PROBABILITY OF DETECTION MODELING FOR ULTRASONIC TESTING

Pradipta Sarkar, William Q. Meeker,
R. Bruce Thompson, Timothy A. Gray
Center for Nondestructive Evaluation
Iowa State University Ames, IA 50011

Warren Junker
Westinghouse Sci. & Tech. Center
R & D Center
1310 Beulah Road
Pittsburgh, PA 15235

INTRODUCTION

Ultrasonic (UT) inspection can be used to detect a wide variety of subsurface discontinuities, such as inclusions, cracks and voids, as well as associated reactive or diffusion zones. In comparison with Eddy-Current inspection, much less work has been done on the determination of the Probability of Detection POD of UT inspections. This is because, unlike Eddy-Current inspection, it is very difficult to produce synthetic sub-surface flaws that adequately represent the acoustic properties of the naturally-occurring flaws (see Burkel et al., 1996). Here traditional methods for POD determination are difficult to apply.

The POD is defined as the probability that the response signal exceeds some detection threshold T as a function of inspection and flaw-related factors that include UT transducer characteristics, focus depth, scan increment, pulse rate, part geometry, type of material being inspected, surface roughness, position and orientation of the crack, crack size, crack morphology, etc. Conceptually, the POD can be obtained exactly if all of the factors affecting signal UT strength are known. In practice, however, we know only some of these factors. In this paper we treat the depth of the crack and some inspection modality parameters as known fixed-effect factors and treat all other parameters as unknown random-effect factors. We develop a methodology to estimate POD as a function of the known fixed-effect factors. Variability is estimated from observed variability from the random effect-factors. We also show how to obtain a corresponding confidence interval to express uncertainty arising from limited data.

The (POD) as a function of the crack size is defined as the probability, given a flaw of size a, that the response signal $Y(a)$ exceeds the detection threshold T,

$$\mathrm{POD}(a) = \Pr\left(Y(a) > T\right).$$

Here our notation suppresses the known inspection modality factors like UT transducer characteristics, part geometry, etc.

Probability of False Alarm (PFA) is defined as the probability that the response signal exceeds the detection threshold T when there is no flaw, that is,

Review of Progress in Quantitative Nondestructive Evaluation, Vol. 17
Edited by D.O. Thompson and D.E. Chimenti,, Plenum Press, New York, 1998

$$\text{PFA} = \Pr\left(Y(\text{no flaw}) > T\right).$$

Figure 1 shows examples of a possible noise distribution, two possible flaw distributions, and the detection threshold. The area under the noise distribution to the right of the detection threshold is the PFA and the area under the flaw distribution to the right of the detection threshold is the POD.

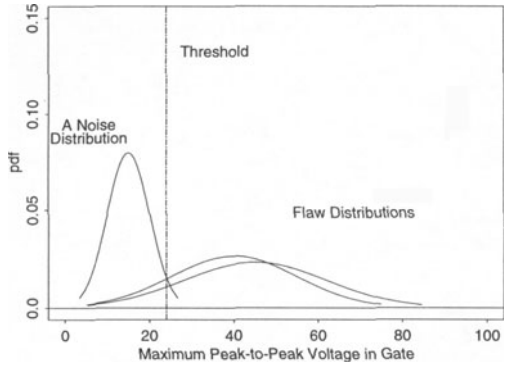

Figure 1. A noise distribution and two flaw distributions.

MODELING POD

With large amounts of data, one can use a model-free empirical approach to estimate POD. This type of POD estimation is common in medical applications where detection and non-detection information is recorded on patients and subsequently, the accuracy of the detection decision is determined.

In engineering applications, however, specimens and experiments are expensive and in many cases, expensive destructive metallographic evaluation is required to determine actual flaw characteristics. It is generally impossible to do experimental evaluations under all or even a substantial proportion of possible testing conditions. An alternative is to use a deterministic physical model to predict inspection results. This paper uses a deterministic POD model to explain the effect that inspection modality factors have on the expected UT signal strength along with a statistical model to quantify the variability in UT signal strength. The combination of these models allows one to predict POD at test conditions that are different from the original.

Example

The data used in this study consist of results from the nondestructive UT testing and subsequent destructive-test evaluation of cracks in heat exchanger tubes to obtain detailed information about crack size. For the purpose of our analysis we will use data taken from three different tubes, each tube providing a number of observations on UT signal strength and crack size. For the nondestructive testing, the transducer was moved along the length of the tube and signal amplitude was recorded. We will refer to these data as the axial position versus amplitude data. Subsequently, in a destructive test, the tubes were subjected to high pressure causing them to burst. The crack depths were then measured at different positions. We will refer to these data as the axial position versus crack depth data.

For each tube, measurements were available for signal amplitude and crack depth at a number of different axial positions. Because the two sets of data (axial position versus signal amplitude, and axial position versus crack depth) were collected at different times and for different purposes, the axial position values were relative and did not match up, in recorded position, between the two data sets. We will refer axial position versus amplitude data set as $(\mathbf{x^{(1)}}, \mathbf{y})$, and axial position versus crack depth data set as $(\mathbf{x^{(2)}}, \mathbf{a})$. Let n_A, n_B and n_C be the number of readings for the $(\mathbf{x^{(1)}}, \mathbf{y})$ data sets for tubes A, B and C respectively. Similarly let m_A, m_B and m_C be the number of readings for the $(\mathbf{x^{(2)}}, \mathbf{a})$ data sets for tubes A, B and C respectively.

Preliminary Analysis

The first step in our analysis was to match these two data sets for each tube. This could be done by using any of the following procedures:

1. Visual inspection: Superimpose the data sets and move one of the two data sets back and forth until a visual good match occurs.

2. Match the $(\mathbf{x^{(1)}}, \mathbf{y})$ and $(\mathbf{x^{(2)}}, \mathbf{a})$ data sets in such a way that the distance between the minimums of the $x^{(1)}$ and $x^{(2)}$ values is same as the distance between the maximums of the $x^{(1)}$ and $x^{(2)}$ values.

   Computationally, $L^{(j)} = \max_i\{x_i^{(j)}\} - \min_i\{x_i^{(j)}\}, \quad j = 1, 2.$

   If $L^{(1)} > L^{(2)}$, then define the interpolated $x^{(2)}$ values as

$$x_{i,new}^{(2)} = x_i^{(2)} - \min_k\{x_k^{(2)}\} + \min_k\{x_k^{(1)}\} + \frac{L^{(1)} - L^{(2)}}{2}.$$

   For notational convenience the new data set $\left(x_{i,new}^{(2)}, a_i\right)$ will be denoted as $\left(x_i^{(2)}, a_i\right)$ for subsequent analysis.

   If $L^{(2)} > L^{(1)}$, then define the interpolated $x^{(1)}$ values as

$$x_{i,new}^{(1)} = x_i^{(1)} - \min_k\{x_k^{(1)}\} + \min_k\{x_k^{(2)}\} + \frac{L^{(2)} - L^{(1)}}{2}.$$

   For notational convenience the new data set $\left(x_{k,new}^{(1)}, y_k\right)$ will be denoted as $\left(x_k^{(1)}, y_k\right)$ for subsequent analysis.

3. Minimize the sum of the squared deviations between the crack-size and UT signal data sets. Superimpose the two data sets with axial position along the horizontal axis. Interpolate/extrapolate the ordinates a at $x^{(1)}$ values (or alternatively, interpolate/extrapolate the ordinates y at $x^{(2)}$ values). Calculate the sum of squared deviations $\sum(\widehat{a}_k - y_k)^2$ (or $\sum(\widehat{y}_i - a_i)^2$ respectively), where $\widehat{a}_k$ and $\widehat{y}_i$ are the estimated value of a at $x_k^{(1)}$ and the estimated value of y at $x_i^{(2)}$ respectively. Move the second data set back and forth and calculate the sum of squared deviations for each position (over a fine grid). Choose the position for which the sum of squared deviations is minimized.

We used method 2 for our analysis. Method 2 is easy to implement and computationally simpler than method 3. Also, method 2 is not as subjective as method 1.

Estimated Signal and Theoretical Model Prediction Curve

Given a signal amplitude we calculated the estimated crack size from the matched data sets. To do so we joined the $\left(x_k^{(1)}, y_k\right)$ points by piecewise straight lines. Similarly the

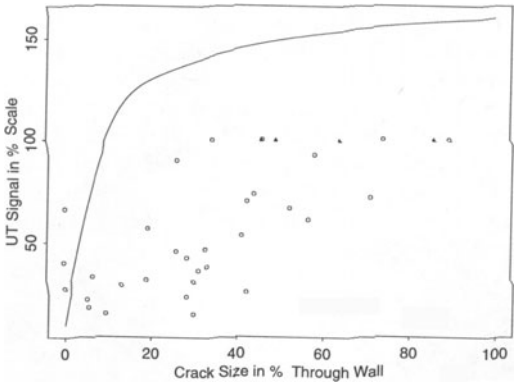

Figure 2. UT measurements and theoretical model prediction.

points $\left(x_i^{(2)}, a_i\right)$ were also joined by piecewise straight lines. The estimated crack size for a given signal y_k, denoted by $\hat{a}_k \equiv a(x_k^{(1)})$ was calculated by interpolation from the piecewise linear graph joining the $\left(x_i^{(2)}, a_i\right)$ points. We do this for all the cracks. This yielded a combined data set $\{(\hat{a}_k, y_k)\,;\, k = 1, \cdots, n\}$, where $n = n_A + n_B + n_C$. Figure 2 shows this combined data set together with the theoretical signal strength prediction curve as a function of the crack size for a given set of inspection modality parameters. The $\hat{a}_k$ values corresponding to the saturated (right censored) observations are indicated by small triangles. The following features are apparent from Figure 2.

1. Except for the three points with zero cracks sizes, the theoretical model predictions are always above the corresponding point in the scatter plot. This is because the theoretical model prediction curve is computed for an ideal reflector, a rectangular slot. Actual flaws have less reflectivity than a rectangular slot.

2. The three points corresponding to zero crack size are thought to have resulted from false indications (noise) in the vicinity of the crack. For this reason, these points are not used in our formal analysis.

3. Although the signal strength will usually increase with an increase in the crack size, both the scatter plot and the theoretical model prediction curve show the existence of a beam limiting effect. The beam limiting effect arises because as the crack size approaches (and eventually exceeds) the UT beam width, the amount of reflected energy does not increase significantly.

4. The wide spread along the vertical axis indicates high variability in the signal amplitudes for a given crack size. Most of the variability is believed to be due to flaw morphology.

Modeling Generalized Deviation

Let $\tilde{y}_k$ denote the predicted signal from the theoretical model for a crack size of a_k. The generalized deviations between the theoretical model predictions and the actual data

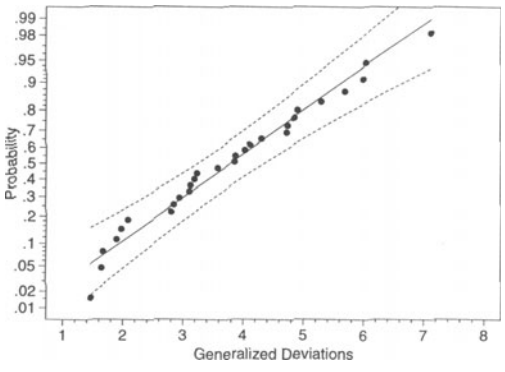

Figure 3. Normal probability plot of the generalized deviations with $\gamma = 0.3$.

are defined as

$$e_k(y_k, \tilde{y}_k; \gamma) = \begin{cases} \dfrac{\tilde{y}_k^{\gamma} - y_k^{\gamma}}{\gamma}, & \gamma \neq 0 \\ \log(\tilde{y}_k) - \log(y_k), & \gamma = 0. \end{cases} \tag{1}$$

The transformation used here is known as the Box-Cox transformation. The e_k's corresponding to the saturated y_k values are also saturated and we treat them as right censored observations for the purpose of analysis.

Previous experiences with experimental UT data on synthetic flaws in titanium (see Meeker et al., 1996) has shown that the value of γ in the neighborhood of 0.3 results in e_k's that are approximately distributed as a normal distribution with constant mean μ and standard deviation σ. In our analysis we also found $\gamma = 0.3$ suitable and use the method of maximum likelihood estimation to estimate μ and σ. Figure 3 shows the normal probability plot for the fitted distribution along with a 95% pointwise confidence interval for the underlying normal distribution.

Estimation of POD

We estimate the POD under the assumption that the distribution of the generalized deviations is same over different transducers and other testing conditions. For each transducer the theoretical model provides a prediction for the typical UT signal value. This can be combined with the fitted probability distribution for the generalized deviations to obtain an estimate of the probability distribution of the signal amplitude for each given true crack size. For a given detection threshold T, the estimated probability of detection as a function of the crack size, say a, is

$$\begin{aligned} \widehat{\mathrm{POD}}(a) &= \Pr\left(Y(a) > T\right) = \Pr\left(\frac{\tilde{Y}^{\gamma}(a) - Y^{\gamma}(a)}{\gamma} < \frac{\tilde{Y}^{\gamma}(a) - T^{\gamma}}{\gamma}\right) \\ &= \Pr\left(\frac{e(Y, \tilde{Y}; \gamma) - \widehat{\mu}}{\widehat{\sigma}} < \frac{\frac{\tilde{Y}^{\gamma}(a) - T^{\gamma}}{\gamma} - \widehat{\mu}}{\widehat{\sigma}}\right) \\ &= \Phi\left(\frac{\frac{\tilde{Y}^{\gamma}(a) - T^{\gamma}}{\gamma} - \widehat{\mu}}{\widehat{\sigma}}\right) \qquad \text{for } \gamma \neq 0, \end{aligned} \tag{2}$$

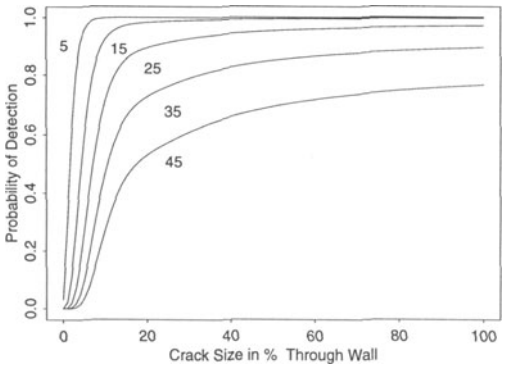

Figure 4. PODcurves for different detection threshold levels.

$$\widehat{\mathrm{POD}}(a) = \Phi\left(\frac{\log\left(\tilde{Y}(a)\right) - \log(T) - \hat{\mu}}{\hat{\sigma}}\right) \qquad \text{for } \gamma = 0, \tag{3}$$

where $\hat{\mu}$ and $\hat{\sigma}$ are the maximum likelihood estimators of μ and σ respectively. Figure 4 presents the estimates of the POD curves for different specified detection thresholds. For higher detection thresholds, the asymptotes of the estimated POD curves are attained at higher crack size. Also as expected, for a given crack size the estimated POD is lower if the detection threshold is higher.

Confidence Interval for POD

The logit function $\hat{d}(a)$ is defined as

$$\hat{d}(a) = \log\left(\frac{\widehat{\mathrm{POD}}(a)}{1 - \widehat{\mathrm{POD}}(a)}\right).$$

$\widehat{\mathrm{POD}}(a)$ is bounded by zero and one, but $\hat{d}(a)$ can be any real number. For this reason, the approximate large sample confidence intervals based on the distribution of the studentized (random quantity minus expected value, divided by an estimate of the standard error) logit $\hat{d}(a)$ transformation of POD is expected to have better statistical properties than the confidence intervals that are constructed using a studentized statistic based on an untransformed POD(a) (see Agresti, 1990, pp 419-422 for details). Therefore, we use the logit transformation of POD to construct the pointwise confidence interval for the POD.

Under regularity conditions that hold for the model in this paper, using the delta method provides an estimator of the variance of $\hat{d}(a)$ as

$$\widehat{\mathrm{Var}}\left[\hat{d}(a)\right] = \frac{\phi^2\left(\dfrac{\tilde{Y}^{\gamma}(a) - T^{\gamma}}{\gamma} - \hat{\mu}\right)}{\hat{\sigma}^2\hat{p}^2(1-\hat{p})^2}\left\{\widehat{\mathrm{Var}}(\hat{\mu}) + \left(\frac{\tilde{Y}^{\gamma}(a) - T^{\gamma}}{\gamma} - \hat{\mu}\right)^2 \frac{\widehat{\mathrm{Var}}(\hat{\sigma})}{\hat{\sigma}^2}\right.$$
$$\left. + 2\left(\frac{\tilde{Y}^{\gamma}(a) - T^{\gamma}}{\gamma} - \hat{\mu}\right)\frac{\widehat{\mathrm{Cov}}(\hat{\mu},\hat{\sigma})}{\hat{\sigma}}\right\}, \qquad \text{for } \gamma \neq 0,$$

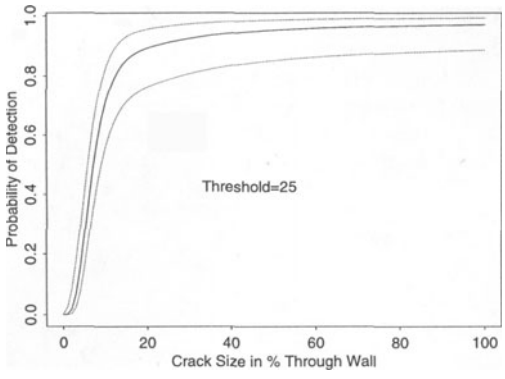

Figure 5. 95% confidence interval for POD.

where $\hat{p} = \widehat{\mathrm{POD}}(a)$. A similar expression can be obtained for $\gamma = 0$ case. Thus a large sample approximate 95% pointwise confidence interval for the logit-transform $\hat{d}(a)$ is

$$\hat{d}(a) - 1.96\sqrt{\widehat{\mathrm{Var}}\left[\hat{d}(a)\right]} \leq d(a) \leq \hat{d}(a) + 1.96\sqrt{\widehat{\mathrm{Var}}\left[\hat{d}(a)\right]}. \tag{4}$$

Because $\mathrm{POD}(a) = (1 + \exp[-d(a)])^{-1}$ is a monotone increasing function of $d(a)$, a large sample approximate 95% pointwise confidence interval for $\mathrm{POD}(a)$ is constructed by inverting the interval (4). Figure 5 shows the 95% large sample approximate pointwise confidence interval for $\mathrm{POD}(a)$.

AREAS FOR FURTHER RESEARCH

- In our analysis we have assumed the Box-Cox parameter $\gamma = 0.3$. Alternatively, one can do a full maximum likelihood estimation to estimate γ, μ, and σ^2 simultaneously.
- Instead of using a normal distribution, other distributions such as Weibull, lognormal might be used to model the generalized deviations.
- The distributions of the generalized deviations under different inspection and flaw/material conditions should be investigated and compared. This will require additional data sets.
- Often it is desired to estimate POD simultaneously for more than one flaw size (or other specified set of fixed-effect conditions). In such cases, uncertainty should be quantified by constructing simultaneous confidence bands for POD. Accurate methods for doing this need to be developed.
- In our present work we have defined crack size to be the depth of the crack. It may be interesting to extend the procedure to the multidimensional situation where other aspects of the flaw geometry are taken into consideration.
- It would be of interest to obtain an estimate of the noise distribution. This would allow for the computation of the probability of false alarms (PFA) and Receiver Operating Curves (ROC).

ACKNOWLEDGMENTS

This material is based upon work performed at Iowa State University, and Westinghouse Science and Technology Center. The work was supported by the program for Integrated Design, NDE and Manufacturing Sciences under U.S. Department of Commerce, National Institute of Standards and Technology Cooperative Agreement No. 70NANB4H1535.

REFERENCES

1. Agresti, A. (1990), Categorical Data Analysis. (John Wiley & Sons, NY, 1990).
2. Burkel, R. H., Sturges, D. J., Tucker, W. T. and Gilmore, R. S. (1996), Probability of Detection for Applied Ultrasonic Inspections, Review of Progress in Quantitative NDE, 15B, D.O. Thompson and D.E. Chimenti, Eds. (Plenum Press, NY, 1996), 1991-1998.
3. Meeker, W. Q., Thompson, R. Bruce, Chiou, C. P., Jeng, S. L. and Tucker, W. T. (1996), Methodology for Nondestructive Evaluation Capability, Review of Progress in Quantitative NDE, 15B, D.O. Thompson and D.E. Chimenti, Eds. (Plenum Press, NY, 1996),1983-1990.