

ANALYSIS OF THE RFC/NDE

SYSTEM PERFORMANCE EVALUATION EXPERIMENTS

A. P. Berens

University of Dayton Research Institute
Dayton, Ohio 45469-0001

INTRODUCTION

The RFC/NDE System comprises automated eddy current and ultrasonic inspection stations developed by Systems Research Laboratories for Air Force depot level inspections of aircraft engines. The system performance is being evaluated through a series of experiments whose objectives are to quantify crack detection capability as a function of crack size in various engine components. This paper presents the eddy current results from the first (preliminary) phase of this evaluation.

The RFC/NDE eddy current system makes accept/reject decisions by comparing peak voltage, \hat{a} , to a pre-defined threshold value. The \hat{a} values from specimens with known crack sizes were recorded and provided the basis for estimating the probability of detection, $POD(a)$, as well as for quantifying the effects of repeated measurements, probe changes, load changes, crack orientation, and operator changes. Analysis of the \hat{a} values also validated the use of the cumulative lognormal distribution as an appropriate model for the $POD(a)$ function.

REVIEW OF \hat{a} vs a ANALYSIS METHOD

Automated NDE systems make accept/reject decisions based on the analysis of a response to an induced inspection stimulus. In most current automated systems, this analysis produces a single numerical value, \hat{a} , which is compared to a threshold, \hat{a}_{TH} . The magnitude of \hat{a} is influenced by many factors related to both the inspection system and the geometry and material state of a flaw. However, the property of interest for damage tolerance analysis is the characteristic flaw size, a , of a crack growth equation. Therefore, the inspection response, \hat{a} , is modeled only as a function of a with all other factors contributing to a random component whose statistical properties determine the probability of detection. That is, if

$$\hat{a} = f(a) + e \quad (1)$$

then

$$\begin{aligned} POD(a) &= P\{\hat{a} > \hat{a}_{TH}\} \\ &= P\{e > \hat{a}_{TH} - f(a)\} . \end{aligned} \quad (2)$$

Since \hat{a} vs a data from inspections of representative cracks can be used to estimate both $f(a)$ and the statistical properties of e , such data can also be used to estimate the POD function. Further, uncertainty in the POD function can be quantified by a confidence bound which is calculated from the statistical properties of the parameter estimates [1].

The POD function will be a cumulative lognormal distribution if the relationship between \hat{a} and a is given by

$$\ln \hat{a} = b_0 + b_1 \ln a + e, \quad (3)$$

where e is normally distributed with zero mean and a variance of S^2 . That is, equations (2) and (3) imply that

$$\text{POD}(a) = \Phi \left[\frac{\ln a - \frac{1}{b_1} (\ln \hat{a}_{\text{TH}} - b_0)}{S/b_1} \right] \quad (4)$$

where $\Phi(Z)$ is the standard normal cumulative distribution.

Analysis of \hat{a} vs a data from several eddy current inspection reliability experiments has shown that equation (3) is often (but not necessarily always) an acceptable model for the \hat{a} vs a relationship. The analysis of the RFC/NDE eddy current reliability data was first directed toward testing hypotheses regarding the applicability of equation (3) and then toward estimating the $\text{POD}(a)$ functions.

PRELIMINARY RFC/NDE EDDY CURRENT RELIABILITY EXPERIMENTS

A preliminary evaluation of the RFC/NDE system was performed prior to its delivery to the Air Force. The objectives of these preliminary experiments were to demonstrate that the system was reasonably capable of meeting specifications and to determine the effect (if any) of five potential sources of scatter in \hat{a} measurements. Real fatigue cracks in three specimen types were used in these preliminary tests. The specimens simulated engine rivet holes, bolt holes, and the web/bore surface. Each set of specimens contained at least 30 fatigue cracks and at least 30 more inspection opportunities without cracks. Crack size, a , in these experiments is crack depth and the \hat{a} value is peak voltage.

By performing repeat measurements under controlled experimental conditions, specific contributions to total scatter in \hat{a} values can be isolated [1]. Accordingly, repeat inspections of each set of specimens were made to quantify the effects of five potential sources of variability: a) repeatability - five measurements without removing specimens from the fixtures; b) transducers - repeat inspection using a second transducer; c) specimen loading - specimens were removed from fixture, reloaded, and re-inspected; d) operator - independent inspections by three operators; and e) crack orientation - three methods of loading the rivet and bolt hole specimens in the fixtures. Only some of these tests were performed on the web/bore surface specimens due to the length of time required to inspect these specimens. It was agreed before the start of testing that system capability for a specimen type would be estimated from the first inspection of the specimens.

Further details on these tests can be found in two other papers in these proceedings [2,3]. The following paragraphs summarize the results for each of the specimen types.

Rivet Hole Specimens

Figure 1 presents a plot of $\ln \hat{a}$ vs $\ln a$ for the rivet hole specimens. The crack detection threshold, \hat{a}_{TH} , was set at 140 mv for these inspections as indicated by the horizontal line. Ten cracks were not detected in this inspection. These cracks had depths of 1, 3, 3, 4, 6, 7, 7, 8, 9, and 16 mils. Four extra indications were recorded but it is not known whether or not these extra indications are false calls.

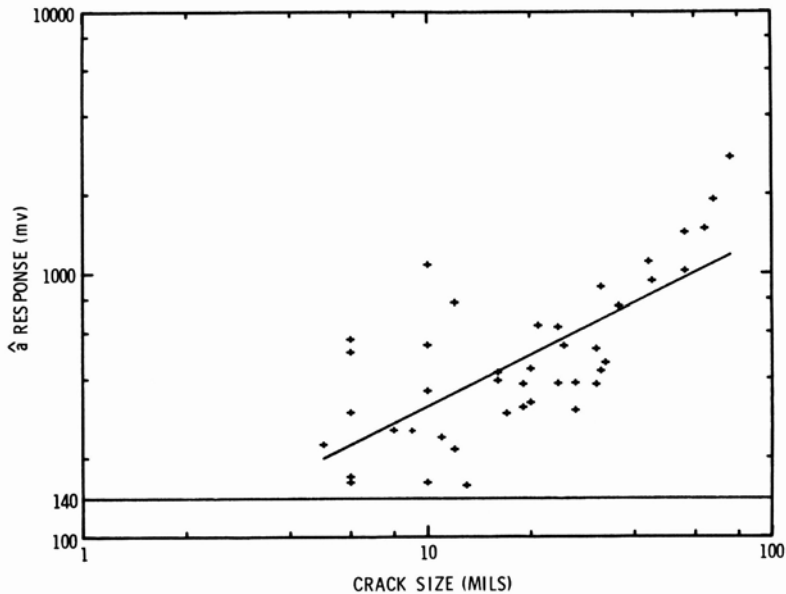


Fig. 1. \hat{a} vs a for the rivet hole specimens.

The linear least squares fit to the $\ln \hat{a}$ vs $\ln a$ data is also shown in Figure 1 and it was judged to adequately fit the data. Deviations (residuals) of individual $\ln \hat{a}$ values from the regression line were obtained and analyzed for normality. Figure 2 displays a plot of the residuals on a form of normal distribution paper. Deviation from linearity in Figure 2 indicates non-normality. The Shapiro-Wilkes test, which is based on the correlation coefficient between the residuals and the normal scores, failed to reject the hypothesis that the $\ln \hat{a}$ values are normally distributed about the $b_0 + b_1 \ln a$ line. Thus, the cumulative lognormal model was judged to be an acceptable model for the POD(a) function. Figure 3 presents the estimated POD function and its lower 95 percent confidence bound for the rivet hole specimens.

Figure 4 presents the \hat{a} vs a test results for five repeat inspections during which the specimens were not removed from the fixtures. The scatter reflected in the \hat{a} values for a particular crack in this figure represents the optimum repeatability of the system for this specimen type during Phase I. The scatter was quantified by pooling the standard deviations of $\log \hat{a}$ values from each of the cracks. The pooled estimate of the standard deviation for the repeatability measurements was 0.050. Assuming a lognormal distribution of \hat{a} values, this degree of repeatability would imply approximately a 5 percent coefficient of variation, i.e., 95 percent of \hat{a} values for a particular crack would be within ± 10 percent of the average.

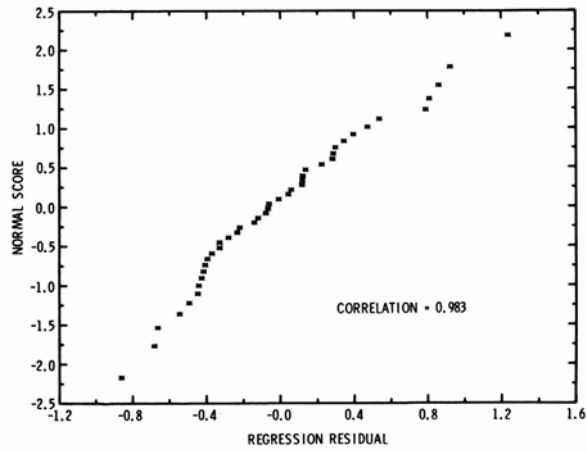


Fig. 2. Normal probability plot of $\ln \hat{a}$ residuals for the rivet specimen.

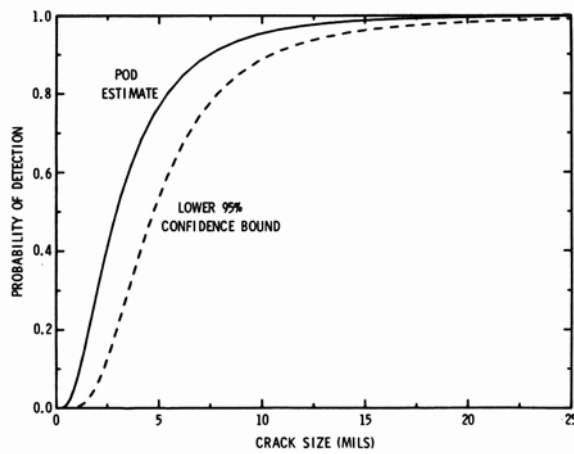


Fig. 3. POD(a) and lower 95% confidence bound for the rivet hole specimens.

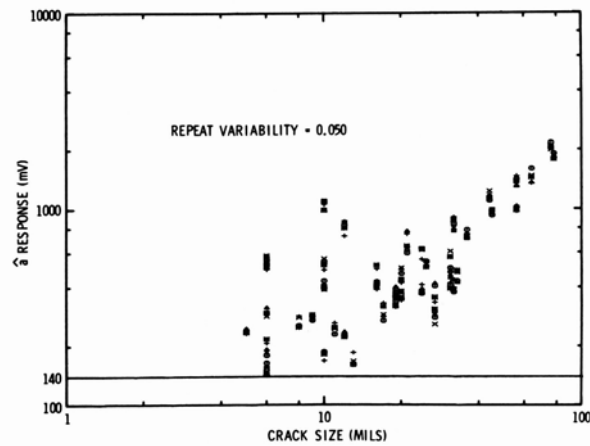


Fig. 4. \hat{a} vs a for five repeat inspections under identical conditions.

Similar analyses were performed on the repeat \hat{a} measurements when inspections were repeated with all but a single factor held constant. Table 1 presents these pooled estimates of the standard deviation of $\ln \hat{a}$ values. The degree of scatter exhibited by multiple loadings of the specimens in the test machines and by different operators is about equal to the minimum scatter as measured by the repeatability standard deviation. The \hat{a} scatter introduced by using a different transducer, however, was significantly greater than the minimum. Changing the orientation of the crack with respect to the transducer also resulted in a significant increase in \hat{a} scatter.

Table 1. Coefficient of Variation for Sources Isolated during Preliminary RFC/NDE Eddy Current Reliability Evaluation

	RIVET HOLE	BOLT HOLE	WEB/ BORE
REPEATABILITY	0.050	0.043	0.017
TRANSDUCER	0.112	0.134	0.177
LOADING	0.056	0.048	
OPERATOR	0.055	0.056	
CRACK ORIENTATION	0.104	0.129	

Bolt Hole Specimens

The \hat{a} vs a data obtained from the inspections of the bolt hole specimens are presented in Figure 5. All except a one mil deep crack in one of the specimens were detected. Extra indications (possibly false calls) were recorded on eight of the specimens during the characterization inspections. The linear $\ln \hat{a}$ vs $\ln a$ relationship was judged to be acceptable and the Shapiro-Wilks tests could not reject the normality hypothesis for the deviations of $\ln \hat{a}$ values about the straight line. Figure 6 presents the estimated lognormal POD(a) function and its lower 95 percent confidence bound for the bolt hole specimens.

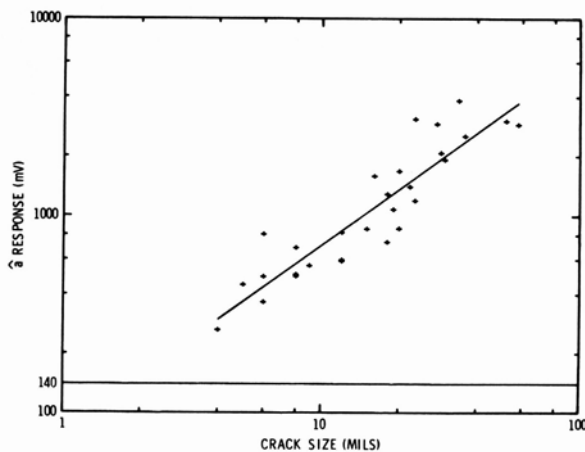


Fig. 5. \hat{a} vs a for the bolt hole specimens.

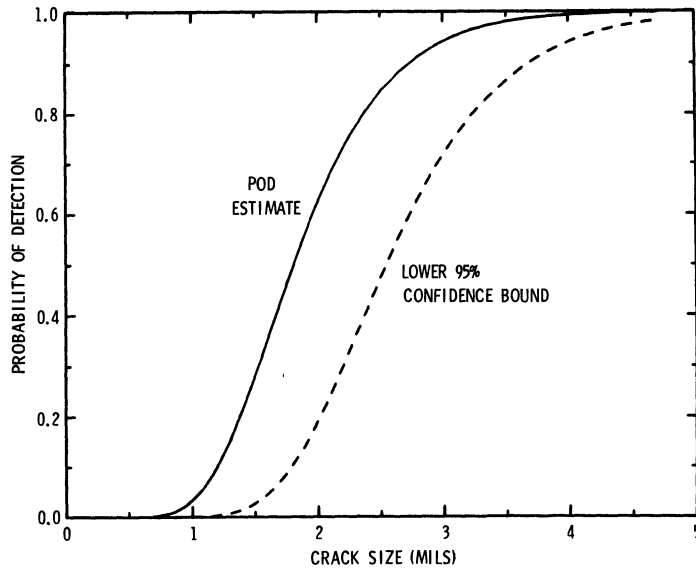


Fig. 6. POD(a) and lower 95% confidence bound for the bolt hole specimens.

The results of the tests to measure the scatter resulting from controlled factors are presented in Table 1. These results are essentially equivalent to those obtained from the rivet hole specimens. Repeat loading and different inspectors do not increase the scatter in \hat{a} values and, hence, do not degrade the POD(a) function. Different transducers, however, do increase the scatter in \hat{a} .

Web/Bore Specimens

Figure 7 presents the \hat{a} vs a data obtained from the specimens which simulated inspections of an engine disk web/bore. Two cracks with depths of one and three mils were not detected in this experiment. Ten extra indications were recorded. Again the $\ln \hat{a}$ vs $\ln a$ relationship was judged to be linear and normality of the deviations of $\ln \hat{a}$ values from the straight line could not be rejected by the Shapiro-Wilkes test. The lognormal POD(a) function and its lower 95 percent confidence bound for these data are presented in Figure 8.

Since scanning the web/bore specimens proceeded slowly, the experimental matrix was greatly reduced for this specimen type. Only one set of repeatability measurements were taken and only the variability due to different transducers was obtained. The results of these experiments are presented in Table 1.

DISCUSSION

Equation (4) implies that the more scatter in \hat{a} values for a fixed crack depth, the flatter the POD(a) function. This is also easily realized from the \hat{a} vs a plots. A consistent trend that has been present for \hat{a} vs

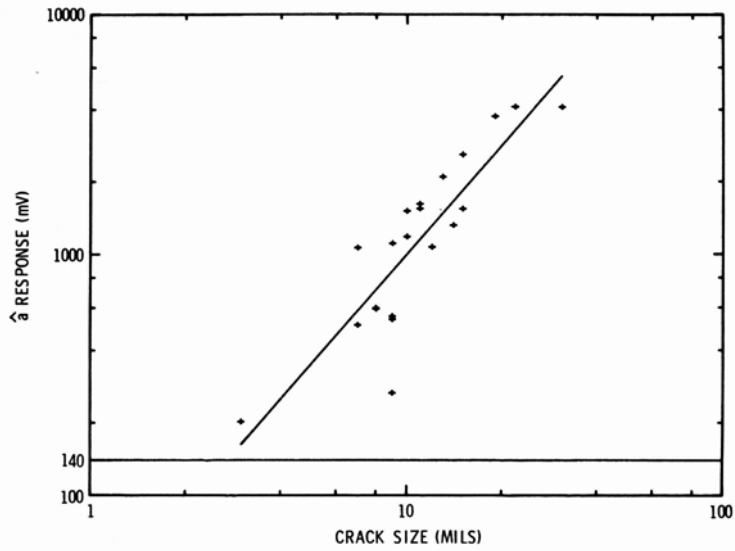


Fig. 7. \hat{a} vs a for the web/bore specimens.

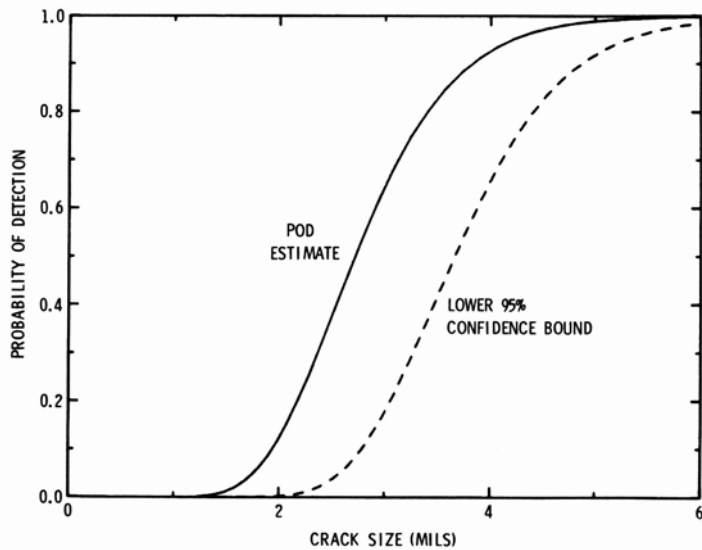


Fig. 8. POD(a) and lower 95% confidence bound for the web/bore specimens.

a data from automated systems is that the biggest cause of this scatter is due to variation associated with different cracks. The data of this study agrees with this trend.

The random error term, e , of equation (3) can be partitioned into components and the contributions of the individual components can be estimated from properly conducted experiments. In the repeat experiments of this study, one factor at a time was varied and the components of the total variation were estimated. Table 1 lists the components due to the

factors that were varied. Estimates of the variation due to cracks were 0.73, 0.32, and 0.26 for the rivet hole, bolt hole, and web/bore specimens, respectively. Thus, the crack-to-crack variation was much larger than that due to the other factors.

The POD(a) function of Figures 3, 6, and 8 included only the crack-to-crack and repeatability variation. Of the factors measured, only the transducer variability could significantly contribute to the total. For example, if it is assumed that a bolt hole is to be inspected and the transducer to be used will be chosen at random, then the proper S to be used in equation (4) would be given by

$$S^2 = (0.32)^2 + (0.13)^2$$

or $S=0.35$. On the other hand, if the bolt hole was to be inspected by the transducer used to generate Figure 6, then $S = 0.32$ is the proper measure of the scatter.

Perhaps it should also be noted that POD(a) functions were calculated for each of the experimental conditions. Resulting differences in the POD(a) functions were judged to be negligible in the retirement-for-cause application.

CONCLUSION

The conclusions drawn from these Phase I tests of the RFC/NDE system are:

- a) The lognormal model was appropriate for these highly automated inspections.
- b) The significant causes of variability in transducer response, $\hat{\alpha}$, were the cracks, the transducers, and the crack orientation with respect to the transducer.
- c) The reloading of the specimens and the changing of operators had no effect on the transducer response (and, thus, no effect on POD).
- d) The test matrix for the Phase II evaluation tests will be greatly reduced.

ACKNOWLEDGMENT

This work was sponsored by the Materials Laboratory of the Air Force Wright Aeronautical Laboratories under Contract F33615-81-C-5002 through a Systems Research Laboratories subcontract.

REFERENCES

1. Berens, A.P., and Hovey, P.W., "Flaw Detection Reliability Criteria," AFWAL-TR-84-4022, Air Force Wright Aeronautical Laboratories, Wright-Patterson Air Force Base, Ohio, April 1984.
2. Shambaugh, R.L., "Design and Execution of the RFC/NDE System Inspection Reliability Tests," these proceedings.
3. Ko, R.T., "RFC Phase I Eddy Current Reliability Tests Results," these proceedings.