ABSTRACT. Damage tolerance-based operation and maintenance of aircraft components require the estimation of probability of detection (POD) of nondestructive inspection (NDI) techniques. The standard approach to estimate POD of an NDI process is to perform demonstration tests on simulated or real components. To generate a realistic POD, MIL-HDBK-1823 suggests that flaw sizes should be uniformly distributed on a log scale covering the expected range of the POD curve in the transition region. This paper presents an alternative approach that is based on Rayleigh distribution of flaw sizes on a log scale covering the expected range of flaws. This approach is only applicable to NDI methods that provide a signal response or “a” for a given flaw size “a”. The use of this approach increases the accuracy of the POD estimate by optimizing the distribution of flaw sizes in the range of transition between very low and very high POD. The paper also presents a new three dimensional representation of POD versus the false call rate, defined by the relative operating characteristic (ROC). The 3D presentation makes it easier to visualize the change in POD and false call rate with the variation of decision threshold set for flaw detection.

INTRODUCTION

Damage tolerance maintenance of aircraft components requires periodic nondestructive inspection (NDI) to ensure continued safe operation. The scheduling of the NDI is based on the desired risk level and the probability of detection of the NDI technique in the particular application. A typical NDI reliability assessment [1,2] includes the following steps: First is the system definition in terms of the limits of operational parameters and range of application. Then, the experimental design for the inspections must be performed. The experimental design defines the conditions for the NDE process and all the process variables. Specimens are inspected, and results are converted into POD relationships.

The goal of the NDI reliability experimental design step is to measure the effect of the change of one or more process variables (input factors) on response variables (outputs). Experimental design may be illustrated by a black box linking the measured outputs response and input factors, as is shown in Figure 1.
This black box model has several discrete or continuous input factors that can be controlled. But it often has to account for a number of uncontrolled factors that may be discrete, such as different inspectors or sensors and/or continuous such as ambient temperature or humidity. Experimental design has to include test specimens (including geometry, material, part processing, number of flawed specimens and flaw sizes, etc.) and inspection process. Inspection process includes such factors as scan rates, scan path, and probe frequency threshold levels to name a few.

This paper discusses issues surrounding the number of flawed specimens and the flaw sizes, as well as the relationship between threshold levels, false calls, and POD of the inspection process. In the first section, the effect of flaw size distribution on the accuracy of POD estimation is demonstrated. In the second section, the role of threshold level in the tradeoff between the false call rate and the ability to detect very small flaws is explained. Finally, a three-dimensional representation of POD-false call rate-flaw size is presented.

**FLAW SIZE DISTRIBUTION FOR POD ESTIMATION**

In the determination of the reliability of NDI by POD estimation, it is desirable to use flawed components from service as test specimens, as these will provide the most representative results [3]. However, flawed components are often not available, and therefore engineered specimens are often used. This allows the experimenter to choose the number and sizes of flaws to be included in the experiment. In the case of complex geometries, specimen manufacture may be a significant portion of the cost of the entire POD experiment, and it is therefore desirable to choose the optimum set of flaw sizes to be produced in order to achieve the best estimate of POD.

MIL-HDBK-1823 recommends that the specimen test set contain at least 60 flawed sites if the system provides only hit/miss result and at least 40 flawed sites if the system provides a quantitative response, \( a \), to a flaw. It also suggests that the flaw sizes should be uniformly distributed on a log scale covering the expected range of increase of the POD function.

The following example of the POD analysis for \( \hat{a} \) vs. \( a \) shows the effect of different distributions of flaw sizes on the accuracy of the estimated POD. Figure 2 shows the histogram, the POD and the lower 95% confidence bound of an arbitrarily chosen eddy current data set [4]. These data were obtained by eddy current inspections of fatigue cracks.
in flat plates. This POD was selected as a "true" or "underlying" POD, and estimates of this underlying POD were simulated using only 19 flaws and three different flaw size distributions. The choice of 19 flaws is again somewhat arbitrary, but for fewer flaws it is difficult to distinguish between different distributions, and for more than about 40 flaws, the choice of distribution becomes less significant.

As MIL-HDBK-1823 notes, the POD is not known before the test, and the experimenter must make an estimate of where the POD changes from near zero to near unity. This is where the majority of the flaws should be distributed. For the different distributions employed here, the same transition region was used, and was obtained from the underlying POD. Qualitative comments are included on the sensitivity of the different flaw distributions to the assumed transition region.

Without a priori knowledge of the POD, the experimenter determining flaw sizes must first determine the $a_{50\%}$ that is the middle of the transition region of POD. It can easily be estimated by assuming the linear relationship between $\ln(d)$ and $\ln(a)$. As shown in Figure 3, a linear curve fitting between 2 or 3 measurements on known crack sizes provides a line that can be used to obtain an estimate of $a_{50\%}$.

![HISTOGRAM](Image)

**FIGURE 2.** Histogram, the POD, and the lower 95% confidence bound of the eddy current data set [4].

![Figure 3](Image)

**FIGURE 3.** NDT system response versus flaw size.
Secondly, the width of the transition region of POD must be estimated to use uniform or normally distributed flaw sizes. However, that is generally impossible to determine in the experimental design step. Therefore, the precision of the POD will again be dependent on choices made by the experimenter, with little data to provide guidance.

The first distribution shown is the uniform distribution of flaw sizes on a log scale, covering the transition region of the POD, that is recommended by MIL-HDBK-1823. Figure 4 shows the histogram, the underlying POD, and the POD estimated from 19 flawed sites that were distributed uniformly on a log scale around the transition region of the POD.

It is also possible to distribute the flaw sizes according to the normal distribution. The 19 flaw sites for this test are normally distributed with mean equal to the flaw size at 50% POD, denoted $a_{50\%}$, and standard deviation equal to the half of the transition region of POD. Using these values, 68% of the flaw sizes fall in the transition region of the POD.

Figure 5 shows the histogram, the underlying POD, and the estimated POD from 19 flawed sites that are normally distributed based on the above procedure.

In order to eliminate the difficulty of estimating a priori the transition region of the POD, a trial was made using a Rayleigh distribution of flaw sizes in the expected range of flaw size. The Rayleigh distribution is a special case of the Weibull distribution that is an appropriate analytical tool for reliability and lifetime modeling [5].

![Figure 4](image1.png)

**FIGURE 4.** Histogram, underlying POD, and the POD estimated from 19 uniformly distributed flaws.

![Figure 5](image2.png)

**FIGURE 5.** Histogram, underlying POD, and the POD estimated from 19 normally distributed flaws.
The Rayleigh probability density function is defined as:

\[ f(x \mid b) = \frac{x}{b^2} e^{-\frac{x^2}{2b^2}} \]  

(1)

where \( b \) is the parameter of Rayleigh distribution. The Rayleigh distribution is chosen because it is a one parameter distribution and often simpler to handle statistically than the other lifetime distributions.

If the flaw sizes are distributed based on a Rayleigh model such that the maximum of the Rayleigh distribution corresponds to the \( a_{50\%} \), the majority of the flaws will fall in the transition region of the POD. In addition, 60% of flaw sizes fall after the \( a_{50\%} \), providing more information in the region where the POD is near unity. Figure 6 shows the histogram, the underlying POD and the POD estimated from 19 flawed sites that are distributed based on the Rayleigh model. The POD from 19 flawed sites distributed based on Rayleigh distribution and the underlying POD are almost overlapped.

In order to compare the effect of different distributions of flaw sizes on the POD, the difference curve between the estimated POD from 19 flaw sizes distributed by different distributions and the underlying POD is plotted. The area under this curve is then a measure of the error in the estimates of POD. Figure 7 shows the difference curve and the area under this curve for uniform, normal, and Rayleigh distributions of flaw sizes.

In comparing the error area from uniform, normal, and Rayleigh distribution of flaw sizes for the above example it can be seen that in case of the normal distribution of flaw sizes the error area is almost half of the error area for the case of the uniform distribution. Moreover, in case of the Rayleigh distribution of flaw sizes the error area is even less than that of the normal distribution.

Thus the distribution of crack sizes based on a Rayleigh model provides not only an excellent estimation of the underlying POD, but also a better result than that obtained from a normal distribution of flaw sizes or the MIL-HDBK-1823 suggestion. In addition, the Rayleigh model has only one parameter \( b \) to adjust while the normal model has two parameters \((\mu, \sigma)\) to adjust.

FIGURE 6. Histogram, underlying POD, and the POD estimated from 19 flawed sites distributed based on a Rayleigh model.
INSPECTION RELIABILITY

An NDI reliability assessment is generally presented as a POD curve. But there are problems associated with characterizing NDI performance solely through POD curves. An arbitrarily high POD can be attained by simply setting a very low decision threshold. This leads to using the probability of false call rate in addition to POD curve to characterize the performance of an NDI technique.

The probability of false calls can be presented through Relative Operating Characteristic (ROC) curves [6,7]. A single ROC curve plots the POD versus the probability of false call rate with decision threshold level as a parameter. The ROC gives a measure of the accuracy of an inspection system by evaluating the findings rate (POD) versus the false call rate for a set of possible decision thresholds (for the decision whether the signal is noise or a defect signal). Figure 8 shows an example of a ROC curve for one flaw size.

As shown in Figure 8, when the decision threshold decreases, the false call rate and the POD increase, meaning the sensitivity of the system rises. Following the ROC curve from the lower left corner to the upper right, more and more of the defects are found but a greater number of false calls are also reported.

FIGURE 8. An example of an ROC (Receiver Operating Characteristic) curve.
Three Dimensional Representation of the POD, False Call Rate and Flaw Size

Since the POD curve gives the relationship between the POD and the flaw size, and the ROC curve gives relationship between the false call rate and the POD; a three dimensional representation of the POD, the false call rate, and the flaw size may help us to better understand the relationship between these three parameters.

Figure 9(d) shows the 3D representation of the POD, false call rate and flaw size. This representation allows us to simultaneously study the behavior of the POD curves and the ROC curves when the threshold level changes. Figure 9(a) shows the POD for different values of decision threshold. Figure 9(b) shows the ROC curves for flaw sizes from $a_{50\%}$ until the $a_{90\%}$. Figure 9(c) is the projection of the 3D representation onto the false call rate-flaw size plane. Each line describes how for one value of POD, the corresponding flaw size decreases as the associated false call rate increases. For example, the flaw size $a_{50\%}$ depends on the chosen decision threshold, denoted $\hat{a}_{dec}$. As $\hat{a}_{dec}$ is increased, the flaw size $a_{50\%}$ increases and the false call rate decreases.

![Figure 9](image-url)
CONCLUSION

In the determination of the reliability of NDI by POD estimation, engineered specimens are often used. This allows the experimenter to choose the number and sizes of flaws to be included in the experiment. In the case of complex geometries, specimen manufacture may be a significant portion of the cost of the entire POD experiment, and it is therefore desirable to choose the right flaw sizes to be produced in order to achieve the best estimate of POD. This paper studied the effect of the flaw size distribution on the accuracy of estimated POD using uniform, normal, and Rayleigh models. The Rayleigh distribution of flaw sizes provided the most accurate estimate of POD for a fixed number of flaws.

In addition, a three dimensional representation of the POD, false call rate, and flaw size was demonstrated in order to provide a better understanding of trade-offs between false call rates and detection rates.

ACKNOWLEDGEMENTS

This work was supported by National Research Council Canada and Department of National Defence Canada.

REFERENCES