Physical Model Assisted Probability of Detection in Nondestructive Evaluation for Detecting of Flaws in Titanium Forgings

Ming Li and William Q. Meeker

Department of Statistics and Center for Nondestructive Evaluation,

Iowa State University, Ames, IA 50011

R. Bruce Thompson

Department of Aerospace Engineering and 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 that need to be inspected while in service. The commonly-used standard statistical model for such data is a simple empirical linear regression between the (possibly transformed) signal response variables and the (possibly transformed) explanatory variable(s). 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 the explanatory variable or 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 defects in titanium forging disks.

Key Words: Bayesian analysis, Censored data, Extrapolation, Hard alpha inclusion, Kirchhoff approximation, Mixed effects, Titanium forging, Ultrasonic testing.

1 INTRODUCTION

1.1 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 [see for example NTSB/AAR-89/03 (1989) and NTSB/AAR-90/06 (1990)]. 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 (2009) 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 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 enables us to extrapolate outside the range of the available data.

As exemplified in NTSB/AAR-90/06 (1990), 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 what 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 Margetan et al. 2007). 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.

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.

1.2 Related Literature

Olin and Meeker (1996) and Spencer (1996) provided an overview of statistical methods for NDE techniques. MIL-HDBK-1823A (2009) described the standard statistical procedures for NDE data analyses and Annis (2009) provided an R package to implement these procedures through maximum likelihood (ML) method.

1.3 Overview

The rest of this paper is organized as follows. Section 2 presents the standard statistical methods used in NDE and the concept of POD. Section 3 gives a summary description of the experimental data. Section 4 describes the details of the physical models used in the analyses. Section 5 presents the physics-based statistical model. Section 6 describes the estimation procedures of the statistical model. Section 7 presents

detailed POD results for different types of defects. Section 8 contains some concluding remarks and extensions for future research work.

2 STANDARD STATISTICAL METHODS IN NONDESTRUCTIVE EVALUATION

In this section, we outline the standard statistical methods and procedures that are commonly used in NDE applications, as described at MIL-HDBK-1823A (2009). There are two types of responses in NDE applications: hit and miss binary responses and continuous responses such as voltage. Given the fact that the UT measurements from the titanium forging SID are continuous, we focus on the statistical model for a continuous response.

2.1 Statistical Models for NDE

We use Y to denote the NDE measurement response (or its transformation) and x to denote the defect size (or its transformation). Other explanatory variables (or their transformation), some of which might be random effects, are denoted by a vector $\mathbf{z}$. Then the statistical model is $Y = f(x, \mathbf{z}, \boldsymbol{\beta}) + \varepsilon$ where $\boldsymbol{\beta}$ is a vector of regression parameters and ε is the measurement error following a normal distribution $N(0, \sigma_y^2)$. With the measurement data (possibly censored or truncated) and specified $f(x, \mathbf{z}, \boldsymbol{\beta})$, estimates of the parameter vector $(\hat{\boldsymbol{\beta}}, \hat{\sigma}_y^2)$ and the estimated variance covariance matrix of these estimates can be obtained through standard ML methods described, for example, in Pawitan (2001). MIL-HDBK-1823A (2009) discussed the commonly used simplest case with $Y = \beta_0 + \beta_1 x + \varepsilon$ and Annis (2009) provided an R package based on the ML method with censored observations for this and more general linear regression models. It is common to use a normal distribution to describe the variability in ε, although it is possible to use alternative appropriate distributions when needed.

2.2 Detection Threshold

For specimens without any defects there are still measurement responses due to background noise and other measurement variations. We use Y_n to denote the resulting noise response (or its transformation) Often the noise (generally using the same transformation as the response) can be modeled adequately with a normal distribution of $Y_n \sim N\left(\mu_n, \sigma_n^2\right)$. NDE noise data can be obtained by taking measurements on units without flaws or from those parts of a unit not containing flaws. These data can then be used to compute ML estimates $\left(\hat{\mu}_n, \hat{\sigma}_n^2\right)$ of the noise parameter. The detection threshold $\left(y_{th}\right)$ is typically set to provide an acceptably small probability (p_f) of a false alarm (e.g., $p_f = 0.01$ or 0.05). In particular, the detection threshold can be chosen such that $\Pr\left(Y_n > y_{th}\right) = p_f$. Specifically, the detection threshold is then chosen as $y_{th} = \hat{\mu}_n + \hat{\sigma}_n \Phi^{-1}\left(1 - p_f\right)$ where $\Phi^{-1}\left(x\right)$ is the standard normal distribution quantile function.

2.3 Probability of Detection

For a specified model $f\left(x, \mathbf{z}, \boldsymbol{\beta}\right)$ and detection threshold, the probability of detection as function of defect size can be obtained as follows:

$$\mathrm{POD}\left(x\right) = \Pr\left(Y > y_{th}\right) = 1 - \Phi\left(\frac{y_{th} - f\left(x, \mathbf{z_0}, \boldsymbol{\beta}\right)}{\sigma_y}\right) \tag{1}$$

where $\mathbf{z_0}$ is a set of fixed explanatory variables and $\Phi\left(x\right)$ is the standard normal cumulative distribution function. Confidence bounds for the POD can be obtained by using delta method (see for example Appendix B in Meeker and Escobar 1998), requiring as inputs, the estimated variance and covariance matrix of the parameter estimates $\left(\hat{\boldsymbol{\beta}}, \hat{\sigma}_y^2\right)$.

3 DATA DESCRIPTION

3.1 Data Overview

The titanium SID that was used in the experiments described in this paper contained a large number of cylindrical FBH and SHA targets. A cross section diagram of the SID is shown in Figure 1 with longer rods indicating FBH inclusions and shorter rods indicating cylindrical SHA inclusions. 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. Detailed information about the SID can be found in Margetan et al. (2007).

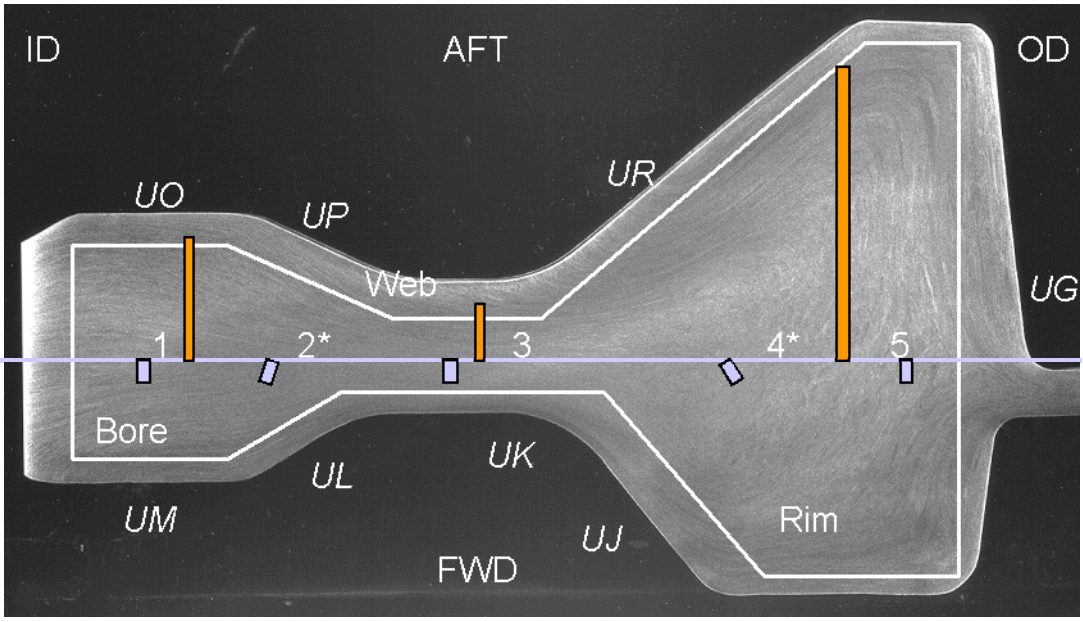

Figure 1. The cross section of the synthetic inclusion disk.

The SID was inspected with two different UT inspection methods which are commonly known as the Conventional method (Figure 2 top) and the Multizone method (Figure 2 bottom). The Conventional

method sets the focal point near the surface of the SID, and the Multizone method uses several transducers simultaneously each of which has a focal point at certain depth of the SID. 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. The EFBH response is defined as the flat bottom hole area that would give a signal response equal to the observed response, assuming a common calibration to certain size FBH. In the case of the SDI experiment the comparison was to a #1 FBH, corresponding to the specified calibration level that was used for all runs of the experiment (i.e., gain was set such that a #1 FBH would have a response that is 80% of a signal that would cause saturation). 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, there were additional noise measurements also converted to EFBH units. Noise data was also acquired in the Conventional inspections and used to define detection limits, so that missed targets could be treated as left-censored observations.

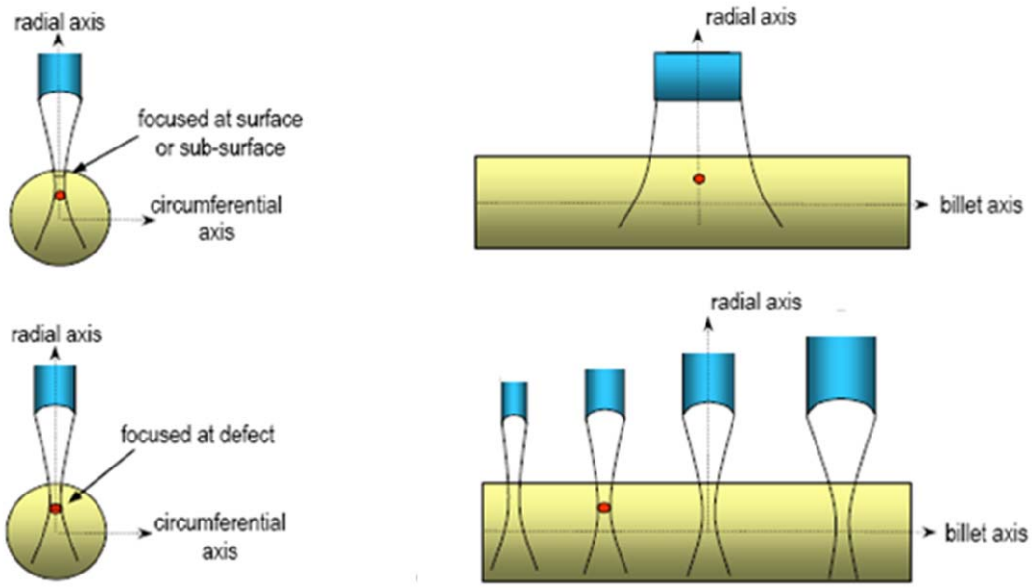

Figure 2. Conceptual illustration of the Conventional inspection system (top) and the Multizone inspection system (bottom) for billet inspection.

For most observations on individual targets within an inspection, we have exact readings that were translated to EFBH. In some of the inspections, however, the signal was below the noise floor and therefore determined to be a "miss." These observations are left censored in that we know only that the actual EFBH response is less than the noise floor EFBH. The noise floor varies from target to target. In some of the Multizone method inspections, the operator did not follow the protocol with respect to saturated observations. The protocol required that, in the case of saturated observations, the operator should reduce the gain in a sequence of steps to a known level where an actual reading could be made. Then this reading could be converted to the actual voltage and corresponding EFBH. When the operator did not follow the protocol, we know only that the EFBH response is larger than the EFBH corresponding to the smallest voltage level that would cause saturation. Figure 3 is a summary plot of the data sets used in the analyses for both the Conventional and the Multizone methods, showing the seven different target types. Because the systems used UT probes that were operating at the same nominal frequency (10 MHz)

and were calibrated in the same manner, it is not surprising that the amplitude values are similar for the two methods.

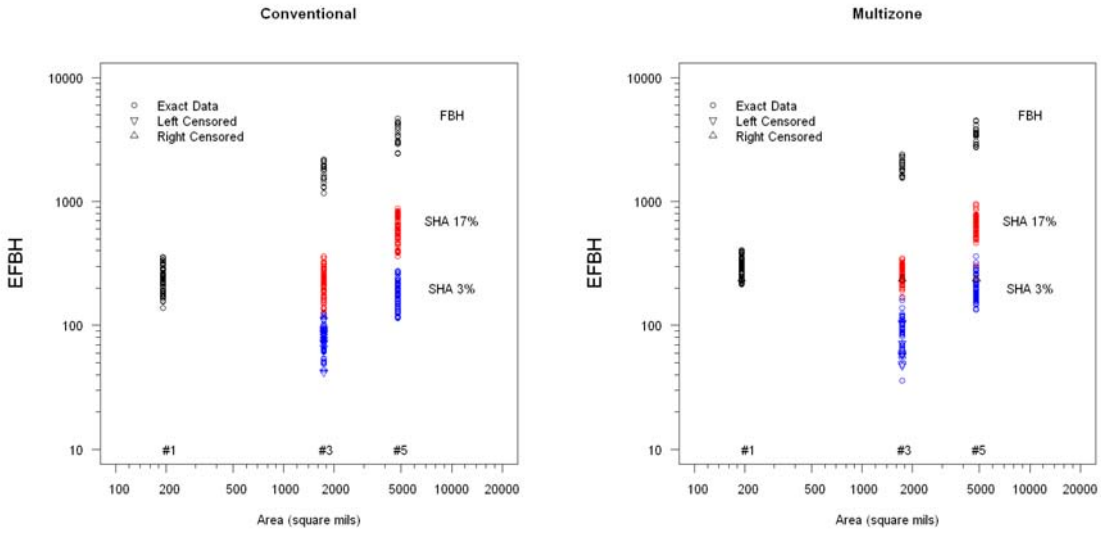

Figure 3. A summary plot of the data from the Conventional (left) and Multizone (right) inspections.

3.2 Operator Plots and Targets Plots

The SID disk was inspected with both the Conventional UT method (two locations, six operators) and the Multizone UT method (three locations, seven operators). Figure 4 (operator plots) shows the EFBH response for target type #5SHA3 plotted versus operator, with one line for each target with Conventional results on the left and Multizone results on the right. These plots show that there is an important amount of operator-to-operator variability (i.e., random operator effect) in the EFBH responses for a given target type. Figure 5 (target plots) shows the EFBH response for target type #5SHA3 plotted versus the individual targets, with one line for each operator, again with the Conventional results on the left and the Multizone results on the right. These plots show that there is an important amount of target-to-target variability (i.e., a random target effect) in the EFBH responses for a given target type. The operator plots and target plots for target types other than #5SHA3 are similar and thus not shown here.

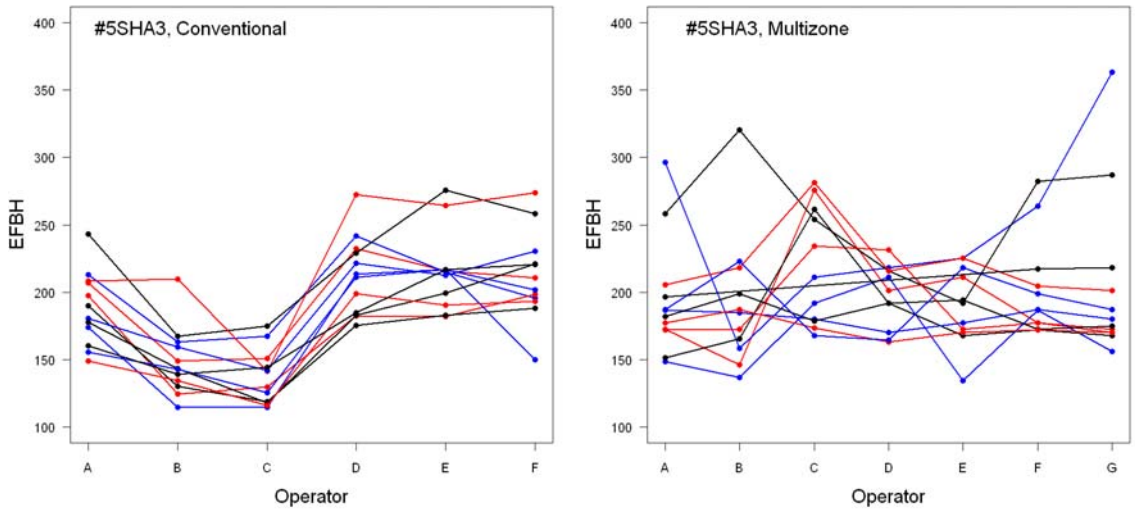

Figure 4. Operator plots for the #5SHA3 targets for Conventional (left) and Multizone (right) inspections, with one path for each target.

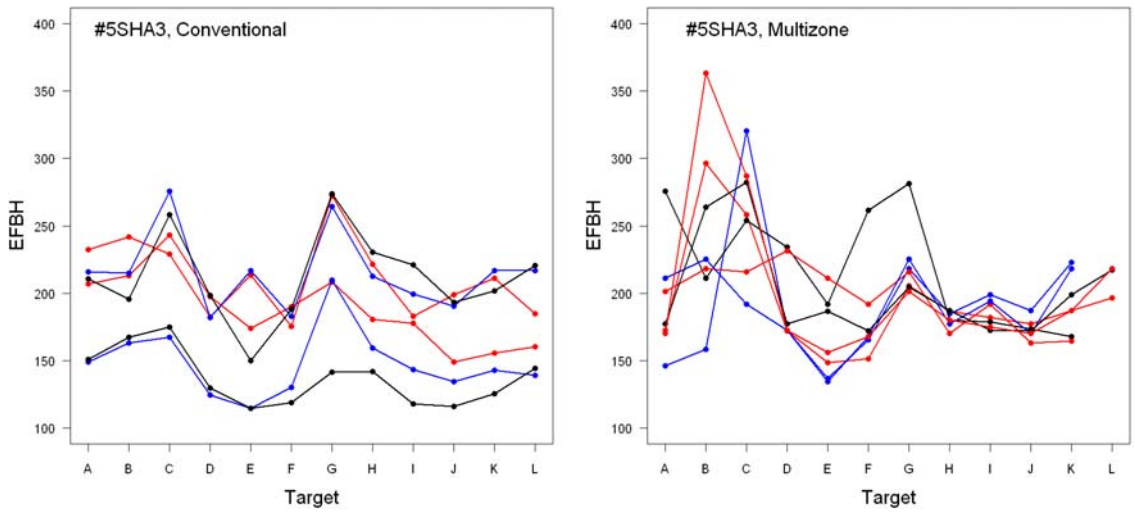

Figure 5. Target plots for the #5SHA3 targets for Conventional (left) and Multizone (right) inspections with one path for each operator.

4 PHYSICAL MODEL DETAIL

A typical UT system includes a pulser, a transducer, and a display screen. Driven by the electrical pulses generated by the pulser, the transducer generates an ultrasonic wave. The ultrasonic wave is coupled into and propagates through the SID being tested. When there is a discontinuity such as one of the SHA or FBH targets in the ultrasonic wave propagation path, part of the energy will be reflected. The reflected energy is then transformed into an electrical signal by the transducer and is shown in the display screen. By analyzing the results at the display screen, the existence of defects (SHA or FBH in the SID study considered here) can be determined, and the location and size of the defects can be further evaluated. In this section, several physical models are discussed to describe the principles behind the UT responses for defects with different composition and various sizes.

4.1 Reflectance Factor

A key characteristic affecting ultrasonic (and other kinds of) reflection from a discontinuity and the resulting signal strength is a function of the material properties on both sides of the discontinuity. The reflectance factor R

$$R = \frac{\rho_i v_i - \rho_m v_m}{\rho_i v_i + \rho_m v_m} \tag{2}$$

is used to describe this characteristic. Here ρ denotes density, v denotes the ultrasonic wave speed and subscripts i and m refer to inclusion and titanium alloy matrix (host material), respectively. Because the density of a flat bottom hole target is essentially zero and the density of the host titanium materials is much larger (i.e., $\rho_i << \rho_m$), it follows that $|R|$ is unity for a FBH. Thus R can be expressed as a function of weight percent nitrogen concentration for a SHA (3% or 17% in this study) target through the coefficients ρ_i and v_i. The effects of SHA nitrogen concentration on the values of ρ_i and v_i in titanium

alloys were studied experimentally by Gigliotti, Gilmore, and Perocchi, (1994). Based on their experiments and analysis, they reported that

$$\begin{aligned} \rho_m &= 4461 \text{ kg/m}^3 \\ v_m &= 6175 \text{ m/s} \\ \rho_i &= \left(4490.9 + 5.03 \times N_{at} - 0.01 \times N_{at}^2\right) \text{ kg/m}^3 \\ v_i &= \left(6002.2 + 61.86 \times N_{at}\right) \text{ m/s} \end{aligned} \tag{3}$$

where N_{at} is the atomic percent nitrogen concentration. The relationship between atomic percent nitrogen concentration and weight percent nitrogen concentration $\left(N_w\right)$ is

$$N_{at} = \frac{342 \times N_w}{100 + 2.42 \times N_w}. \tag{4}$$

Thus R can be used to link the signal responses of the data from the FBH, SHA3, and SHA17 targets and make predictions for intermediate values of weight percent nitrogen.

4.2 Kirchhoff Approximation

4.2.1 General background

When the duration of the incident ultrasonic pulse is sufficiently small with respect to the delay of the back surface echo of the targets, the echoes from the front and back surfaces of the targets can be resolved in time. Under such cases the elastodynamic Kirchhoff approximation (Adler and Achenbach 1980) is appropriate to model the measurement response. With the 10MHz UT system that was used for the Conventional and Multizone inspections, the seven types of target studied in this paper fall in the Kirchhoff regime. Thompson and Lopez (1984) introduced the beam radiation pattern Gaussian approximation concept and concluded the electrical signal $\text{Voltage}\left(\omega\right)$ observed in a pulse-echo experiment for a circular planar surface target can be described by using the following form:

$$\text{Voltage}(\omega) = A(\omega, z)\, R \frac{\pi w^2}{2}\left(1 - e^{-2(b/w)^2}\right) \tag{5}$$

where ω is ultrasonic frequency, z is the propagation distance, b is the circular target radius, and w is the ultrasonic beam radius.

4.2.2 Effective Flat Bottom Hole

In production inspections, calibrations are performed to eliminate the effects of the factor $A(\omega, z)$ in (5) which account for variations in transducer performance and the effects of propagation distance. Especially when there is need to combine data from measurements that are taken under different calibration levels, it is common practice to scale UT data into what is known as an EFBH response. The EFBH response is intended to represent the FBH area that would produce a signal equal to that which was observed from the target. More precisely, the EFBH is defined as

$$\text{EFBH} = \frac{S}{S_c}\frac{\pi}{4}\left(\frac{D_c}{64}\right)^2 = \frac{S}{S_c}\pi b_c^2 \tag{6}$$

where S is the peak defect signal strength [proportional to $\text{Voltage}(\omega)$ but in units of percentage of full screen height], S_C is the peak calibration signal strength (in units of percentage of full screen height), D_c is the diameter of the calibration hole (in units of 1/64 inch diameter), and b_c is the radius of the calibration hole in inches. It is easy to show that, for a FBH with size in the Kirchhoff approximation regime, in the absence of noise, the EFBH would be equal to the area of the FBH, consistent with the intent of the definition. Combining (5) and (6), and assuming calibration to a #1 FBH, and targets in the Kirchhoff approximation regime, the predicted response in units of EFBH would be

$$\text{EFBH} = |R| \frac{\pi w^2}{2}\left(1 - e^{-2(b/w)^2}\right). \tag{7}$$

This is a powerful result in the context of this study because (7) can be used to predict the response of all types of targets in the SID. The size enters through the value of b and the composition (i.e. weight percent nitrogen concentration) through the factor R (which is taken to have a value of 1 for a FBH). This approach allows the data from all of the targets in the SID to be described by a single statistical model, thereby increasing the power of the regression analysis and tightening the confidence bounds. Based on the physical model, valid extrapolation of EFBH values for targets with a radius between #1 and #5 and beyond #5 can be obtained for a range of weight percent nitrogen concentrations in SHAs.

4.2.3 Beam Limiting Kirchhoff Approximation

From (7), we can see the beam limiting effect that arises when the defect size b becomes large compared to the ultrasonic beam size w. When $b >> w$, only a fraction of the defect surface is illuminated and the EFBH response in (7), can be simplified as $\mathrm{EFBH}_{b>>w} = |R| \pi w^2 / 2$. That is, the response is no longer a function of target size b but is only a function of reflectance factor R and beam size w.

4.3 Rayleigh Scattering Regime

The Kirchhoff approximation is appropriate for the targets that are present in the SID under study in this paper. An additional consideration is the Rayleigh scattering regime where the defect size is small with respect to the ultrasonic wavelength. Although none of the targets in the SID fall in the Rayleigh scattering regime, it is necessary to consider the different response mechanism to avoid improper extrapolations of the Kirchhoff model to defect sizes smaller than the size of a #1 target. Huang, Schmerr, and Sedov (2006) developed the modified Born approximation from which the model for EFBH is given by $\mathrm{EFBH}_B = 2\pi b^2 \left(v_i / v_m \right) \left| R \sin\left(2b\omega / v_i \right) \right|$. When the defect radius b is sufficiently small such that $\sin\left(2b\omega / v_i \right) \approx 2b\omega / v_i$, the corresponding EFBH model is given by $\mathrm{EFBH}_R = 4\pi b^3 |R| \omega / v_m$ and we say that the response from a target or defect in this size region is described by the Rayleigh limit regime.

4.4 Physical Model Summary

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 with a complex signal pattern is dependent 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.

5 STATISTICAL MODEL

5.1 Mean Response

Section 4 described the physical models for different size regimes with respect to wave length and beam size. The targets in the SIDs fall into the Kirchhoff regime and the physical response function is written in (7) in units of EFBH. By adding a fitting parameter and taking log transformation of (7) we have:

$$\log_{10}\left[\mathrm{EFBH}\left(x\right)\right] = \log_{10}\left(\alpha\right) + \log_{10}\left[\left|R\right|\frac{\pi w^2}{2}\left(1 - e^{-2(x/w)^2}\right)\right] \tag{8}$$

where α is the scaling fitting parameter that accounts for the overall factor of the Kirchhoff approximation, R 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).

5.2 Weight Percent Nitrogen Concentration Correction

The reflectance factor (R) is equal to 1 for FBH targets and is a function of weight percent nitrogen concentration for SHA targets, as described at Section 4.1. The original weight percent nitrogen concentration (N_w) values for SHA targets were 3% and 17%. However our statistical analysis revealed some systematic lack-of-fit relative to the model (Eq. 8), especially for the 17% SHA nitrogen concentration targets. During fabrication, the SID was sliced like a bagel, the FBH targetss were drilled, and SHAs were inserted into holes for that purpose. Then the two pieces of the SID were HIPped (Hot Isostatic Pressing) together at high temperature and pressure to form of a single disk. Ultrasonic velocity measurements (before fabrication and after the experiment) confirmed that the overall nitrogen concentrations in the center of the SHA targets were still 3% or 17%. Thus it is believed that the HIPing process led to some diffusion of nitrogen into the titanium alloy matrix around the SHA target. This would cause a small gradient in nitrogen concentration, reducing the acoustic impedance and the effective reflectivity. It is interesting to note that naturally occurring hard alpha inclusions are also surrounded by a diffusion zone.

A detailed physical model to describe the diffusion process and the change of the reflectance factor is unavailable and thus we adopt an empirical approach of applying corrections to the original weight percent nitrogen concentration. In this empirical approach, we assume that the SHA target size remains unchanged, but introduce a uniform correction to the nitrogen concentration for the entire target area, even though the concentration remains the same in the body of the SHA and only decreases around the edge. Both experimentally and from physical theory it is known that the amount of diffusion will depend

on the original concentration of nitrogen. Here we assume a typical quadratic correction term to the original weight percent nitrogen concentration as

$$N_{wc} = \left[1 - \beta \times \left(\frac{N_w}{100}\right)^2\right] N_w \tag{9}$$

where β is a parameter to be estimated from the data. Then instead of using the original weight percent nitrogen concentration, the corrected weight percent nitrogen concentration in (9) is now used in (4) as follows

$$N_{at} = \frac{342 \times N_{wc}}{100 + 2.42 \times N_{wc}}. \tag{10}$$

.

5.3 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 (8) was extended as follows:

$$\begin{aligned} &\log_{10}\left(\mathrm{EFBH}\left(x\right)\right) = \log_{10}\left(\alpha\right) + \log_{10}\left(R \bullet \frac{\pi}{2} w^2 \left(1 - e^{-2(x/w)^2}\right)\right) + \tau + \gamma + \varepsilon \\ &\text{with} \quad \tau \sim N\left(0, \sigma_\tau^2\right),\ \gamma \sim N\left(0, \sigma_\gamma^2\right),\ \varepsilon \sim N\left(0, \sigma_\varepsilon^2\right) \end{aligned} \tag{11}$$

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 (8), we now have three more variance component parameters to be estimated.

To simplify the expression of the statistical model in (11), we define the $\mu_{\log_{10}(\text{EFBH})}(x)$ as

$$\mu_{\log_{10}(\text{EFBH})}(x) = \log_{10}(\alpha) + \log_{10}\left(R \bullet \frac{\pi}{2} w^2 \left(1 - e^{-2(x/w)^2}\right)\right). \tag{12}$$

Then the statistical model can be expressed as $Y(x) = \log_{10}(\text{EFBH}(x)) = \mu_{\log_{10}(\text{EFBH})}(x) + \tau + \gamma + \varepsilon$.

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

$$Y(x) \sim N\left(\mu_{\log_{10}(\text{EFBH})}(x), \sigma_{\text{total}}^2\right). \tag{13}$$

6 ESTIMATION

6.1 Estimation of the Model Parameters

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 (e.g., Pawitan 2001) 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 (e.g., Gelman,Carlin, Stern, and Rubin. 2003) 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 (2007) 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. Summary results for all model parameters are shown at Table 1 for both Conventional method and Multizone method.

6.2 Estimation of Functions of Model Parameters

Besides the model parameters, we can also find the posterior distribution for functions of the model parameters. For example the corrected weight percent nitrogen concentration (N_{wc}) is a function of the model parameter β defined in (9). By substituting in the M sampling draws of β into (9) we can get the M sampling draws of N_{wc} for any fixed N_w. We can further get the M sampling draws of reflectance factor (R) based on the sampling draws of N_{wc} through (2), (3) and (10). The posterior mean and standard deviation (in parenthesis) for corrected weight percent nitrogen concentration and reflectance factor are shown at Table 2 for both the Conventional method and the Multizone method. As

mentioned in Section 5.2, we used an empirical quadratic nitrogen concentration correction in our model. We only have two original nitrogen concentrations of 3% and 17%. Although our model allows predictions at any level of percent nitrogen, because of the empirical nature of this correction factor, one would have to be cautious in the use of predictions at other values of percent nitrogen.

Table 1. Posterior mean and standard deviation for all of the model parameters

Model	Conventional Method		Multizone Method	
Parameter	Posterior Mean	Standard Dev.	Posterior Mean	Standard Dev.
α	1.168	0.1441	1.468	0.07689
β	14.77	0.9054	13.69	1.300
w	81.46	9.407	58.69	4.776
σ_τ	0.1016	0.04644	0.03314	0.01274
σ_γ	0.05398	0.005752	0.06920	0.007475
σ_ε	0.05343	0.002183	0.07186	0.002737

Table 2. Posterior mean and standard deviation for the corrected weight percent nitrogen concentration and reflectance factor for 3% and 17% original weight percent nitrogen concentrations

N_w	Conventional Method		Multizone Method	
	N_{wc}	R	N_{wc}	R
3%	2.96 (0.0024)	0.041 (0.00004)	2.96 (0.0035)	0.041 (0.00005)

17%	9.74 (0.445)	0.125 (0.0044)	10.3 (0.638)	0.130 (0.0061)

6.3 Estimation of the Response Function

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 response function for each target/operator combination). The NDE community traditionally focuses on the average 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. 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. In this section we describe the procedures to estimate the mean response function and a quantile of response function distribution. Section 7 describes procedures to estimate the mean POD and a quantile of the POD distribution.

6.3.1 Mean of the response function distribution

As described in Section 6.1, we used the internal MCMC simulation algorithm in WinBUGs (2007) to generate sampling draws for all of the parameters in the statistical model (i.e., α, β, w, σ_τ^2, σ_γ^2 and σ_ε^2). We then used these sampling draws to generate the sampling draws of $\mu_{\log_{10}(\mathrm{EFBH})}(x)$ through (12). Figure 6 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 top and Multizone results on the bottom. 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, as described in Section 7. Also shown in these plots are the

exact, left censored and right censored data points denoted by circles, down triangles and up triangles, respectively.

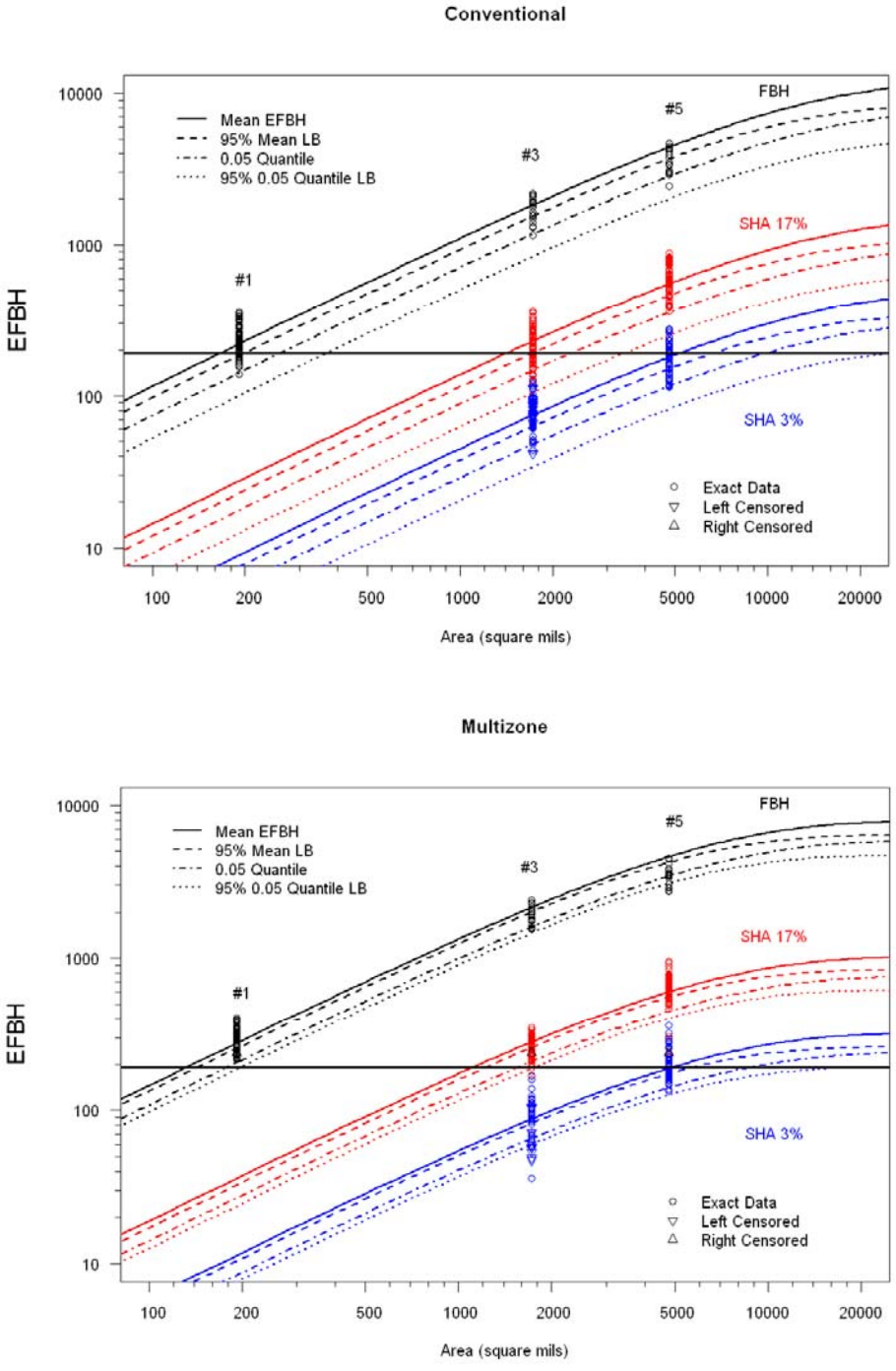

Figure 6. Estimates of the mean response functions, the 0.05 quantiles of response function distribution and their corresponding 95% LCBs for the Conventional (top) and Multizone (bottom) inspection methods.

6.3.2 Quantile of the response function distribution

In many applications it is important to obtain estimates of quantities in the tail of a distribution, as opposed to the mean or other measure of central tendency. For example, in the SID inspection experiment the data tell us that some targets and some operators tend to result in weaker signals than others. 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 $\log_{10}\left(\mathrm{EFBH}\left(x\right)\right)|_{\tau,\gamma}$ and $\mathrm{POD}\left(x\right)|_{\tau,\gamma}$, 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

$$Y\left(x\right)|_{\tau,\gamma} = \mu_{\log_{10}(\mathrm{EFBH})}\left(x\right) + \tau + \gamma. \tag{14}$$

with mean $\mu_Y\left(x\right) = \mu_{\log_{10}(\mathrm{EFBH})}\left(x\right)$ and variance $\sigma_Y^2 = \sigma_\tau^2 + \sigma_\gamma^2$. The p quantile of $Y\left(x\right)|_{\tau,\gamma}$ is $y_p\left(x\right) = \mu_Y\left(x\right) + z_p\sigma_Y$ 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.

In our examples we focus on the 0.05 quantile of the response function. This quantile can be interpreted as the mean response value that will be exceeded by 95% of the target and operator combinations from the population of targets and operators. Because $y_{0.05}\left(x\right)$ is a function of the model parameters we can estimate its mean and compute a corresponding 95% LCB by using the sampling draws of $y_{0.05}\left(x\right) = \mu_Y\left(x\right) + z_{0.05}\sigma_Y$. Figure 6 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 top and Multizone results on the bottom. 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.

6.4 Diagnostics

It is important to assess how well the statistical model fits the experimental data. Figure 7 shows the residuals versus fitted values with Conventional results on the left and Multizone results on the right. The residuals are evenly distributed except for those from the #5FBH targets. The reason for this deviation is that there are relatively few data points for #3FBH and #5FBH and thus these observations are not influential in fitting the model. For #5FBH targets the residuals are below zero which indicates an upward bias in estimation for a #5FBH. This does not raise serious practical concerns because there is little practical interest in predicting POD in the target space region anywhere near to the #5 FBHs. We also compared the results between including #5FBH targets and excluding #5FBH targets when doing the analysis. The results showed little change in the mean response function, and the PODs were more conservative for analysis that includes the #5 FBH targets and thus all the results in this paper are based on analyses that include the #5FBH targets.

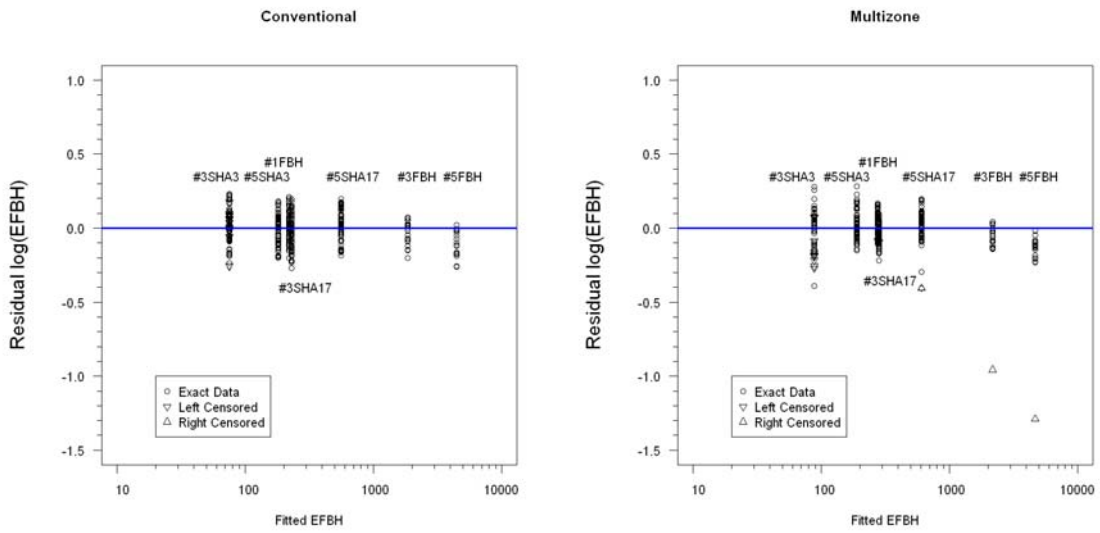

Figure 7. Residual plots as function of fitted value for Conventional (left) and Multizone (right) inspection methods.

7 PROBABILTY OF DETECTION

Given the response function and the detection threshold, the mean POD, the quantile of POD distribution and the corresponding LCBs can be obtained. In this section we first describe the procedures to estimate POD for Conventional and Multizone respectively. Then we present the POD plots of both inspection methods for all types of targets.

7.1 The Conventional Inspection Method

7.1.1 Mean 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 (13). Specifically, the $\mathrm{POD}(x)$ is evaluated as follows:

$$\mathrm{POD}(x) = \Pr(Y(x) > y_{th}) = \Phi\left(\frac{\mu_{\log_{10}(\mathrm{EFBH})}(x) - y_{th}}{\sigma_{\mathrm{total}}}\right) \tag{15}$$

where $\Phi(x)$ is the standard normal cumulative distribution function and $\mu_{\log_{10}(\mathrm{EFBH})}(x)$ is defined in (12). 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.

7.1.2 Quantile of the POD distribution

Again, consider a random draw of an operator and a target. Some combinations will result in higher POD than others. As with the derivation of the quantile of response function distribution in Section 6.3.2, we can take account of this variability by computing a quantile of the POD distribution. 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 (15). In particular, the p quantile of the POD distribution for target size x is

$$\left[\mathrm{POD}(x)\right]_p = \Phi\left(\frac{y_p(x) - y_{th}}{\sigma_\varepsilon}\right). \tag{16}$$

Again, estimates of the 0.05 quantile of the POD distribution and corresponding 95% LCB were obtained by computing the sampling draws of the 0.05 quantile of the $\mathrm{POD}(x)$ distribution for different values of x.

7.2 The Multizone Inspection Method

7.2.1 Mean 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. (1995) used the following definition for SNR and a corresponding detection limit.

$$\text{SNR} = \frac{Y - N_a}{N_p - N_a}$$

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 $\text{SNR} > 2.5$. 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 results of the Multizone experimental results. The variability in the noise threshold data can be described by a normal distribution:

$$N_{\text{th}} \sim N\left(\mu_{\text{noise}}, \sigma^2_{\text{noise}}\right). \tag{17}$$

Figure 8 illustrates this two-dimensional Multizone detection criterion. There is an amplitude detection if the amplitude is above the horizontal line. There is also a SNR detection if the amplitude is above the noise threshold (i.e., if the amplitude/noise threshold point lies above the diagonal line in Figure 8).

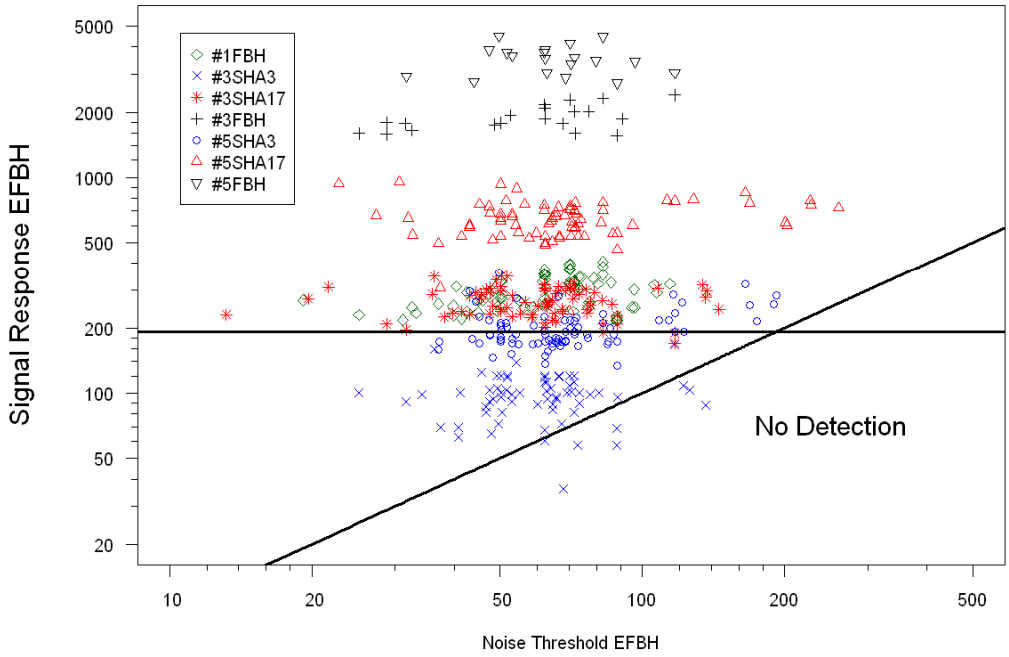

Figure 8. Illustration of the Multizone detection criteria.

The POD for SNR noise threshold detection criterion is:

$$\mathrm{POD}_1(x) = \Pr\left(Y(x) > N_{\mathrm{th}}\right) = \Phi\left(\frac{\mu_{\log_{10}(\mathrm{EFBH})}(x) - \mu_{\mathrm{noise}}}{\sqrt{\sigma^2_{\mathrm{total}} + \sigma^2_{\mathrm{noise}}}}\right)$$

with the random variable $Y(x)$ defined in (13) and N_{th} is defined in (17). Given the independent relationship between $Y(x)$ and N_{th}, the joint density for the response function and noise threshold is:

$$f(y, n_{th}) = \phi\left(y, \mu_{\log_{10}(\mathrm{EFBH})}, \sigma^2_{\mathrm{total}}\right)\phi\left(n_{th}, \mu_{\mathrm{noise}}, \sigma^2_{\mathrm{noise}}\right)$$

with $\phi\left(x, \mu, \sigma^2\right)$ the normal density function with mean μ and variance σ^2. The POD for regions with $Y(x) \le N_{\mathrm{th}}$ but $Y(x) > y_{\mathrm{th}}$ (i.e. the triangle at right edge of Figure 8) is

$$\mathrm{POD}_2(x) = \Pr\left(Y(x) > y_{\mathrm{th}} \text{ and } Y(x) \le N_{\mathrm{th}}\right) = \int_{y_{th}}^{\infty} dn_{th} \int_{y_{th}}^{n_{th}} f(y, n_{th})\, dy$$

which is calculated by numerical integration. Then the Multizone mean $\mathrm{POD}(x)$ is determined as

$$\mathrm{POD}(x) = \mathrm{POD}_1(x) + \mathrm{POD}_2(x). \tag{18}$$

The estimate of the noise mean is $\hat{\mu}_{noise} = 1.7990$. The target-to-target noise variance estimate is $\hat{\sigma}^2_{noise,tt} = 0.02860$. We do not have data that would provide a disk-to-disk noise variance estimate, but, for purposes of illustration, we assume the disk-to-disk variance estimate is $\hat{\sigma}^2_{noise,dd} = 0.5 \times \hat{\sigma}^2_{noise,dd} = 0.01430$. Thus the estimate of total noise variance is $\hat{\sigma}^2_{noise} = 1.5 \times \hat{\sigma}^2_{noise,tt} = 0.04290$. The sampling draws of μ_{noise} and σ^2_{noise} were used to compute the mean POD estimate and the corresponding 95% LCB for Multizone inspection through (18).

7.2.2 Quantile of POD Distribution

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 (18) we can get the Multizone p quantile of POD distribution for any target size x.

7.3 POD Plots

Figure 9 contains plots of the estimates of the mean of the POD distribution (solid lines), the corresponding 95% LCBs (dashed lines), the 0.05 quantile of POD distribution (dashed-dotted lines) and the corresponding 95% LCBs (dotted lines) for 3% SHA, 17% SHA and FBH targets for both inspection methods. Figure 9 shows that although there was little difference between the inspection methods when looking at the signal-response functions estimates in Figure 6, there are large differences between the

estimates of the POD functions. This is due to the important increase in detection power provided by the more complicated SNR detection criterion used in the Multizone inspection method and to some degree because there is less operator-to-operator variability in the Multizone inspection method.

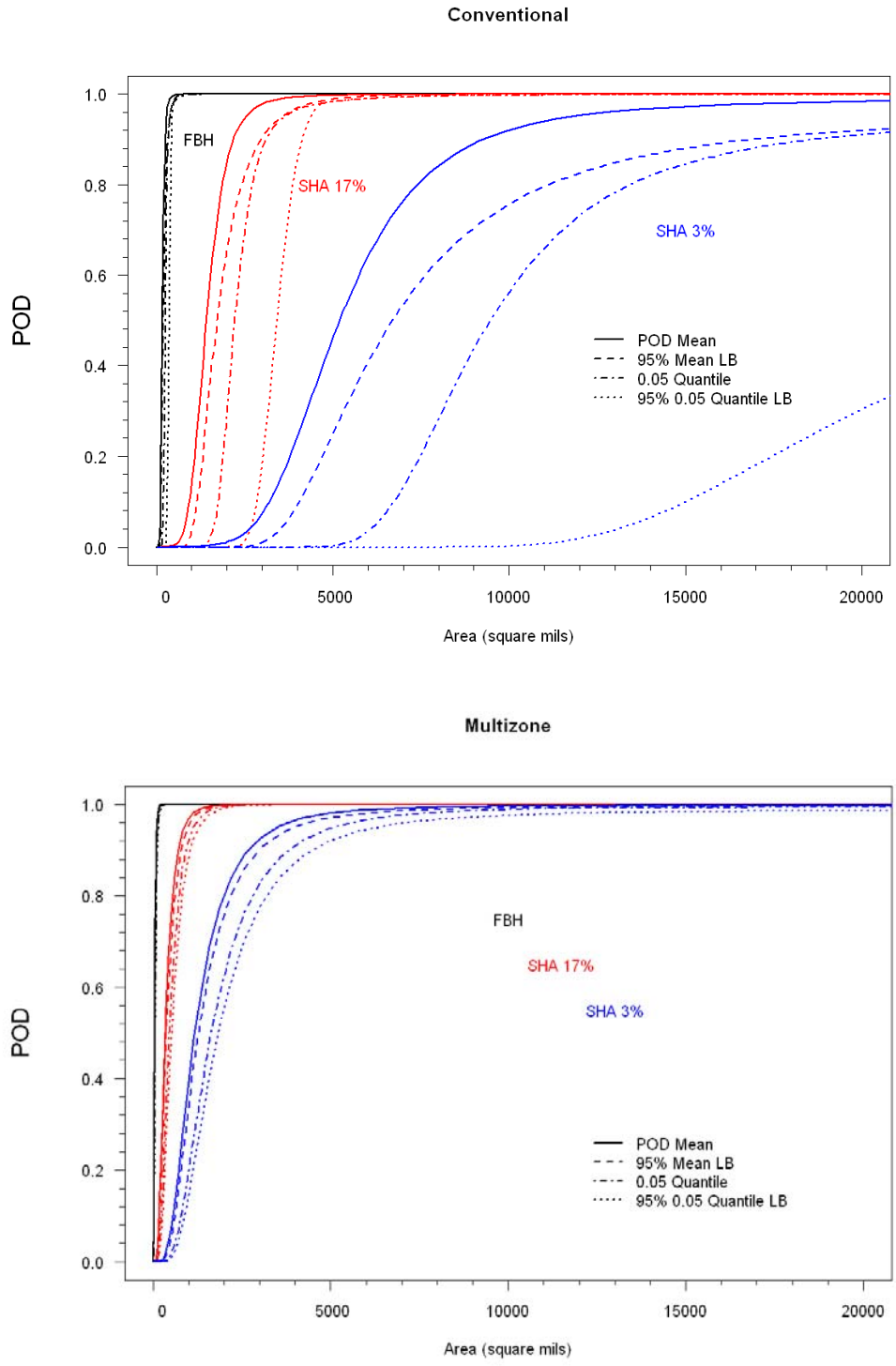

Figure 9. Estimates of the mean PODs, the 0.05 quantiles of POD distribution and their corresponding 95% LCBs for the Conventional (top) and Multizone (bottom) inspection methods.

8 SUMMARY AND CONCLUSION

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.

There are a number of extensions for the methodology presented in this article that suggest future research directions. These include the following:

1. The target sizes and flaw sizes of interest in this study were within the range where the Kirchhoff approximation provides a good description of ultrasonic testing signals. Although not adopted in this study, an explicit extrapolation procedure based on the Rayleigh scattering regime could be developed when needed, allowing extrapolation to smaller flaw sizes.
2. The quadratic term correction for the weight percent nitrogen concentration was used to account for nitrogen diffusion during HIPping process. A correction based on a physical

principle could be implemented if we had more knowledge about the mechanism behind the diffusion arising in the HIPing process.

3. For applications where there is useful prior information about the model parameters (e.g., from previous experience with a particular kind of inspection), a Bayesian analysis with informative prior distributions could be implemented.
4. The current model showed some lack of fit for the #5 FBH condition. Further experimentation on a different inclusion sample with targets having reflectance in the gap between the 17% nitrogen and the flat bottom holes might make it possible to resolve the reasons for this deviation from the physics-based model. Presently, no such sample block is known to exist.

ACKNOWLEDGMENTS

This work was performed at the Center for Nondestructive Evaluation at Iowa State University with support of the Federal Aviation Administration under contract number 08-C-00005. We would also like to thank other members of the project team for their valuable inputs and guidance and for providing details of the inspection methodology that was important when making decisions about data processing and model development. These include Lisa Brasche (Iowa State), Waled Hassan (Rolls-Royce), Richard Klaassen, (GE Aviation), Tim Mouzakis (FAA), Cu Nguyen (FAA), Surendra Singh (Honeywell), Floyd Spencer (FAA consultant), and Harpreet Wasan (Pratt and Whitney). We would like to thank Floyd Spencer for suggesting that, for some purposes, it is important to estimate the quantile of the POD distribution.

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. Annis, C. (2009), R package: mh1823, Version 2.5.4.2, http://www.statisticalengineering.com/, downloaded March 2010.
2. Adler, L. and Achenbach, J. D. (1980), "Elastic Wave Diffraction by Elliptical Cracks: Theory and Experiment," *Journal of Nondestructive Evaluation*, 1, 87-99.
3. Gelman, A., Carlin, J. B., Stern, H. S. and Rubin, D. B. (2003), *Bayesian Data Analysis (Second Edition)*, Florida: Chapman and Hall.
4. Gigliotti, M.F.X., Gilmore, R.S., and Perocchi, L.C. (1994), "Microstructure and Sound Velocity of Ti-N-O Synthetic Inclusions in Ti-6Al-4V," *Metallurgical and Materials Transactions A*, 25, 2321-2329.
5. Huang, R., Schmerr, L. W., and Sedov, A. (2006), "A Modified Born Approximation for Scattering in Isotropic and Anistotropic Elastic Solids," *Journal of Nondestructive Evaluation*, 25, 139-154.
6. Margetan, F.J., Umbach, J., Roberts, R., Friedl, J., Degtyar, A., Keller, M., Hassan, W., Brasche, L., Klassen, A., Wasan, H., and Kinney, A. (2007), "Inspection Developments for Titanium Forgings," DOT/FAA/AR-05/46, Air Traffic Organization Operations Planning Office of Aviation Research and Development, Washington, DC 20591.
7. Meeker, W. Q. and Escobar, L. A. (1998), *Statistical Methods for Reliability Data*, New York: Wiley.
8. Nieters, E. J., Gilmore, R. S., Trzaskos, R. C., Young, J. D., Copley, D. C., Howard, P. J., Keller, M. E., and Leach, W. J. (1995), "A Multizone Technique for Billet Inspection," *Review of Progress in Quantitative Nondestructive Evaluation*, 14, 2137-2144.

9. NTSB/AAR-89/03 (1989), *Aircraft Accident Report for Aloha Airline Flight 243*, National Transportation Safety Board, 800 Independence Avenue S.W., Washington, D.C. 20594.
10. 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.
11. MIL-HDBK-1823A (2009), *Nondestructive evaluation system reliability assessment*, Standardization Order Desk, Building 4D, 700 Roberts Avenue, Philadelphia, PA 19111-5094.
12. Olin, D. B. and Meeker, W. Q. (1996), "Application of Statistical Methods to Nondestructive Evaluation," *Technometrics*, 38, 95-112.
13. Pawitan, Y. (2001), *In All Likelihood: Statistical Modelling and Inference Using Likelihood*, New York: Oxford University Press.
14. Spencer, F. W. (1996), "Discussion: Application of Statistical Methods to Nondestructive Evaluation," *Technometrics*, 38, 122-124.
15. Thompson, R. B. and Lopez, E. F. (1984), "The Effects of Focusing and Refraction on Gaussian Ultrasonic Beam," *Journal of Nondestructive Evaluation*, 4, 107-123.
16. WinBUGs (2007), Version 1.4.2, http://www.mrc-bsu.cam.ac.uk/bugs. Downloaded March 2010.