HIPping process when making the SID.

STATISTICAL MODEL

Mean Response

The targets in the SIDs fall into the Kirchhoff regime and the physical response function is written in (1) in units of EFBH. By adding a fitting parameter and taking a log transformation of (1) we have

$$\log_{10}\left[\mathrm{EFBH}(x)\right] = \log_{10}(\alpha) + \log_{10}\left[\left|R(\beta)\right|\frac{\pi w^2}{2}\left(1-e^{-2(x/w)^2}\right)\right] \quad (2)$$

where α is the scaling fitting parameter that accounts for the overall factor of the Kirchhoff approximation, $R(\beta)$ is the reflectance factor, w is the beam radius, and x is the target radius. The beam radius (w) is to be estimated from the data and the target radius (x) is in units of mils (a mil is .001 inch).

Random Effects

At each inspection location, there were several operators, each of whom inspected the entire disk. There were operator-to-operator variations in the measurement responses even for the same target. There were also target-to-target variations, probably due to variability in the SID fabrication processes and spatial variability in materials properties throughout the SID. To account for these variations, we assumed a random operator effect and a random target effect in addition to the measurement error. We also assume that any differences from site-to-site were due primarily to differences among the operators. To account for these random effects, the physical model in (2) was extended as follows:

$$\log_{10}\left(\mathrm{EFBH}(x)\right) = \log_{10}(\alpha) + \log_{10}\left(R(\beta)\bullet\frac{\pi}{2}w^2\left(1-e^{-2(x/w)^2}\right)\right) + \tau + \gamma + \varepsilon$$
$$\text{with} \quad \tau \sim N\left(0,\sigma_\tau^2\right),\ \gamma \sim N\left(0,\sigma_\gamma^2\right),\ \varepsilon \sim N\left(0,\sigma_\varepsilon^2\right) \quad (3)$$

where τ, γ and ε are the corresponding operator random effect, target random effect and measurement error, respectively. We assume a normal distribution with mean zero for the operator random effect, the target random effect, and the measurement error. The variances for operator random effect, target random effect and measurement error are σ_τ^2, σ_γ^2 and σ_ε^2, respectively. Thus, in addition to the three parameters (α,β,w) in the physical model in (2), we now have three more variance component parameters to be estimated. To simplify the expression of the statistical model in (3), we define the $\mu_{\log_{10}(\mathrm{EFBH})}(x)$ as

$$\mu_{\log_{10}(\mathrm{EFBH})}(x) = \log_{10}(\alpha) + \log_{10}\left(R(\beta)\bullet\frac{\pi}{2}w^2\left(1-e^{-2(x/w)^2}\right)\right). \quad (4)$$

Then the statistical model can be expressed as $Y(x) = \log_{10}\left(\mathrm{EFBH}(x)\right) = \mu_{\log_{10}(\mathrm{EFBH})}(x) + \tau + \gamma + \varepsilon$. By defining the total variance as $\sigma_{\text{total}}^2 = \sigma_\tau^2 + \sigma_\gamma^2 + \sigma_\varepsilon^2$, we can write the log response function in terms of a normal distribution as

$$Y(x) \sim N\left(\mu_{\log_{10}(\mathrm{EFBH})}(x), \sigma_{\text{total}}^2\right). \quad (5)$$

BAYESIAN ESTIMATION

Estimation Model Parameter

The features in our statistical model and data involve a non-linear response function from the physical model, left and right censored data, random effects, and a need to provide point estimate and bounds to reflect statistical uncertainty. Likelihood based methods could be use to handle all the above needs and data/model features. No commercial software, however, exists to do such an analysis, and developing such software was not feasible within the timing constraints of our funding sponsor. Bayesian methods provide a useful alternative method of analysis. It is well known that with flat

prior distributions the joint posterior distribution is proportional to the likelihood function. Thus with a moderately large amount of data, and diffuse prior distributions, Bayesian methods will produce inferences on functions of the parameters that are similar to what would be obtained by using likelihood-based methods. Furthermore, the software package WinBUGs is flexible enough (with just a little programming being needed) to handle the data/model features needed for the analysis of the SID data.

In our Bayesian analysis, a Markov Chain Monte Carlo (MCMC) algorithm is used, through WinBUGs, to generate a large number of sampling draws from the joint posterior distribution of the model parameters. After the MCMC algorithm has converged, we have M sampling draws for each model parameter. These M sampling draws are samples from the joint posterior distribution of the parameters. These can in turn be used to compute statistics of interest such as mean, standard deviation, median, 2.5% and 97.5% quantiles of the posterior distribution for each model parameter.

Estimation of Mean Response Function

Besides the model parameters, we can also find the posterior distribution for functions of the model parameters. In an inspection process with random effects, the true response function and true POD are random (e.g., in our application there would be a different true, unknown response function for each target/operator combination). The NDE community traditionally focuses on mean quantities in reporting the response function and POD, in effect, averaging over the random effects. We refer to these averages as the mean response function and the mean POD function, respectively. The MCMC samples for mean response function are obtained through $\mu_{\log_{10}(\mathrm{EFBH})}(x)$ with the MCMC sample draws of model parameters.

Figure 2 shows the mean response functions (solid lines) with 95% lower credible bounds (LCBs) (dashed lines) versus target areas for each target type with Conventional results on the left and Multizone results on the right. The amplitude detection criterion is the same for Conventional and Multizone inspections and is indicated as horizontal solid lines at these plots. The Multizone inspection uses, in addition, a signal-to-noise ratio criterion.

Estimation of Quantile Response Function

In some applications, however, there is interest in the worst case scenario among the population of operators and targets. Under such cases a small quantile of response function distribution and a small quantile of POD function distribution for operator and target random effects would be more appropriate metrics to report. Consider a random draw of an operator $\tau \sim N\left(0,\sigma_\tau^2\right)$ and a target $\gamma \sim N\left(0,\sigma_\gamma^2\right)$. Important functions of these random effects such as $\log_{10}\left(\mathrm{EFBH}(x)\right)|_{\tau,\gamma}$ and $\mathrm{POD}(x)|_{\tau,\gamma}$, where x is the target area, will have their own distributions. The mean response for a particular operator and target (averaging over measurement error) can be described by the random variable

$$Y(x)|_{\tau,\gamma} = \mu_{\log_{10}(\mathrm{EFBH})}(x) + \tau + \gamma. \tag{6}$$

with mean $\mu_Y(x) = \mu_{\log_{10}(\mathrm{EFBH})}(x)$ and variance $\sigma_Y^2 = \sigma_\tau^2 + \sigma_\gamma^2$. The p quantile of $Y(x)|_{\tau,\gamma}$ is $y_p(x) = \mu_Y(x) + z_p\sigma_Y$ where z_p is the standard normal p quantile. Here $\varepsilon \sim N\left(0,\sigma_\varepsilon^2\right)$ is the consolidation of all other variations in the measurement after a particular operator and target are selected.

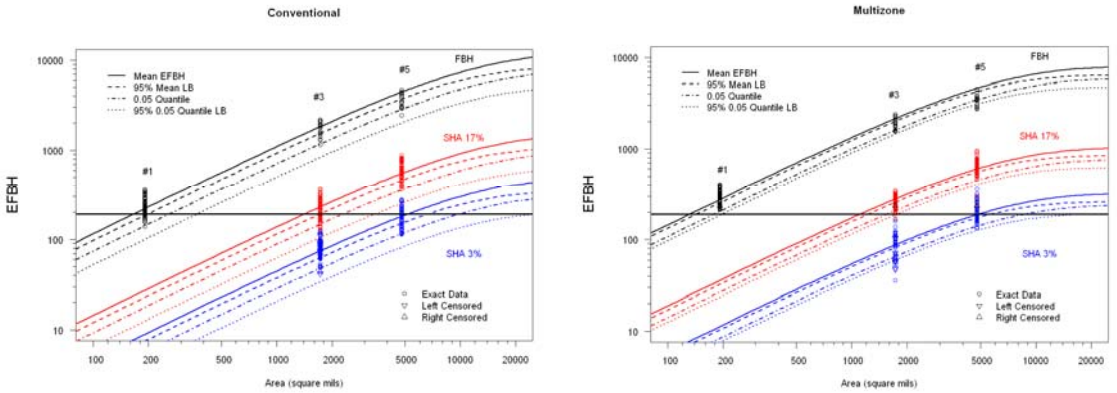

FIGURE 2. Estimates of the mean response functions, the 0.05 quantiles of response function distribution and their corresponding 95% LCBs for the Conventional (left) and Multizone (right) inspection methods.

In our examples we focus on the 0.05 quantile of the response function. This quantile can be interpreted as the response value that will be exceeded by 95% of the target and operator combinations from the population of targets and operators. Because $y_{0.05}(x)$ is a function of the model parameters we can estimate its mean and compute a corresponding 95% LCB by using the sampling draws of $y_{0.05}(x) = \mu_Y(x) + z_{0.05}\sigma_Y$. Figure 2 shows the mean of the 0.05 quantiles of the response function distribution (dashed-dotted lines) and their LCBs (dotted lines) versus target areas for the 3% SHA, 17% SHA, and FBH targets respectively with Conventional results on the left and Multizone results on the right. Compared to the tight 95% LCBs on the mean response function, the LCBs for the 0.05 quantiles of the response function distribution are further away from the estimate of the response quantile. One reason for this is that uncertainty in the quantile involves uncertainty in both the parameters of (4) and in the variance components being estimated.

PROBABILTY OF DETECTION

Conventional POD

For the Conventional method, the detection threshold is set as $y_{th} = \log_{10}(191.75) = 2.2827$, where 191.75 is the area of a #1 FBH in units of square mils. Sensitivity to a #1 FBH was the inspection sensitivity agreed upon by jet engine manufacturers and the Federal Aviation Administration. POD can be found by computing $\mathrm{POD}(x) = \Pr(Y(x) > y_{th})$ where x is the target area and the random variable $Y(x)$ is defined in (5). Specifically, the $\mathrm{POD}(x)$ is evaluated as follows:

$$\mathrm{POD}(x) = \Pr(Y(x) > y_{th}) = \Phi\left(\frac{\mu_{\log_{10}(\mathrm{EFBH})}(x) - y_{th}}{\sigma_{\mathrm{total}}}\right) \qquad (7)$$

where $\Phi(x)$ is the standard normal cumulative distribution function and $\mu_{\log_{10}(\mathrm{EFBH})}(x)$ is defined in (4). With the sampling draws of the $\mu_{\log_{10}(\mathrm{EFBH})}(x)$ and σ_{total}, we can compute the corresponding sampling draws of $\mathrm{POD}(x)$. Estimates of the mean POD and a corresponding 95% LCB can be found by computing the sampling draws of $\mathrm{POD}(x)$ over a range of x values.

An expression for the p quantile of the POD distribution for the Conventional method is obtained by replacing $\mu_{\log_{10}(\mathrm{EFBH})}(x)$ with $y_p(x)=\mu_Y(x)+z_p\sigma_Y$ and replacing $\sigma^2_{\mathrm{total}}$ with σ^2_ε in (7). In particular, the p quantile of the POD distribution for target size x is

$$\left[\mathrm{POD}(x)\right]_p=\Phi\left(\frac{y_p(x)-y_{th}}{\sigma_\varepsilon}\right) \tag{8}$$

Multizone POD

The Multizone inspection method uses a signal-to-noise ratio (SNR) detection rule in addition to the amplitude detection criterion used in the Conventional method. Nieters et al. [5] used $\mathrm{SNR}=(Y-N_a)/(N_p-N_a)$ to define SNR. Here Y is the UT signal measurement, N_a is the noise average, and N_p is the noise peak in a defined rectangular region with the rectangle containing the target signal cut out. The industry standard detection criterion for SNR detection in Multizone inspection is $\mathrm{SNR}>2.5$. Then the SNR criterion is equivalent to $Y>N_{th}\equiv 2.5N_p-1.5N_a$ where N_{th} is defined as the noise threshold. Instead of modeling the SNR, it is easier to estimate the signal distribution and the noise-threshold distribution directly. The noise threshold varies from target to target and from disk to disk and can be computed from the Multizone experimental results. The variability in the noise threshold data can be described by a normal distribution $N_{\mathrm{th}}\sim N(\mu_{\mathrm{noise}},\sigma^2_{\mathrm{noise}})$.

The Multizone POD is composed from two parts: the amplitude detection POD and POD contribution from the noise threshold detection. That is, we need to compute the probability of $Y(x)\le N_{\mathrm{th}}$ or $Y(x)>y_{\mathrm{th}}$. This probability can be expressed as

$$\mathrm{POD}(x)=\Phi\left(\frac{\mu_{\log_{10}(\mathrm{EFBH})}(x)-\mu_{\mathrm{noise}}}{\sqrt{\sigma^2_{\mathrm{total}}+\sigma^2_{\mathrm{noise}}}}\right)+\int_{y_{th}}^{\infty}dn_{th}\int_{y_{th}}^{n_{th}}f(y,n_{th})\,dy \tag{9}$$

where $f(y,n_{th})=\phi\left(y,\mu_{\log_{10}(\mathrm{EFBH})},\sigma^2_{\mathrm{total}}\right)\phi\left(n_{th},\mu_{\mathrm{noise}},\sigma^2_{\mathrm{noise}}\right)$ and $\phi(x,\mu,\sigma^2)$ is the normal density function with mean μ and variance σ^2.

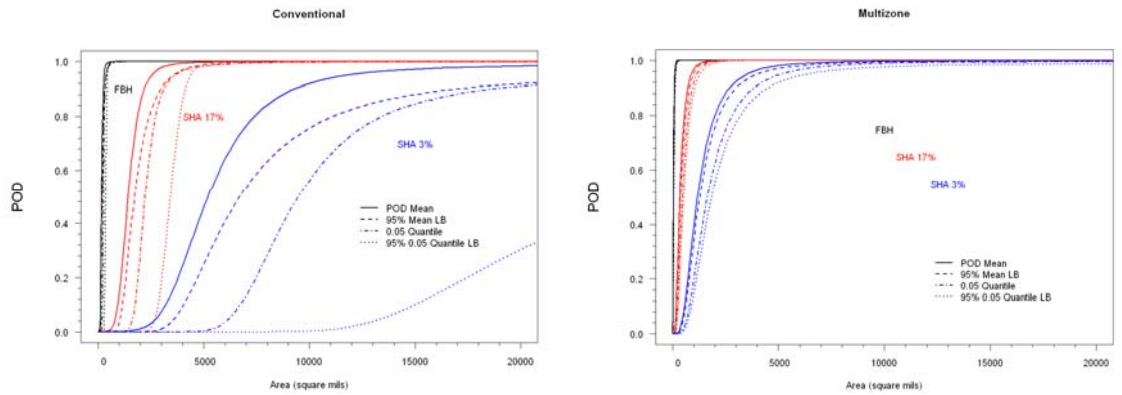

FIGURE 3. Estimates of the mean PODs, the 0.05 quantiles of POD distribution and their corresponding 95% LCBs for the Conventional (left) and Multizone (right) inspection methods.

Similar to the quantile of POD distribution for the Conventional method, by replacing $\mu_{\log_{10}(\mathrm{EFBH})}(x)$ with $y_p(x)$ and replacing $\sigma^2_{\mathrm{total}}$ with σ^2_ε in (9) we can get the Multizone p quantile of POD distribution for any target size x. The mean and 0.05 quantile POD for both Conventional and Multizone method are shown at Fig. 3.

CONCLUSIONS

In this paper, we described the establishment and application of a statistical model for quantifying inspection capability and estimating POD, based on the physical mechanisms of an ultrasonic testing process. The physics-based statistical model enabled needed information extraction from data taken on the limited types and sizes of the synthetic inclusion targets in the synthetic inclusion titanium disk that was available for the experiment. The physics-based model further made possible the needed interpolation and extrapolation for a wider range of flaw sizes and nitrogen concentrations. The nonlinear response function, random effects, and the censored observations were accommodated in the statistical part of the physics-based model. The Markov Chain Monte Carlo based Bayesian software WinBUGs was utilized with a diffuse prior distribution for estimation of the model-tuning parameters. The mean and 0.05 quantile of the response functions and the POD curves for a representative set of target areas and target types were presented. The results from this study provide useful information about the ability to detect hard alpha inclusions in titanium forgings. The methodology provided here is, however, more general and could be used to study NDE inspection capability in other areas of application and for other kinds of inspection.

ACKNOWLEDGEMENTS

This material is based upon work supported by the Federal Aviation Administration under Contract Number DTFACT-08-C-00005 and performed at Iowa State University's Center for NDE. The experimental results presented in this paper are based on the efforts of the aforementioned team. However, as a final report has not been completed and approved by all, these results should be considered to represent the current views of the authors of this paper and not an endorsement by the total team or the funding agency.

REFERENCES

1. NTSB/AAR-90/06 (1990), *Aircraft Accident Report for United Airline Flight 232*, National Transportation Safety Board, 800 Independence Avenue S.W., Washington, D.C. 20594.
2. MIL-HDBK-1823A (2009), *Nondestructive Evaluation System Reliability Assessment*, Standardization Order Desk, 700 Roberts Avenue, Philadelphia, PA 19111.
3. DOT/FAA/AR-05/46 (2007), *Inspection Developments for Titanium Forgings*, Air Traffic Organization Operations Planning Office of Aviation Research and Development, Washington, DC 20591.
4. R. Thompson, W. Meeker, and L. Brasche (2010), "Determination of the POD of Synthetic Hard-Alpha Inclusions in Titanium Forgings for Aircraft Engines", in this proceedings.
5. E. Nieters, R. Gilmore, R. Trzaskos, J. Young, D. Copley, P. Howard, M. Keller, and W. Leach, "A Multizone Technique for Billet Inspection," *Review of Progress in Quantitative Nondestructive Evaluation*, **14**, 2137-2144 (1995).