AD-P000 013

# FLAW DETECTION RELIABILITY ASSESSMENT AND ANALYSIS

by Ward D. Rummel
Manager Advanced Quality Technology
and
Richard A. Rathke
Staff Quality Engineer
Martin Marietta Corporation
Denver Aerospace

## ABSTRACT

The assessment of nondestructive flaw detection reliability is complex in character due to the varied engineering and scientific disciplines involved. The evolution of nondestructive flaw detection reliability demonstration and assessment has involved varied efforts by workers in various industries, applications and environments. A significant data base has been established and has contributed to a general understanding of the elements of inspection reliability. A considerable number of analyses have been performed to effect a better understanding of the problem and to identify critical factors in both the inspection process performance and in reliability assessment.

This paper reviews principle factors in nondestructive flaw detection process performance and suggests an alternative approach to the analysis of performance data. The approach includes consideration of the conditional probability character of flaw detection and consideration for predictive modeling based on signal and noise analyses of flaw detection by instrumental techniques and by human operators.

## INTRODUCTION

Nondestructive inspection has been incorporated as an integral part of modern engineering structures design in both critical and non-critical applications. The assessment of nondestructive flaw detection reliability is complex in character due to the number of parameters that must be accounted for and to the varied disciplines involved. For critical applications, the reliability of inspection processes must be assured to provide confidence in the functional integrity and performance of critical materials, structures or components. Measurement and assessment of nondestructive inspection reliability requires multi-parameter assessment and documentation under controlled conditions. Indeed, much of the reliability data and data analyses that have been generated are confusing and appear to be contradictory. The following discussion provides an approach to the understanding and modeling of nondestructive inspection processes with respect to overall process reliability.

## THE NATURE OF INSPECTION RELIABILITY ASSESSMENT

The task of measuring inspection reliability differs from that of initial inspection selection by a shift in emphasis from the smallest flaw detected to the largest flaw missed. An inspection process constitutes an exercise in conditional probability as opposed to joint probability due to the interdependence of inspection stimuli and inspection responses. A schematic presentation of such interdependence is shown in the following:

| | | STIMULI<br>POS a | STIMULI<br>NEG n |
|---|---|---|---|
| RESPONSE | POS<br>A | M(Aa)<br>T.P.<br>P(A,a) | M(An)<br>F.P.<br>P(A,n) |
| RESPONSE | NEG<br>N | M(Na)<br>F.N.<br>P(N,a) | M(Nn)<br>T.N.<br>P(N,n) |

The outcome of the inspection test may be:

TRUE POSITIVE (T.P.),
where M(Aa) is the total number of T.P. calls;
and P(A,a) is the probability of T.P. calls.

FALSE POSITIVE (F.P.),
Where M(An) is the total number of F.P. calls;
and P(A,n) is the probability of F.P. calls.

FALSE NEGATIVE (F.N.),
where M(Na) is the total number of F.N. calls;
and P(N,a) is the probability of F.N. calls.

TRUE NEGATIVE (T.N.),
where M(Nn) is the total number of T.N. calls;
and P(N,n) is the probability of T.N. calls.

Interdependence of the matrix quantities is denoted by:

T.P. + F.N. = Total opportunities for positive calls.
F.P. + T.N. = Total opportunities for negative calls.

Therefore, only two independent probabilities must be considered in alternative inspection / decision tasks.

The SPECIFICITY of the technique or the PROBABILITY OF DETECTION of flaws may be expressed as:

$$POD = \frac{\text{T.P.}}{\text{T.P.} + \text{F.N.}} \quad \text{or} \quad \frac{\text{total positive calls}}{\text{opportunities for positive calls}}$$

Likewise, the NONSPECIFICITY of the technique of the PROBABILITY OF FALSE ALARMS may be expressed as:

$$POFA = \frac{\text{F.P.}}{\text{T.N.} + \text{F.P.}} \quad \text{or} \quad \frac{\text{total false alarms}}{\text{opportunities for false alarms}}$$

Confidence limits for the probability of detection value may be calculated from standard tables for a given sample size and calculated value from experimental sample data. This technique establishes an estimate for performance at one flaw size value, calibration level and acceptance criteria level. Data of most interest to the design engineer, nondestructive inspection engineer and systems manager is plotted as a composite of the discrete values calculated for individual operating points.

## PROBABILITY OF DETECTION (POD) CURVES

The established method of assessing and presenting inspection reliability data is by means of a probability of detection or POD curve as shown in FIGURE 1. A POD curve is generated by passing a series of specimens that contain a large number of flaws of varying flaw sizes, through a nondestructive inspection process and documenting the success in detecting all flaws. Flaws are then ordered from large to small in terms of decreasing flaw size and are grouped to provide a statistically significant sample size for analysis. Sampling to provide a 95% confidence level (MIL Handbook No. 5, B values) is attained by grouping samples into lots of 60 observations (REF 1). The point estimate of detection for the sample group is calculated by dividing the total number of opportunities for the sample group into the total number of successes (flaws detected). The point estimate (probability of detection) is plotted at the largest flaw size in the sample group. The process is repeated to generate a curve that denotes the probability of detection as a function of flaw size (FIGURE 1). This method of data presentation was introduced by Rummel et al (REF 2) in 1971 and has been adopted as the standard method for inspection reliability data presentation. The method has been used by various investigators to plot both controlled experimental data sets and uncontrolled experimental data sets. A measure of the capability of a specific inspection technique for flaw detection can be derived when the method is used to analyze controlled experimental data. A measure to the overall baseline capability for a facility, organization, etc. can be derived when the method is applied to unlike (uncontrolled) inspection operations.

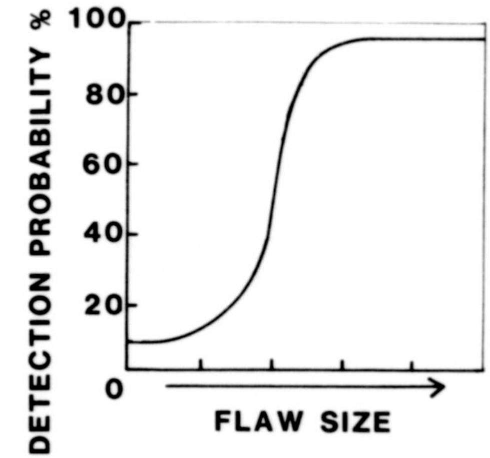


Figure 1 Typical Form of a Probability of Detection (POD) Curve

The shape of the POD curve provides a qualitative basis for assessment of the degree of control for a given data set and by the criteria for grouping of similar data sets. FIGURE 2, illustrates the POD curves for the data sets generated under varying condition of control. Curve A is typical of an inspection process that is under control and that is discriminatory (specific) to the desired output. Curve B is typical of a process that is approaching control. The mode and type of variance denote the influence of factors not accounted for in the direct correlation of process performance with flaw size. Curve C is typical of a process that is out of control but whose performance is influenced by flaw size. Curve A is worthy of further statistical rigor. Curve B is worthy of further analyses to ascertain the nature of secondary variances. Curve C is worthy of further analyses to improve the process or to provide a measure of inspection discrimination by sampling. Flaw size is a secondary variance in Curve C at the operating point for the inspection. Identification and control of the primary variance will change the nature of the data set, the specificity of the technique and the resultant POD curve.

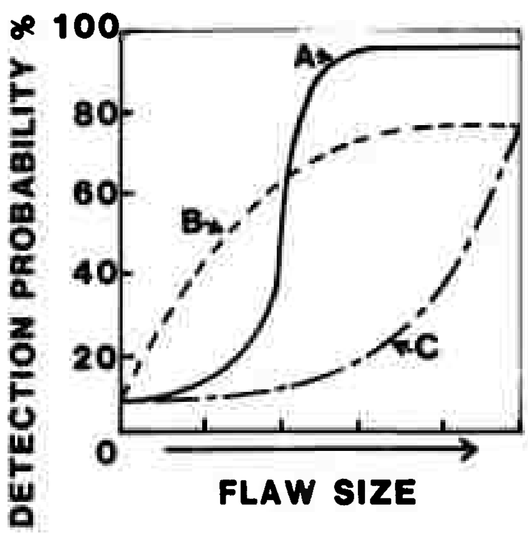


Figure 2 Typical Probability of Detection (POD) Curve Under Varying Conditions of Process Control

Each POD curve is unique to the specificity of the inspection process, the degree of control effected in the inspection process and to the nature and distribution of flaws being assessed. Rigorous use of the data in specific applications is limited to the specific process, control and flaw distribution conditions used in data generation. The cost of generation, precision in data collection and the discipline required for specific applications have fostered many attempts to generalize and model POD curve prediction. To date, no satisfactory model has been developed and some modeling attempts have contributed to the confusion in application and in data generation. Since many critical inspections are currently performed by skilled operators using manual techniques, human factors are most frequently cited as the source of unreliability. Although human factors are a primary contributor to unreliability, nondestructive test engineering and engineering management (selection and control of the right tool for the job) are proposed by the authors to be greater sources of unreliability in general applications. Such errors will not be alleviated by the automation of inspection processes. An understanding of the nature and character of inspection processes is necessary to predict and to effect improvements. Automation without understanding will only lead to multiplication of errors.

## CONSIDERATIONS FOR MODELING OF INSPECTION PROCESSES

The POD curve provides a convenient method for comparison of inspection process performance. It provides visualization of the discrimination capability of a given technique in a form that communicates to the designer, the system manager and to the nondestructive inspection engineer. The POD curve does not, however, provide an indication of the calibration performed to establish the baseline process, the acceptance criteria imposed on the process or the level of incorrect rejections (false calls) inherent to the process / application.

Selecting an exact operating point from the POD curve is difficult and has not proven to be meaningful in many applications. Variation in a point on the curve is due to variation in response of the system and variation in the reproducibility of the inspection process application. For example, flaw size measurement by nondestructive inspection processes has been shown to be variable within a technique and between techniques (REF 3,4). The distribution of response of an interrogating energy field accounts for variation in flaw sizing by nondestructive inspection processes and to part of the variation in POD curve generation. This "third dimension" of analysis must be accounted for in an inspection model. Variation in response along a POD curve is shown schematically in FIGURE 3.

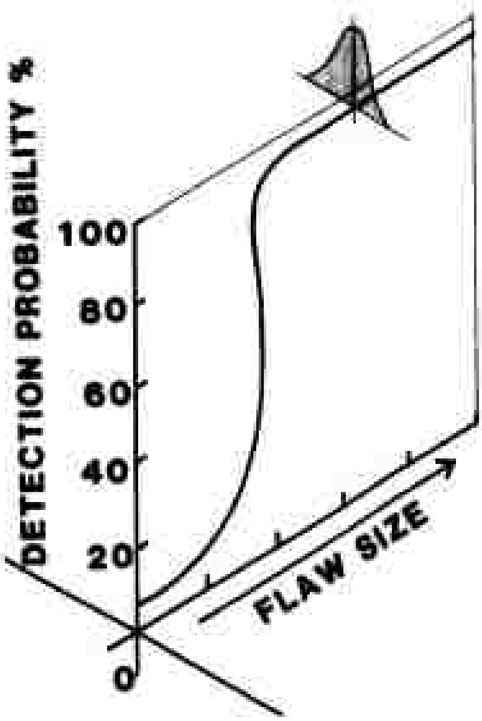


Figure 3 Interaction of the Probability of Detection (POD) Curve with the Distribution of Flaw Response

## SIGNAL / NOISE VARIATIONS AND THE POD CURVE

Consider a case where the response (signal) from a flaw is "Gaussian" in nature and where process noise is well separated from the signal (FIGURE 4). Such an inspection has high specificity for discrimination of signals that are due to flaw responses from background or process noise signals that are inherent to the process.

Consider a second case where the response (signal) from a flaw is "Gaussian" in nature with process noise signals overlapping the flaw response envelope (FIGURE 5). A threshold (signal) discrimination level may be set for this process to provide a degree of separation of flaw responses from inherent process noise. Some flaws will be missed by such a system and some false calls (rejections) will be inherent to the process. The lack of specificity will cloud the use of the process as a final discriminator.

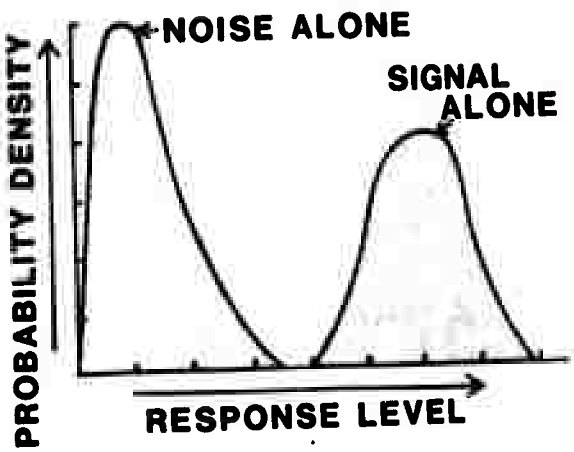


Figure 4 Signal / Noise Response for Discrimination with a High Degree of Specificity

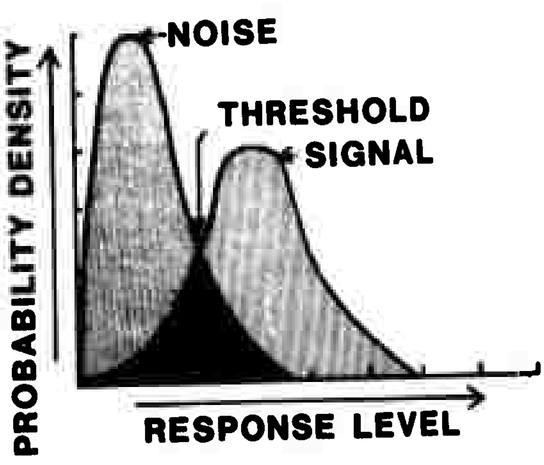


Figure 5 Signal / Noise Response for Discrimination with Overlapping Signal and Noise Stimuli

Consider a third case where the response (signal) from a flaw is coincident with the process noise signals (FIGURE 6). Such a process provides a random discrimination of flaws and is not considered to be a valid process. Indeed, better separation is likely by simple coin flipping.

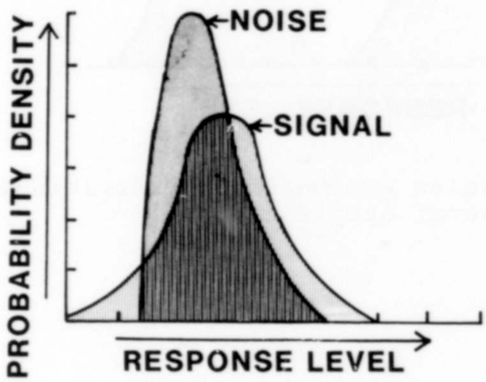


Figure 6 Signal / Noise Response for Coincident Stimuli

A POD curve typically reflects all of the variations in signal / noise response and discrimination levels as shown schematically in FIGURE 7. A continuing variation in signal / noise response is reflected by variation in the discrimination level along the POD curve. The signal / noise response and the discrimination level appear to be "common denominators" for all inspection processes and hence all POD curves generated for respective processes.

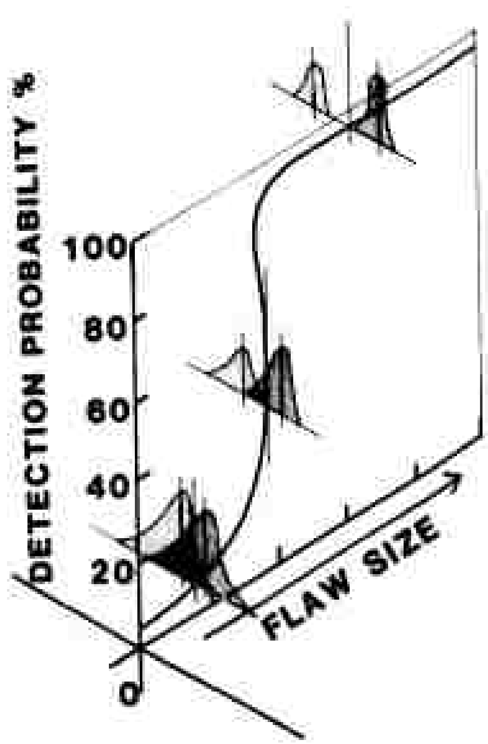


Figure 7 Interaction of Signal / Noise Discrimination with the Probability of Detection (POD)

## CRITERIA DISCRIMINATION RESPONSE AND THE POD CURVE

A second factor (common denominator) that may be shown to affect the mode and specificity of an inspection process is the criteria level selected. Consider an inspection process with a measurable separation in noise and flaw signal responses as shown in FIGURE 8. If the acceptance (discrimination) criteria level for this inspection (indicated by the vertical arrow) is set too high, some flaws will be accepted (missed) by application of the process and "EVERYBODY WILL BE UNHAPPY". If the acceptance criteria is set at a level that provides clear separation of noise signal from flaw signal, all flaws will be rejected, few false calls (rejections) will occur and "EVERYBODY WILL BE HAPPY". If the acceptance criteria is set too low, all flaws will be rejected, some false calls (rejections) will occur and "MANAGEMENT WILL BE UNHAPPY".

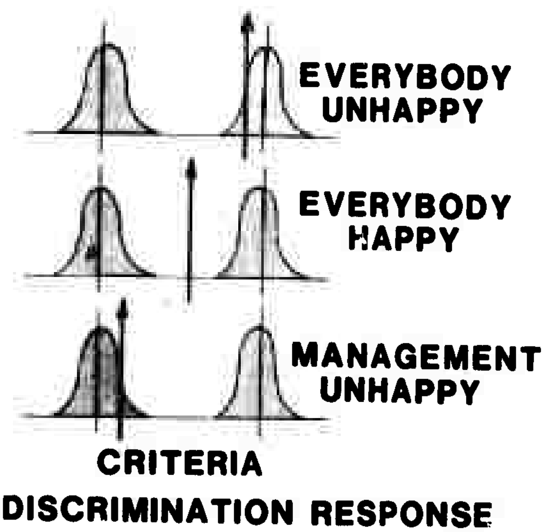


Figure 8 Influence of Acceptance Criteria Level (Vertical Arrow) on Process Discrimination (Specificity)

The process specificity and hence its POD curve may be affected by changes in the acceptance criteria level. FIGURE 9 illustrates the effects of varying levels of criteria discrimination levels on performance as denoted by the POD curve. It is important to note that the criteria discrimination level is a function not only of the rejection level imposed on an inspection process but also of the calibration reference standards and criteria used to set-up and validate inspection process performance.

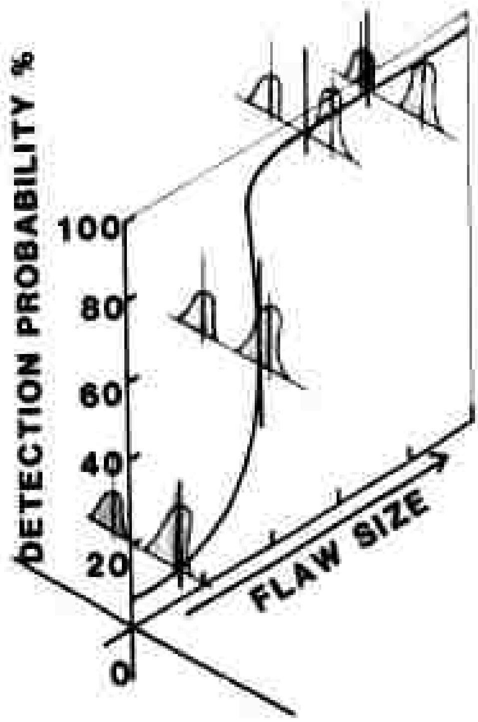


Figure 9 Interaction of Acceptance Criteria with the Probability of Detection (POD)

## INSPECTION FACTORS AND SIGNAL / NOISE RESPONSE

Variations in the condition of the flaws to be interrogated and variations in inspection conditions will affect the signal / noise function of the inspection process and its resultant discrimination level. FIGURE 10 illustrates some known and projected variations in flaws and process applications on the signal / noise response. Experimental data on the effects of variation of a single parameter on the overall signal response have been documented by various investigators. (REF 5,6) It is now clear that documentation of the calibration technique and the process noise for the inspection is necessary to account for parameter variations in a predictive model.

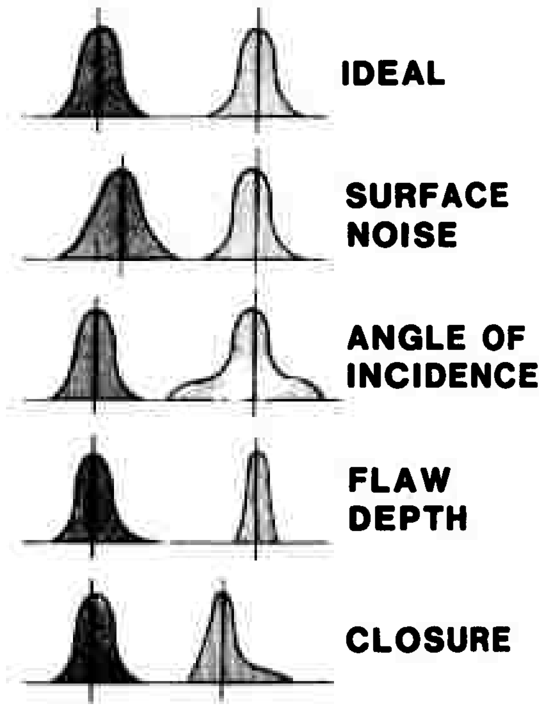


Figure 10 Interaction of Flaw Condition with Signal / Noise Discrimination

## PREDICTIVE MODELING OF INSPECTION PROCESS PERFORMANCE

Generation of POD curves and qualification of inspection processes are tedious, time consuming and expensive. At present, POD curves are unique to the inspection process and process application and cannot be used for a second process or process application. For critical applications, experimental qualification and validation are required and must be completed for each process and process application.

Current work is under way to approach predictive modeling based on "first principles" to calculate behavior and interaction of an energy field in a given application (REF 7). The approach and emphasis of this important work will provide a prediction of the performance level (POD) for an inspection process for calibration and validation at a given signal / noise level. Ultrasonic (REF 8) and eddy current (REF 9) models have been initiated as first steps in providing the engineering tools for future nondestructive process applications.

Predictive modeling would be of significant advantage in both the qualification of additional inspection processes and in reconsideration of current processes. Consider the case of cracks emanating from a radius area in a slot as shown in FIGURE 11. An eddy current inspection had been developed and qualified for cracks emanating from the center of the radius. The inspection consisted of inserting an eddy current probe, with a small ferrite core, into the radius area such that it touched the center of the radius in a plane passing through the center of curvature of the radius. After qualification and validation, crack intiation was discovered at both points of tangency of the radius area. Predictive modeling / analysis tools could have been used to calculate the size of cracks that could be reasonably detected by the center probe technique. Actual requalification and validation were necessary to establish the performance level with the analysis tools that are currently available.

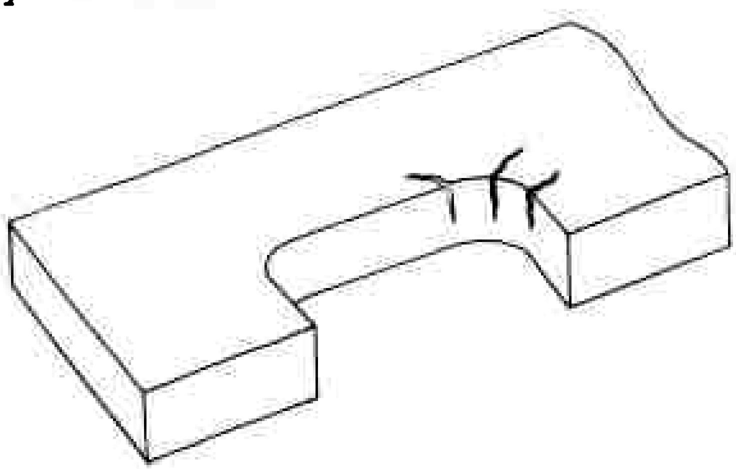

Figure 11 Variation in Location of Service Cracks in Typical Engineering Hardware

## HUMAN FACTORS VARIATIONS ON INSPECTION RELIABILITY

The impact of human factors on inspection reliability have been discussed by various investigators and have been considered to be a primary factor in "unreliability" of a given inspection process. It is therefore important to separate and measure human factors response and variations in application of inspection processes. Recent work by Swets and Pickett (REF 12) provide a logical basis for assessment of the human factors variable in an inspection process. The technique consists of establishing an inspection process with known variation in signal response and known discrimination response criteria. By repetitive measurement of human response / discrimination to the signal stimulus, a performance level may be established and used in prediction for a variety of applications. Swets and Pickett propose development of a "RELATIVE OPERATING CHARACTERISTICS (ROC)" curve (FIGURE 12) as the method of displaying and analyzing human factors data. This curve visualizes the human factors contribution at an established signal / noise level, and various discrimination levels, for a single point on the POD curve. The primary advantages of this method are the use of signal / noise and discrimination criteria as factors in assessment, and in previous use of the method in similar inspection / detection processes.

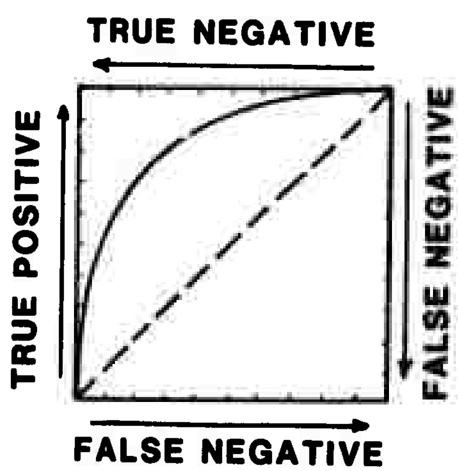


Figure 12 Typical Relative Operating Characteristic (ROC) Curve (REF 10)

## CONCLUSIONS

Flaw detection reliability assessment is indeed a complex process that requires consideration of many factors in both application processes and in assessment. Signal / noise response at a given discrimination (criteria) level have been introduced as "common denominators" to both characteristic inspection process performance and human factors performance. The combined performance levels may be summarized by plotting a probability of detection curve (POD) curve for the inspection process and application. The human factors contribution to the output must be minimized for "controlled data" that is used for assessment of the capability of a process.

Reliable flaw detection may be effected by knowledge of the nature and boundary conditions for signal response in a given inspection task. Such analysis is necessary to provide the nondestructive inspection engineering that is necessary for application to critical inspection processes. Experience by the authors has shown that inspection process qualification is probable for those applications where a consistent response is obtained at a minimum signal to noise response of 3 to 1 for flaws one-half the minimum size required by design acceptance criteria.

Progress is being made for predictive inspection modeling based on "first principles" of energy interaction and scattering. Such techniques will be a primary tool for all future nondestructive inspection engineering analyses. Process modeling, together with human factors modeling will provide the necessary tools for improvement of productivity and for automation of inspection processes in future applications.

## REFERENCES


1. Beyer, W.H., Handbook of Tables for Probability and Statistics, 2nd Edition, CRC Press, Inc., Boca Raton, Florida, 1966.

2. Rummel, W.D., Todd, P.H. Jr., Frecska, S.A. and Rathke, R.A., "The Detection of Fatigue Cracks by Nondestructive Testing Methods", NASA CR-2369, February 1974.

3. Rummel, W.D. and Rathke, R.A., "Detection and Measurement of Fatigue Cracks in Aluminum Alloy Sheet by Nondestructive Evaluation Techniques," Transactions, Second Annual ASM Materials/Design Forum on Prevention of Structural Failure, April 9-11, Port St. Lucie, Florida.

4. Knorr, E., "Reliability of the Detection of Flaws and of the Determination of Flaw Size", AGARD Fracture Mechanics Survey, AGARD - AG - 176, January, 1974.

5. Martin, B.G. and Adams, C.J.,"Detection of Lack-of-Fusion in Aluminum Weldments by Ultrasonic Shear Waves," Materials Research and Standards, v.6, no. 9, September, 1966, 440-442.

6. Corbly, D.M., Packman, P.F., and Pearson, H.S., "The Accuracy and Precision of Ultrasonic Shear Wave Flaw Measurements As a Function of Stress on the Flaw", Materials Evaluation, v. 30, no.5, May 1970, 104-110.

7. Thompson, D.O., Review of Progress in Quantitative NDE, Ames Laboratory, Iowa State University, August 1982.

8. Fertig, K. Jr. and Richardson, J.M., "Computer Simulation of Probability of Detection", Review of Progress of Quantitative NDE, Ames Laboratory, Iowa State University, August 1982, 20-21.

9. Bahr, A.J. and Cooley, D.W., "Analysis and Design of Eddy Current Measurement Systems", Review of Progress of Quantitative NDE, Ames Laboratory, Iowa State University, August 1982, 28-29.

10. Swets, J.A. and Pickett, R.M., Evaluation of Diagnositic Systems, Academic Press, New York, 1982, 19-26.