3.1 NDI Demonstration Of Crack Detection Capability

Non-destructive inspection (NDI) methods are commonly used to determine the condition of a structure during production and at in-service inspections. Detailed descriptions of the principles and use of these methods are generally available (see, for example, Nondestructive Testing Handbook, Volume Ten [1996]). The objectives of this subsection are to briefly describe and compare the common NDI methods and to discuss the statistically based demonstration programs that are required to quantify the detection capability of an NDI system. A distinction between inspections for cracks and for corrosion is made for capability evaluations even though the physical principles of the techniques are common.

3.1.1 NDI Methods

There are six commonly-used NDI techniques, and two others are expected to receive widespread acceptance and application. These eight methods are visual, liquid penetrant, eddy current, ultrasonic, magnetic particle, radiography, thermographic and acoustic emission inspections. Each of these is briefly described in the following paragraphs.

3.1.1.1 Visual Inspection

In a sense, all inspections in which the find/no find decision is made by a human are visual. However, in categorizing NDI methods, visual inspections are generally interpreted as inspections in which the inspector is aided, at most, by optical devices, such as magnifying glasses and mirrors. In the context of JSSG-2006, the in-flight evident, ground evident, walk-around evident, and special visual inspections are all visual inspections and play an important role in the maintenance of structural integrity through damage tolerance.

Visual inspections are the most common and the most economical of the inspections to perform. However, visual inspections are also the least reliable in terms of the size of the cracks that can be detected. Since the efficacy of visual inspections is highly dependent on the alertness and acuity of the inspector, an additional level of human factors is introduced in discerning the physical attributes of a crack from its environment.

3.1.1.2 Liquid Penetrant Inspection

Liquid penetrant inspection is a non-destructive method for finding discontinuities that are open to the surface of parts fabricated from essentially nonporous materials. After cleaning the surface, the penetrant is applied and will seep or be drawn into various types of minute surface openings. The excess penetrant is removed and a developer is applied which highlights the cracks under ultraviolet light. The process is well-suited for the detection of all types of surface cracks, laps, porosity, shrinkage area, laminations, and similar discontinuities.

Indications of cracks can be found regardless of size, configuration, internal structure, or chemical composition of the workpiece being inspected and regardless of the orientation of the crack to the workpiece.

Liquid penetrant inspections are relatively simple and inexpensive (as compared to the other NDI methods) and can be applied to a broad range of materials. Very small cracks can be found. However, they can only detect surface cracks and their effectiveness can be adversely influenced by surface coatings, surface roughness, and porosity. Extreme care is required in pre- and post-inspection cleaning and, in some cases, etching may be required prior to inspection.

3.1.1.3 Eddy Current Inspection

The principles of electromagnetic induction are used in eddy current inspections to detect surface and near-surface cracks in electrically-conductive metals. When an electrically-conductive material is subjected to an alternating magnetic field, small circulating electric currents are generated in the material. Since these eddy currents are affected by variations in conductivity, magnetic permeability, mass, and material homogeneity, the conditions that affect these characteristics can be sensed by measuring the eddy current response of the material. In practice, eddy currents are induced in the part to be inspected with a coil carrying an alternating current. The induced eddy currents generate their own magnetic field, which interacts with the magnetic field of the exciting coil, and changes the impedance of the exciting coil. By measuring the impedance of the exciting coil, or a separate indicating coil, the inspector can infer the presence of cracks in the material.

An important use of the eddy current NDI method has been in the detection of fatigue or stress corrosion cracks around fastener holes after the cracks have grown beyond the fastener head. Special bolt hole probes have also been devised for use after the fastener has been removed for locating cracks emanating from the wall of the fastener hole. This inspection process has been automated to remove operator influence, speed inspections, and produce a permanent inspection record.

Eddy current methods do not require contact with the specimen or clean up, and are generally faster than liquid penetrant and radiographic methods. Although eddy current methods can detect both surface and subsurface cracks, the depth of inspection below the material surface is limited (approximately 0.25 in.). Since eddy currents are influenced by many material variables, masked or false indications can easily be caused by sensitivity to part geometry, lift-off, edge effects and permeability variations. Finally, eddy current methods require well-trained operators to man the test instruments and reference standards are necessary.

3.1.1.4 Ultrasonic Inspection

Ultrasonic inspection uses high frequency sound waves as a probing medium to detect subsurface, as well as surface cracks. The sound waves travel through the part with attendant energy loss and are reflected at material-crack interfaces. Ultrasonic inspection devices detect cracks by monitoring one or more of the following: (a) reflection of energy from interfaces or discontinuities within the metal; (b) time of transit of a sound wave through the test piece; and (c) attenuation of the beams by absorption and scattering within the test piece.

Ultrasonic inspection is one of the most widely used NDI methods. Cracks, laminations, shrinkage cavities, bursts, flakes, pores, bonding faults, and other discontinuities that act as metal-gas interfaces can be detected. Inclusions and other non-homogeneity in the metal being inspected can also be detected by causing partial reflection or scattering of the wave, even though they may not act as a metal-gas interface. Although the primary application of ultrasonic inspection in metals is the detection and characterization of internal cracks, it is also used to detect surface cracks, define bond characteristics, measure extent of corrosion and, (much less frequently) determine physical properties such as structure, grain size, and elastic constants. The penetrating power of ultrasound waves allows the detection of cracks deep within a part. Due to the sensitivity of the instruments, very small cracks can be detected but, if the gain is set too high, at the expense of many false indications. Ultrasonic methods provide greater accuracy than other NDI methods in determining the position of internal cracks, estimating their size, and characterizing their orientation, shape and nature. The limitations of ultrasonic methods are governed by the requirement for experienced technicians, the difficulty in developing inspection procedures, the need for reference standards for equipment calibration, and the physical limitations of the hardware. Since couplants (light oil or water) are needed to provide effective transfer of ultrasonic wave energy between transducers and material, parts that are rough or irregular in shape are difficult to inspect. Similarly, parts that are very small are difficult to inspect. Finally, since discontinuities in a shallow layer immediately below the surface may not be detectable, inspection results of very thin components are questionable.

3.1.1.5 Magnetic Particle Inspection

Magnetic particle inspection is effective in the detection of surface and near-surface cracks in ferromagnetic parts. The inspection is accomplished by inducing a magnetic field in the part and applying either a dry magnetic powder or a liquid suspension of iron particles to the surface being inspected. Defects in the part cause local bipolar perturbations in the magnetic field which attract the magnetic particles, producing visible indications by color contrast or by fluorescence under "black light". The magnetically-held particles form the outline of the discontinuity and generally indicate its location, size, shape, and extent to an experienced inspector.

The magnetic particle method is a relatively fast and inexpensive method for locating small and shallow surface cracks in ferromagnetic materials. Discontinuities that do not break the surface are detectable, but deeper cracks must be larger to be found. Elaborate pre-cleaning is not necessary, but thin coatings of paint or other non-magnetic coverings, such as plating, adversely affect the sensitivity of this inspection technique. Following the inspection, the material must often be de-magnetized, and post-cleaning to remove the clinging magnetic particles is usually necessary. This NDI method can be used only on ferromagnetic materials, which include most of the iron, nickel and cobalt alloys. Many of the precipitation-hardening steels, such as 17-4PH, 17-7PH, and 15-4PH stainless steels, are magnetic after aging. Non-ferromagnetic materials that cannot be inspected by this method include aluminum, magnesium, copper, and titanium alloys and austenitic stainless steels.

3.1.1.6 Radiographic Inspection

Radiographic NDI is based on the differential absorption of penetrating radiation by the structure being inspected. In conventional radiography, the object is bombarded by a beam of X-rays and the portion of the radiation that is not absorbed by the object impinges on a sheet of film. The unabsorbed radiation exposes the film emulsion similar to the way light exposes film in photography. Development of the film produces an image that is a two-dimensional "shadow picture" of the entire volume of the object. Variations in density, thickness, and composition of the object being inspected cause variations in the intensity of the unabsorbed radiation and appear as variations in shades of gray in the developed film. Evaluation of the radiograph is based on a comparison of the differences in photographic density with known characteristics of the object or with standards derived from radiographs of similar objects of acceptable quality.

Radiographic inspection provides the capability to probe the internal characteristics of materials and components. It can disclose structural weaknesses, assembly errors, and mechanical malfunctions, as well as revealing voids, long cracks, and other material anomalies. Radiography is, however, expensive, slow, and not sensitive to detecting certain type cracks. Cracks cannot be detected unless they are parallel to the radiation beam. Tight cracks in thick sections cannot usually be detected even when properly oriented. Laminations are almost always non-detectable. Minute discontinuities such as inclusions in wrought material, flakes, microporosity and microfissures cannot be detected unless they are sufficiently segregated to produce a detectable gross effect. Finally, due to the hazards of exposure to X-rays, strict controls are required to prevent biological damage to the inspectors.

3.1.1.7 Thermographic Inspection

Thermographic inspection uses relative differences in heat transmission to detect internal features and defects, such as delaminations in layered materials. In active thermography, heat is applied to the object under test and surface temperatures are monitored by an infrared camera as the heat propagates through the object. In the reflection method, heat is applied to the surface that is monitored; relatively warm areas indicate possible internal defects. In the transmission method, heat is applied to the opposite side of a panel from a detector, and relatively cooler areas will indicate areas of poor thermal transmission. Heat may be applied by a laser, warm air, heat lamps, flash lamps, or other methods. While heating is most common, cooling may also be used to create thermal transients within the material. Passive thermography, with no external heat source, may be used if thermal contrasts are produced within the object under test by other means, such as electrical heating at a poor solder joint.

Thermographic methods are most appropriate for use with materials that have low thermal conductivity, such as ceramics and polymers. Heat propagates more slowly in these materials, which decreases the image acquisition rate needed from the infrared camera. In addition, interference with the thermal excitation can obscure near-surface data, with "near-surface" measured in time of thermal transmission, so that much less data is lost when the heat is propagating slowly. High emissivity surfaces radiate heat better and, therefore, produce better sensitivity. Coatings, such as a flat black paint, may be applied to low emissivity or reflective surfaces to increase emissivity. Flaws, such as delaminations that are perpendicular to the propagation of thermal energy through the object, are the best candidates for detection by thermography. Other flaws that disrupt heat flow, such as fluids trapped in honeycomb materials, can also be detected with relative ease. The resolution of this method decreases with depth, because the thermal energy is conducted in all directions, not just directly through the material. Surface flaws, such as cracks, may be detected if heat can be forced to propagate along the surface of the material. Thickness or composition variations may be detected by transmission thermography.

Thermography has a long history, but has not achieved the widespread use of other methods, such as ultrasonic, eddy current, and radiography. Disadvantages of this method include the expense of equipment, the reliance on surface emissivity, and the generally low signal-to-noise ratio. Advantages include area inspection nature of the technology, speed, noncontact nature, and versatility. Currently, thermographic methods are used for delamination detection in layered composites, coatings evaluation, honeycomb inspection, thermal barriers, bond evaluation, and thickness evaluation. Improvement in the sensitivity of infrared detectors and better thermal sources indicate that the use of thermographic methods will increase as the supporting technologies continue to mature.

3.1.1.8 Acoustic Emission Inspection

Acoustic emission (AE) is the term used for dynamic stress waves that are created within a material due to the application of a force. Some examples are the sound of fibers breaking when a piece of wood is bent, high-frequency stress waves created when a crack grows in a metal structure undergoing mechanical fatigue, and the pulse of stress waves emanating from the impact site of a meteorite colliding with a spaceship hull. AE differs from most of the other NDI methods in that no directed energy is put into the test object. Whole-body forces create the localized stress waves that propagate through the test object to AE sensors.

AE NDI is done by placing multiple acoustic sensors on the object being inspected and then recording and correlating the signals generated when stress waves reach the sensors. The sensors typically are responsive to acoustic frequencies between 50 kHz and 1 MHz. The lower limit is important in order to limit acoustic noise, although it should be noted that common objects such as jingling car keys or grinding wheels produce acoustic energy above 100 kHz. The upper limit is strongly dependent on the bandwidth of the AE sensor. Occasionally, AE tests utilize sensors with the upper limit extending into the 2-3 MHz range. The sensors are connected to AE instruments that amplify, filter, store, and process the signals produced by the sensors. Typical results from AE tests are the number of AE "events" recorded; the energy, time, and duration of each event; and the location of the event within the test object.

Some advantages of AE NDI are: 1) the method is sensitive to stress waves emanating from anywhere within the test object; the sensors do not have to be focused or scanned across the object; 2) triangulation of the time of detection of the stress wave at different sensors allows identification of the location of the emission, and 3) sensors can be placed on objects with very limited access.

Disadvantages of AE NDI are: 1) the instrumentation is expensive, 2) appropriate signal processing to eliminate unimportant signals can be complicated, 3) large amounts of data often are generated, creating data storage problems.

3.1.1.9 NDI Methods Summary

Figure 3.1.1 summarizes and compares attributes of the five principal non-visual NDI methods that are in widespread use. This subjective comparison describes the types of defects that can be characterized, the structural applications, the advantages, and limitations of each of the methods. For damage tolerance considerations, the key characteristic of an NDI system is the size of the flaws that can be missed when the system is applied in the field. Quantifying inspection capability in terms of flaw size is referred to as inspection or NDI reliability. Because of the many differences in material and geometry of structural details and the many approaches to the application of any of the methods, there is no single characterization of capability in terms of a reliably-detected crack size for any of the methods. Further, because of the difficulty and cost of quantifying NDI reliability, relatively few capability demonstrations have been conducted. Only very general statements can be made comparing the NDI reliability of the five methods.

Because of the random nature of inspection response to flaws of ostensibly the same size, NDI capability is characterized in probabilistic terms and estimated using statistical methods. In particular, NDI reliability is quantified in terms of the probability of detection as a function of flaw size, POD(a). There is no practical flaw size for which there is 100 percent assured detection. For damage tolerance applications in the aircraft industry, it has become customary to characterize inspection capability in terms of the crack size for which there is a 90 percent probability of detection, the a_{90} crack size. To reflect the statistical uncertainty in the estimate of a_{90}, a 95 percent confidence bound can be calculated yielding the $a_{90/95}$ crack size characterization of capability. There is 95 percent confidence that at least 90 percent of all cracks of size $a_{90/95}$ will be detected. The reliably detected crack size for a system is usually taken to be either a_{90} or $a_{90/95}$. Note that cracks smaller than $a_{90/95}$ are readily detected by the NDI systems since POD(a) functions for production inspections increase over a relatively large crack size region. Typically, the 50 percent detectable crack size is less than half the a_{90} crack size for a non-automated inspection. Subsection 3.1.2 describes in considerable detail the approach to demonstrating NDI reliability for an application.

Method	Measures or Defects	Applications	Advantages	Limitations
Magnetic Particles	Surface and slightly subsurface defects; cracks, seams, porosity, inclusions Permeability variations Extremely sensitive for locating small, tight cracks	Ferromagnetic materials, bar, forgings, weldments, extrusions, etc.	Advantage over penetrant in that it indicates subsurface defects, particularly inclusions Relatively fast and low cost May be portable	Alignment of magnetic field is critical Demagnetization of parts required after tests Parts must be cleaned before and after inspection Masking by surface coatings
Liquid Penetrant	Defects open to surface of parts; cracks, porosity, seams, laps, etc. Through-wall leaks	All parts with non-absorbing surfaces (forgings, weldments, castings, etc.) Note: Bleed-out from porous surfaces can mask indications of defects	Low cost Portable Indications may be further examined visually Results easily interpreted	Surface films, such as coatings, scale, and smeared metal may prevent detection of defects Parts must be cleaned before and after inspection Defect must be open to surface
Ultrasonic (0.125 MHz)	Internal defects and variations, cracks, lack of fusion, porosity, inclusions, delaminations, lack of bond, texturing Thickness or velocity Poisson's ratio, elastic modulus	Wrought metals Welds Brazed joints Adhesive-bonded joints Nonmetallics In-service parts	Most sensitive to cracks Test results known immediately Automating and permanent recording capability Portable High penetration	Couplant required Small, thin, complex parts may be difficult to check Reference standards required Trained operators for manual inspection Special probes
Eddy Current (200 Hz to 6 MHz)	Surface and subsurface cracks and seams Alloy content Heat treatment variations Wall thickness, coating thickness Crack depth Conductivity Permeability	Tubing Wire Ball bearings "Spot checks" on all types of surfaces Proximity gage Metal detector Metal sorting Measure conductivity in % IACS	No special operator skills required High speed, low cost Automation possible for symmetrical parts Permanent record capability for symmetrical parts No couplant or probe contact required	Conductive materials Shallow depth of penetration (thin walls only) Masked or false indications caused by sensitivity to variation, such as part geometry, lift-off Reference standards required Permeability variations
Radiography (X-rays-film)	Internal defects and variations; porosity, inclusions; cracks; lack of fusion; geometry variations; corrosion thinning Density variations Thickness, gap and position Misassembly Misalignment	Castings Electrical assemblies Weldments Small, thin, complex wrought products Nonmetallics Solid propellant rocket motors Composites	Permanent records; film Adjustable energy levels (5 kv-25 mev) High sensitivity to density changes No couplant required Geometry variations do not effect directions of X-ray beam	High initial costs Orientation of linear defects in part may not be favorable Radiation hazard Depth of defect not indicated Sensitivity decreases with increase in scattered radiation

Figure 3.1.1. Summary and Comparison of Principal Nondestructive Testing Methods [Walker, et al., 1979]

Table 3.1.1 presents approximate lower limits of reliably-detected crack sizes for the NDI methods in common use in the aircraft industry. These limits are achievable on some structures by well-trained inspectors working in a good production environment. Because the crack sizes of Table 3.1.1 represent the limits of the methods, such capabilities must be demonstrated before use in a damage tolerance based inspection schedule. Note that most routine inspections are not designed for these target crack sizes.

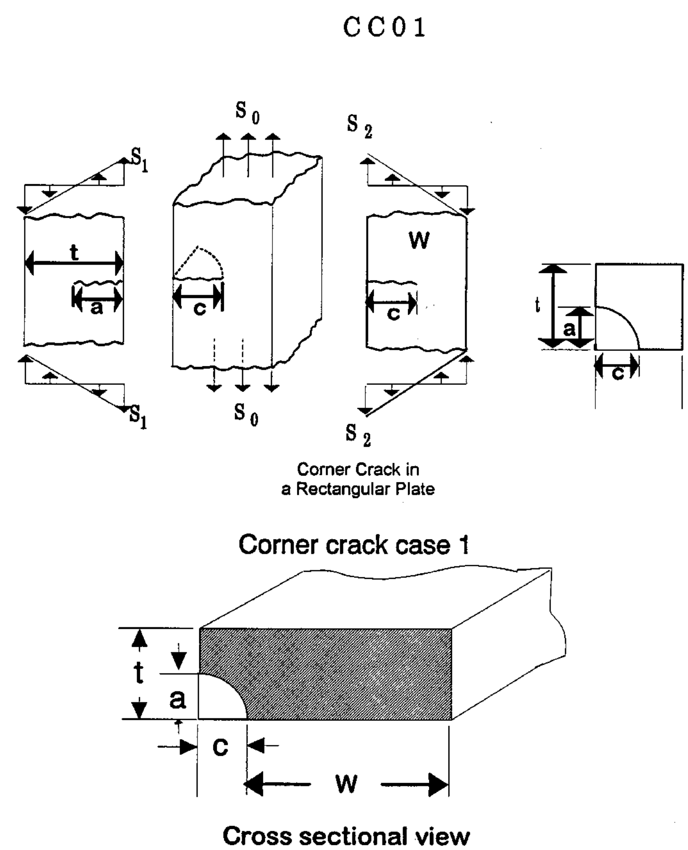
Table 3.1.1. Approximate Limits of Reliably Detectable Crack Sizes

Method		Location	Dimension	Size (in.)
Eddy Current	Manual	Near Surface	Length	0.030-0.040
	Semi-Automated	Near Surface	Length	0.020-0.030
	Automated	Near Surface	Length	0.005-0.010
Ultrasonic	Manual	Subsurface	FBH*	0.032-0.064
	Automated	Subsurface	FBH*	0.016-0.032
Fluorpenetrant	Manual	Surface	Length	0.075-0.100
	Automated	Surface	Length	0.060-0.075
Magnetic Particle	Manual	Near Surface	Length	0.010-0.020

*FBH – capability based on flat bottom holes

There have been a number of demonstrations of NDI reliability for different structures and NDI methods. An early compilation of such results can be found in Yee, et al. [1974], but the analysis methods for POD data were still evolving at that time and the quoted $a_{90/95}$ values in this report are not compatible with those of more recent vintage. A major study sponsored by the United States Air Force was that of a program known as "Have Cracks, Will Travel" [Lewis, et al., 1978]. This study evaluated inspection capability at Air Force facilities and demonstrated the need for improving NDI reliability. More recently, Rummel and Matzkanin [1997] have produced a data book that lists POD results for aluminum and titanium flat plates and panels and steel turbine engine bolt holes. Among others, this data book contains the results of NDI demonstrations produced by the Aging Aircraft NDI Development and Demonstration Center at Sandia National Laboratories (see for example Spencer & Schurman [1995] and those of an AGARD round robin [Fahr, et al., 1995]). A number of POD evaluations have been performed on the Retirement for Cause Eddy Current Inspection System (RFC/ECIS) for the inspection of turbine engine components but the results of these evaluations have not been released.

Another quantitative comparison of the various NDI methods is represented by the default reliably detected crack sizes that can be used in structural design. See, for example, NASA/FLAGRO Version 2.03, in which such default crack sizes are listed for 24 different crack types and the five common NDI methods. As an example of such default reliably detected crack sizes, Figure 3.1.2, from Rummel & Matzkanin [1997] and NASA/FLAGRO Version 2.03, presents one of the crack types and the corresponding default crack sizes.



Crack Case	NDE Inspection Technique or Flaw Size Criterion	Thickness Range (in.)	Crack depth, a	Size (in.) crack length, c
CC01, (edge)	EC	t > 0.075	0.075	0.075
	P	t > 0.100	0.100	0.100
	MP	t > 0.075	0.075	0.075
	U	t > 0.100	0.100	0.100

Notes:
EC = eddy current (ET)
P = dye/fluorescent penetrant (PT)
MP = magnetic particle (MT)
U = ultrasonic (UT)

Figure 3.1.2. Standard NDE Flaw Sizes for STS Payloads – Edge Corner Cracks [Rummel & Matzkanin, 1997]

3.1.2 NDI Capability Evaluation for Cracks

While all of the NDI systems are capable of finding "small" cracks, damage tolerance analyses are based on the largest crack that might be in the structure after an inspection. Thus, the focus of NDI capability evaluation for damage tolerance is the largest crack that might be missed at an inspection. NDI techniques do not always produce a correct indication when applied by inspectors to cracks of the same size. The ability and attitude of the operator, the geometry and material of the structure, the environment in which the inspection takes place, and the location, orientation, geometry and size of the crack all influence the chances of detection. When considering the efficacy of an NDI system as a function of only crack size, uncertainty is introduced as a result of ignoring the other factors. This uncertainty is quantified in terms of the probability of detection (POD) of cracks of a fixed size. POD(a) is defined as the proportion of all cracks of size a that will be detected by the NDI system when applied by representative inspectors to the population of structural elements in a defined environment. At present, demonstrating the capability of an NDI system for a specific application requires a carefully controlled experiment with a valid statistical analysis of the resulting data. Figure 3.1.3 presents an example POD(a) function with a 95 percent confidence bound for a liquid penetrant inspection of turbine engine blades. Each data point represents the proportion of times cracks of the indicated size were detected.

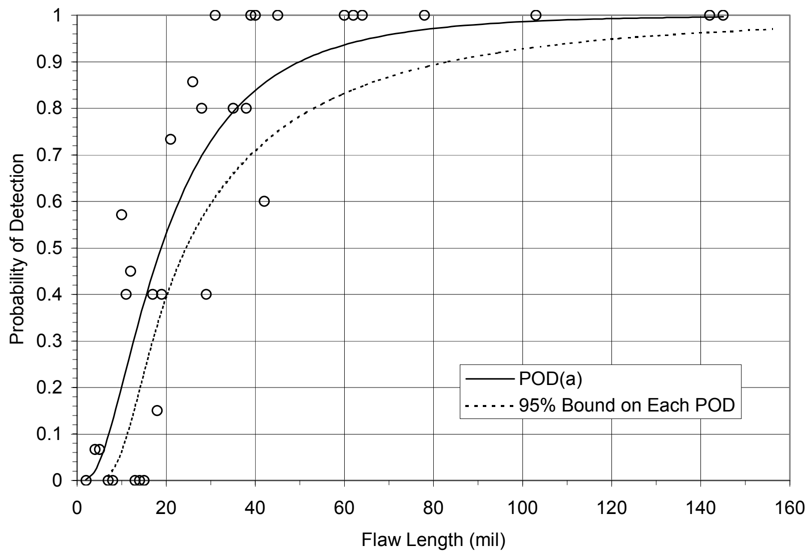


Figure 3.1.3. Example POD(a) Curve with Confidence Bound for Liquid Penetrant Inspections

The statistically-based characterization of NDI capability has two significant ramifications. First, for a given NDI application, the true probability of detection as a function of crack size (or for a single crack size) will never be known exactly. The capability of an NDI system can only be demonstrated by inspecting representative structures with known crack sizes. The true POD(a) function is estimated from the responses to the inspection stimuli or by the observed percentages of correct positive indications. The estimated POD(a) is subject to the statistical variation that can result from all of the uncontrolled factors that lead to variability in positive indications for all cracks of a particular size. However, statistical methods (which depend on the experimental procedure) are available which yield confidence limits on the true probability of detection. Protection against making a wrong decision on the basis of a set of non-typical results is provided by the confidence limits.

Second, in the real-world structural integrity problem, no inspection procedure will provide 100 percent assurance that all cracks greater than some useful size will be detected. Current NDI capabilities at the short crack lengths of interest in aircraft applications dictate that a reliably detectable crack size, can only be specified in terms of a size for which a high percentage of cracks will be detected. To reflect the statistical uncertainty, a confidence bound is often placed on this estimate of crack size. Such single crack size characterizations of NDI capability are expressed in terms of the crack sizes for which there is at least a given POD at a defined level of confidence (the POD/CL crack size). Such characterizations provide a stand-alone measure of the NDI system that is valid for applications represented by the demonstration test conditions. For example, JSSG-2006 states that smaller initial crack sizes can be used for slow crack growth structures if it can be shown that there is 95 percent confidence that at least 90 percent of all cracks of the smaller size will be detected by the manufacturers' NDI system.

There are three major limitations associated with the POD/CL type characterization:

1) The choice of particular POD and confidence limits has been made on a rather arbitrary basis. For example, 90/95 values were selected for JSSG-2006 recommended crack sizes even though there is no real interest in a crack length that is detected only 90 percent of the time. Rather, 90/95 limits were selected because higher POD or confidence limit values would have required much larger sample sizes in the demonstration programs for the analysis methods being used. The 95 percent confidence limit is assumed to provide the required degree of conservatism.
2) A POD/CL limit is not a single, uniquely defined number but, rather, is a statistical or random quantity. Any particular POD/CL estimate is only one realization from a conceptually-large number of repeats of the demonstration program. Berens & Hovey [1981] showed there can be a large degree of scatter in these POD/CL estimates and the scatter depends on the POD function, analysis method, POD value, confidence level and number of cracks in the demonstration program.
3) The POD/CL characterization is not related to the size of cracks that may be present in the structure after an inspection. To calculate the probability of missing a large crack requires knowledge of both POD(a) for all cracks sizes and the distribution of the sizes of the cracks in the population of structural details being inspected.

MIL-HDBK-1823 and Berens [1988] present in considerable detail an acceptable approach to demonstrating NDI capability in terms of a POD/CL characterization. Other approaches are also in use. After a brief description of the design of NDI capability experiments, the following paragraphs present a description of the analyses that are in current use for calculating POD/CL limits.

3.1.2.1 Basic Considerations in Quantifying NDI Capability

There are two distinct strategies for quantifying NDI capability for damage tolerance analyses. These are: a) estimating POD(a) as a function of crack size and b) demonstrating capability for a fixed crack size. To estimate a POD(a) function, the structural details to be inspected would comprise a range of crack sizes in the expected domain of increasing POD. A parametric equation is assumed for the POD(a) function, the parameters of the equation are estimated from the inspection results, and the statistical properties of the estimates are used to place a confidence limit on the selected detection probability. To demonstrate capability for a fixed crack size, only cracks of the size of interest are inspected. The proportion of the cracks that are detected is the estimate of POD (for cracks of that size) and binomial theory is used to place a lower confidence bound on the detection probability. Because of the greater utility of the POD(a) function, the approach based on estimating the entire function is preferred by many, including the Air Force [MIL-HDBK-1823]. The fixed crack size approach is used by NASA for qualifying the inspection capability of vendors [Salkowski, 1993]. It might be noted that a binomial approach to estimating POD as a function of crack size was extensively considered in the 1970's, but later abandoned. Very large numbers of cracked specimens were needed to ensure an adequate sample size within reasonably small intervals of crack size.

The analysis of data for demonstrating capability at a fixed crack size using the binomial approach will be discussed, but the major thrust of the capability evaluation is focused on estimating the POD(a) function. Similar considerations apply to the planning of both types of capability demonstrations.

Inspection results are recorded in two distinct formats and the format determines the analysis method to be used in modeling the POD(a) function. When the results of an inspection are expressed only in terms of whether or not a crack was detected, the data are known as find/no find, hit/miss, or pass/fail data. Such dichotomous inspection results are represented by the data pair (a_i, Z_i) where a_i is the size of the i^{th} crack and Z_i represents the outcome of the inspection of the i^{th} crack: $Z_i = 1$ for the crack being found (hit or pass) and $Z_i = 0$ for the crack not being found (miss or fail). Examples of such data would be the results of visual, magnetic particle, or fluorescent penetrant inspections or any inspection for which the magnitude of the response to the inspection stimulant was not recorded. POD(a) analysis for data of this nature is often called hit/miss or pass/fail analysis. Maximum likelihood estimates of the parameters of the POD(a) model are obtained from the (a_i, Z_i) data. Asymptotic properties of the maximum likelihood estimates are used to calculate the confidence bound on the estimate of the reliably detected crack size.

When the results of the inspection are based on the quantified magnitude of a response to the inspection stimulus and the response is recorded, the POD(a) function can be estimated from the statistical scatter in the response magnitudes as a function of crack size. The data pair comprising size and signal response are designated as (a_i, $\hat{a}_i$) in which $\hat{a}_i$ is the response to the inspection stimulus for the i^{th} crack. If $\hat{a}_i$ is greater than a pre-set threshold, $\hat{a}_{th}$, a crack is indicated. Data of this nature are often referred to as $\hat{a}$ vs a (a-hat vs a). Data from automated eddy current systems are of this nature. Data from ultrasonic and liquid penetrant inspections have also been recorded and analyzed in the $\hat{a}$ vs a format. The parameters of the POD(a) function are estimated from the scatter in $\hat{a}$ values about the mean response to cracks of size a. Maximum likelihood is used to estimate the parameters and to place confidence bounds on the estimate of the reliably detected crack size when desired [MIL-HDBK-1823; Berens, 1988].

The demonstration of NDI capability is a consumer or quality concern. The primary objective of such demonstrations for a particular application is to estimate the POD(a) function and, consequently, the reliably detected crack size, say a_{NDI}. For damage tolerance considerations, a_{NDI} is commonly accepted to be the crack sizes designated as a_{90} or $a_{90/95}$. The a_{90} crack size is defined as the size for which POD(a_{90}) = 0.90 and $a_{90/95}$ is the upper (conservative) 95% confidence bound on the estimate of a_{90}. (The estimate of the a_{90} crack size is often referred to as the $a_{90/50}$ crack size under the wrong assumption that the estimate of a_{90} is the median of the sampling distribution of the estimates.)

NDI reliability experiments have also been conducted to optimize the inspection protocol and to ensure process control. System optimization with respect to POD(a) would have the objective of determining system configurations that produce acceptable a_{90} or $a_{90/95}$ values. The design of system optimization programs is of a different character and beyond the scope of demonstrating the capability of the system.

3.1.2.2 Design of NDI Capability Demonstrations

NDI capability is typically quantified through a capability demonstration program. The concept for such a demonstration is to mimic the real inspection as closely as possible on representative specimens that contain cracks of known sizes that span the range of increase of the POD(a)

function. A comprehensive description for the execution of such a demonstration program and the analysis of the resulting data is presented in MIL-HDBK-1823 (see also Berens [1988] and Berens [2000]). The analysis of the data from an NDI demonstration uses the maximum likelihood estimates of the parameters of the POD(a) model and the asymptotic properties of such estimates. This subsection briefly reviews the design and execution of a generic capability demonstration.

An NDI reliability demonstration comprises the execution of a test matrix of inspections on a set of specimens with known crack locations and sizes. The inspection results, either $\hat{a}$ or hit/miss, are then analyzed to estimate the parameters of the POD(a) function and the reliably detected crack size for the inspection application. The specimens are inspected under a test protocol that simulates as closely as practical the actual application conditions. Establishing test protocols for eddy current, fluorescent penetrant, ultrasonic and magnetic particle inspection systems are discussed in MIL-HDBK-1823.

The objectives and costs of an NDI demonstration determine the matrix of inspections to be performed. From the analysis viewpoint, there are two major categories of concerns that must be addressed in establishing the experimental design. These are: a) the generality of inferences that can be made from the controlled and uncontrolled inspection and material parameters; and, b) the number and sizes of cracks and the number of uncracked inspection sites in the specimens.

<u>Controlled and Uncontrolled Factors</u>

To demonstrate capability for an application, it is assumed that: a) the complete protocol for conducting the inspection is well defined for the application; b) the inspection process is under control; and, c) all other factors which introduce variability in an inspection decision are reasonably representative of the application. The representativeness of these other factors limits the scope of the POD(a) characterization and is addressed by controlling the factors during the inspection or by randomly sampling the factors to be used in the demonstration. The methods of accounting for these factors are important aspects of the statistical design of the demonstration and significantly influence the statistical properties of the estimates of the POD(a) function parameters.

The important classes of the factors that introduce variation in crack detectability are:

a) the inherent degree of repeatability of the magnitude of the NDI signal response when a specific crack is independently inspected many times with all controllable factors held constant;
b) the material and geometrical properties of the specimens and the differences in the physical properties of cracks of nominally identical "size";
c) the variation introduced by different hardware components in the inspection system; and,
d) the summation of all the human factors associated with the particular population of inspectors that might be used in the application.

The effects of these factors are present in every NDI reliability demonstration and they should be explicitly considered in the design of the demonstration and the interpretation of the results.

Little can be done about the variation of the response to the NDI excitation at the demonstration stage when inspections are repeated under fixed conditions. This variation might be reduced if the system was modified or better optimized but that is a different objective. Repeat inspections under identical conditions will provide a measure of the inherent variability that is a lower bound on the variability to be expected in applications of the system.

The character of the cracks in the structure being inspected will have a significant influence on the inspection outcome. There are two elements of crack character that impact the demonstration: the physical characteristics of the specimens containing the cracks and the physical properties of the cracks in the specimens. The inspection system will be designed to detect cracks of a defined size range at a location in a structural element defined at least by a material type and geometrical configuration combination. A fixed set of specimens containing cracks will be inspected and these specimens either must be of this combination or the assumption must be made that differences in inspection response in the specimens is identical to that obtained in the real application.

The cracks in the specimens must be as close as possible to the cracks that will be in the real structures and of sizes that span the region of interest for the POD(a) analysis. The assumption of equivalent response to the real inspection is implied when the results of the demonstration are implemented. Experience with the inspection will dictate the degree of acceptance of the assumption. For example, EDM notches are not good substitutes for eddy current inspections of surface fatigue cracks but may be the only possible choice for subsurface ultrasonic inspections.

Inspection capability is expressed in terms of crack size but not all cracks of the same "size" will produce the same magnitude of inspection response. In general, the specimens used in NDI reliability demonstrations are very expensive to obtain and characterize in terms of the sizes of the cracks in the specimens. Each set of specimens will be inspected multiple times if other factors are being considered in the demonstration. From a statistical viewpoint, this restriction on the experimental design limits the sample size to the number of cracks in the specimen set. Multiple independent inspections of the same crack only provide information about the detection probability of that crack and do not provide any information about the variability of inspection responses between different cracks. Stated another way, k inspections on n cracks is not equivalent to inspections of $n \bullet k$ different cracks, even if the inspections are totally independent. The number and sizes of cracks will be addressed later.

Accounting for the variability due to differences in inspection hardware must first be considered in terms of the scope of the capability evaluation. Each component of the inspection system can be expected to have some, albeit small, effect on inspection response. The combinations of particular components into sub-systems and complete inspection stations can also be expected to influence the response. Recognizing that individual hardware combinations might have different POD(a) capabilities, a general capability objective must be set. Each combination can be characterized, each facility comprising many combinations can be characterized, or many facilities can be characterized. Ideally, the available hardware combinations would be randomly sampled for the scope of the desired characterization and a weighted average of responses would be used to estimate the POD(a) function. On a practical level this is seldom done for ostensibly identical equipment. (Note that an analogous problem exists when accounting for the human factors which will be discussed in the following.) More commonly, capability demonstrations are performed on combination of hardware and the assumption is made that the characterization would apply to all combinations. That is, the POD(a) differences between combinations are assumed to be negligible.

The above is directed at a complete individual inspection system (however defined), but the variability of interchangeable components of a system can often be directly assessed. For example, experience has shown that different eddy current probes produce different responses when all other factors are constant. If a single probe is used to demonstrate the capability of an eddy current system, the estimated POD(a) function applies to the relevant inspections using that probe.

However, if the POD characterization is to be used for in-service inspections using any such probe, an assumption is required that the probe is representative of the entire population. If a larger demonstration is affordable, the inspections could be performed using a random sample of probes from the available population. The analysis method must then account for the fact that multiple inspections of each crack were made with the different probes. The resulting characterization would better represent an inspection for a randomly selected probe.

Accounting for the variation from more than one source is more complex. Care must taken to ensure that the multiple sources are balanced in the analysis of the data and that the correct analysis procedures are used. For example, in the early evaluations of an automated eddy current system for turbine engine disks (the ECIS system for the ENSIP/RFC applications), there was considerable interest in the inherent variability in response from repeated, identical inspections and from different probes with their associated re-calibration changes. (Other factors were initially considered but were later ignored after it was shown that they had no affect on POD(a) for the system.) The specimen sets would be inspected three times: twice with one probe and once with a second probe. The data from the three inspections, however, could not be combined in a single analysis since such an analysis would skew the results toward the probe with double representation. Thus, one analysis would be performed to estimate the inherent repeat variability and a second analysis would be performed to estimate the probe to probe variation. The results would then be combined to arrive at the POD(a) function that accounted for both sources of variation. It might be noted in this context that the repeat variability was negligible as compared to the variability that results from re-calibration and probe changes. The demonstration plan was later modified to better estimate the more significant between probe variation by performing the third inspection with a third probe.

Factorial-type demonstrations are an efficient approach to simultaneously account for several significant factors. However, such demonstrations for more than a couple of factors require many inspections of the specimen set. More sophisticated statistical experimental designs might be employed but the actual choice of such a design and the analysis of the data are driven by the specific objectives of a particular experiment. Discussion of such designs is beyond the scope of this discussion.

Human Factors

When the inspector plays a significant role in the find/no find decision, he or she is an integral component of the NDI system. In such common inspection scenarios, human factors can contribute significantly to the variability in inspection results. In this context, human factors refer to both the dynamic capabilities of individual inspectors and the user friendliness of the inspection tools in the environment of the application. Experiments have been conducted to quantify some of the environmental effects of human factors and data from some demonstration experiments have been interpreted in terms of the level of training and experience of the inspectors (see, for example, Spencer & Schurman [1994]). However, the effects and interactions of human factors on inspection results have not been characterized. Rather, to the extent possible, NDI systems are automated to minimize the effect attributed to the inspector.

In a non-automated inspection, many human factors potentially influence the inspection decision and they cannot all be accounted for in a capability demonstration. At some level, the representative inspection assumption will be required. Given that the mechanical aspects of the NDI system and inspection environment are held constant, differences between inspectors can cause a biased capability characterization if ignored. Again, the objective of the capability characterization must be stated in advance. If each inspector is being evaluated, a separate POD(*a*) function for each is estimated. If a single POD(*a*) function is wanted for an entire facility, the inspectors in the demonstration must be randomly sampled in proportion to the percent of such inspections each performs. Alternatively, inspectors might be categorized by, say, capability as implied by certification level. A random sample of the inspectors from each level could be selected to arrive at a composite POD(*a*) for the level and a weighted average would be calculated based on the percent of inspections performed by each level. An example of designing such a demonstration is given in Hovey, et al. [1989]. Example results from the evaluation of a population of inspectors can also be found in Davis [1988].

3.1.2.3 Sample Size Requirements

Sample sizes in NDI reliability experiments are driven more by the economics of specimen fabrication and crack characterization than by the desired degree of precision in the estimate of the POD(*a*) function. Reasonable appearing POD(*a*) functions can often be obtained from applying the maximum likelihood analysis to an inspection of relatively few specimens. Totally unacceptable results can also be obtained from inspecting specimens containing too few cracks or from inspection results that are not reasonably represented by the assumptions of the models. Therefore, it must be recognized that the confidence bound calculation for a POD(*a*) analysis is based on asymptotic (large sample) properties of the estimates and that there are minimal sample size requirements that must be met to provide a degree of reasonable assurance in the characterization of the capability of the system.

Larger sample sizes in NDI reliability experiments will, in general, provide greater precision in the estimate of the POD(*a*) function. However, the sample size is determined from the number of cracks in the experiment and there is an information content coupling with the crack sizes that must also be considered. The effect of this coupling manifests itself differently for the $\hat{a}$ versus *a* and hit/miss analyses.

Sample sizes for the binomial analysis that is used to demonstrate a capability at a single crack size are dictated strictly by the selected value of the target POD and the degree of confidence.

Sample Size Requirements for $\hat{a}$ versus *a* Analysis

When the crack decision is made on the basis of a recorded response, $\hat{a}$, to the inspection stimulus, the data are known as $\hat{a}$ versus *a* inspection results and a better POD(*a*) analysis is available. An example of $\hat{a}$ versus *a* data from a capability demonstration is presented in Figure 3.1.4. When the inspection response is greater than a pre-set detection threshold, a crack is indicated for the site. In a capability demonstration, the minimum signal threshold is set as low as possible with respect to noise. Detection thresholds are later set that will yield a desired a_{90} value with an acceptable rate of extra indications. Extra indications are crack indications at sites with no known cracks. Extra indications can be the result of noise or large responses from insignificant cracks. However, they can also result from anomalies that do not impair structural integrity.

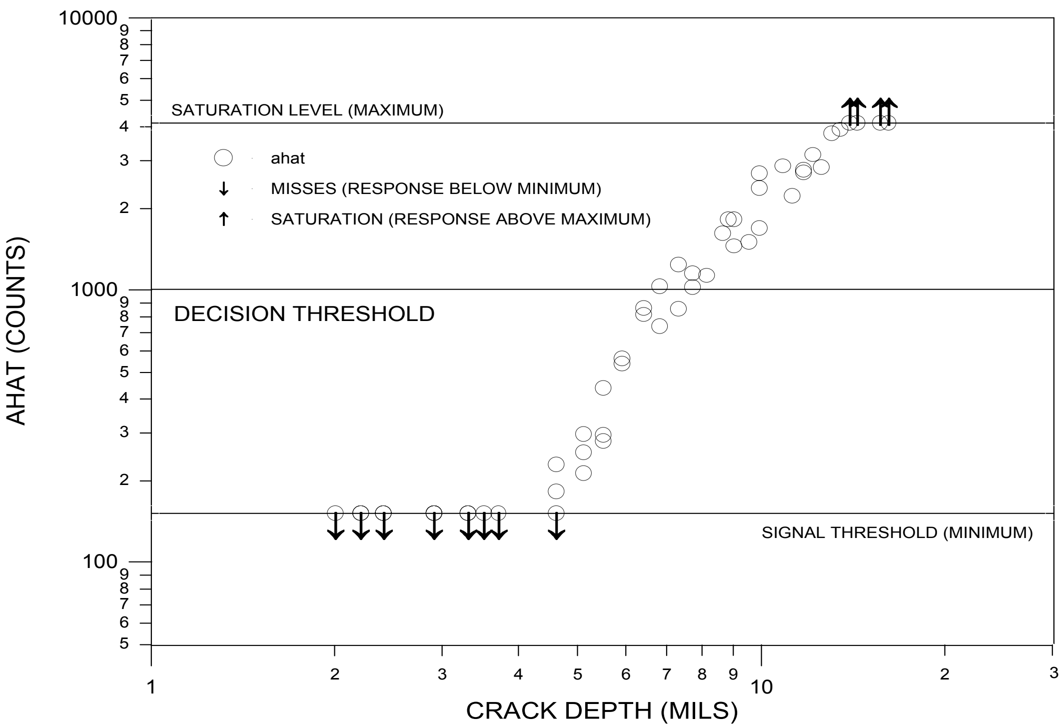


Figure 3.1.4. Example Plot of $\hat{a}$ versus a Data

The recorded signal response, $\hat{a}$, provides significantly more information for analysis than a simple crack or no crack decision of a hit/miss inspection response. The POD(a) model is derived from the correlation of the $\hat{a}$ versus a data and the assumptions concerning the POD(a) model can be tested using the signal response data. Further, the pattern of $\hat{a}$ responses can indicate an acceptable range of extrapolation. Therefore, the range of crack sizes in the experiment is not as critical in an $\hat{a}$ versus a analysis as in a hit/miss analysis. For example, if the decision threshold in Figure 3.1.4 was set at 1000 counts, only the cracks with depths between about 6 and 10 mils would provide information that contributes to the estimate of the POD(a) function. The larger and smaller cracks are always found or missed and would have provided little information about the POD(a) function in a hit/miss analysis. In the $\hat{a}$ analysis, however, all of the recorded $\hat{a}$ values provided full information concerning the relation between signal response and crack size and the censored values at the signal minimum and maximum limits provided partial information. The parameters of the POD(a) function are derived from the distribution of $\hat{a}$ values about the median response for cracks of size a. Assumptions necessary for characterizing this distribution are readily evaluated with the $\hat{a}$ versus a data.

Because of the added information in the $\hat{a}$ data, a valid characterization of the POD(a) function with confidence bounds can be obtained with fewer cracks than are required for the hit/miss analysis. It is recommended that at least 30 cracks be available for demonstrations whose results can be recorded in $\hat{a}$ versus a form. Increasing the number of cracks increases the precision of estimates. Perhaps, more importantly, increasing the number of cracks provides a broader population of the different types of cracks that the inspection will address. Therefore, the demonstration specimen test set should contain as many cracked sites as economically feasible. The analysis will provide parameter estimates for smaller sample sizes but the adequacy of the asymptotic distributions of the estimates is not known.

Sample Size Requirements for Pass/Fail Analysis

In a hit/miss capability demonstration, the inspection results are expressed only in terms of whether or not the crack of known size was detected. There are detection probabilities associated with each inspection outcome and the analysis assumes that the detection probability increases with crack size. Since it is assumed that the inspection process is in a state of control, there is a range of crack sizes over which the POD(a) function is rising. In this crack size range of inspection uncertainty, the inspection system has limited discriminating power in the sense that detecting or failing to detect would not be unusual. Such a range might be defined by the interval ($a_{0.10}$, $a_{0.90}$), where a_p denotes the crack size that has probability of detection equal to p; that is, POD(a_p) = p. Cracks smaller than $a_{0.10}$ would then be expected to be missed and cracks greater than $a_{0.90}$ would be expected to be detected.

In a hit/miss capability demonstration, cracks outside the range of uncertainty do not provide as much information concerning the POD(a) function as cracks within this range. Cracks in the almost certain detection range and almost certain miss range provide very little information concerning probability of detection. In the hit/miss demonstration, not all cracks convey the same amount of information and the "effective" sample size is not necessarily the total number of cracks in the experiment. For example, adding a large number of very large cracks does not increase the precision in the estimate of the parameters of the POD(a) function.

Ideally, all of the cracks in a hit/miss demonstration would have 80 percent of their sizes in the ($a_{0.10}$, $a_{0.90}$) range of the POD(a) function. However, it is not generally possible to have a set of specimens with such optimal sizes for all demonstrations. The demonstrations are being conducted to determine this unknown range of sizes for the NDI system being evaluated. Further, because of the high cost of producing specimens, the same sets of specimens are often used in many different demonstrations. To minimize the chances of completely missing the crack size range of maximum information and to accommodate the multiple uses of specimens, the sizes of cracks in a specimen set should be uniformly distributed between the minimum and maximum of the sizes of potential interest. A minimum of 60 cracks should be distributed in this range, MIL-HDBK-1823, but as many as are affordable should be used. This minimum sample size recommendation was the result of subjective considerations as to the number needed to make the asymptotic assumptions reasonable, experience in applying the model to data, and the results of analysis from a number of simulated POD demonstrations [Berens & Hovey, 1981; Berens & Hovey, 1984; and Berens & Hovey, 1985].

Sample Size Requirements for Binomial Analysis

When capability is to be demonstrated by using specimens with cracks of the same size and the binomial analysis, the number of cracks in the specimens can be determined exactly from the POD level and the desired degree of confidence. The best (maximum likelihood) estimate of the POD at the crack length of interest is the proportion of cracks in the specimen set that are detected. A lower bound on the estimate is then calculated for the desired confidence level using binomial distribution theory. For example, to demonstrate that there is 95 percent confidence that at least 90 percent of all cracks of the size under consideration will be detected requires at least 29 cracks of that size. If all 29 cracks are detected, the maximum likelihood estimate of POD is 1.0 and the lower 95 percent confidence bound is slightly greater than 0.9. If any crack is missed, the lower confidence bound on the estimate of POD is less than 0.9. Sample sizes for the binomial analysis will be discussed further in the subsection on analysis methods.

It must be emphasized that the sample size is determined by the number of different cracks, not the number of inspections. Different cracks can respond differently to inspection stimuli. Multiple inspections of the same crack are not independent and, therefore, cannot be treated as independent samples from the population of cracks of the given size. There is a tendency to re-inspect specimens to increase the sample size. For example, if one of 29 cracks is not detected, the inspection does not qualify for an $a_{90/95}$ capability at that size. The specimen set cannot be re-inspected with the expectation of passing the test for a sample size of 58. New specimens with different cracks must be used or the analysis is not valid.

Uncracked Inspection Sites

In the context of the preceding discussion, sample size refers to the number of known cracks in the specimens to be inspected during the capability demonstration. The complete specimen set should also contain inspection sites that do not contain any known cracks. If the inspection results are of the hit/miss nature, at least twice as many uncracked sites as sites are recommended. The uncracked sites are necessary to ensure that the NDI procedure is truly discriminating between cracked and uncracked sites and to provide an estimate of the false call rate. If the NDI system is based on a totally automated $\hat{a}$ versus a decision process, many fewer uncracked sites will be required. If any $\hat{a}$ values are recorded at the uncracked sites, their magnitude would provide an indication of the minimum thresholds that might be implemented in the application.

3.1.2.4 POD Analysis

As noted there are two approaches to quantifying NDI capability – fitting a model that expresses probability of detection as a function of crack size and demonstrating a POD capability for a particular crack size. Data from the single crack size demonstration approach are analyzed using a straightforward binomial distribution analysis. Fitting a POD(a) model to the results of an NDI demonstration depends on the nature of the data (hit/miss or $\hat{a}$ versus a), the function chosen to represent POD(a), and the method for fitting the parameters of the function and determining the confidence bound on the reliably detected crack size. Experience with $\hat{a}$ versus a data from eddy current inspections has shown that a cumulative normal equation provides a reasonable model for the POD(a) function when transformations of crack size or inspection signal response are considered. Further, Berens and Hovey [1981], showed that the lognormal cumulative distribution provided as good or better a model than the eight others that were considered. Accordingly, the Air Force has generally adopted the cumulative normal distribution function as the model for POD(a) analyses. Note that the cumulative lognormal model is the cumulative normal model after crack size is transformed. The log odds equation is also often used to fit NDI data. The log odds equation and the cumulative lognormal equation are essentially indistinguishable.

A computer program, POD Version 3, is recommended by MIL-HDBK-1823 for the analysis of both $\hat{a}$ versus a and hit/miss POD(a) analyses (see also Berens [2000]). The program calculates the maximum likelihood estimates of the cumulative normal model as well as confidence bounds on estimates of a_p. The program permits transformations of the data. Since the default analysis is based on the natural logarithm transformation, the default analysis is for the cumulative lognormal POD(a) function. In POD Version 3, data are input through an Excel spreadsheet and output is provided as separate tables and graphs in the spreadsheet.

The following paragraphs present a general description of the analysis methods.

$\hat{a}$ Versus a Analysis

All NDE systems make find/no find decisions by interpreting the response to an inspection excitation. In some inspections, the response is a recordable metric, $\hat{a}$, that is related to the flaw size. Find/no find decisions are made by comparing the magnitude of $\hat{a}$ to the decision threshold value, $\hat{a}_{dec}$. The $\hat{a}$ versus flaw size analysis is a method of estimating the POD(a) function based on the correlation between $\hat{a}$ and flaws of known size, a. The general formulation of the $\hat{a}$ versus a model is expressed as

$$\hat{a} = f(a) + \delta \tag{3.1}$$

where $f(a)$ represents the average (or median) response to a crack of size a and δ represents the sum of all the random effects that makes the inspection of a particular crack of size a different from the average of all cracks of size a. In principle, any f(a) and distribution of δ that fit the observations can be used. However, if f(a) is linear in a,

$$\hat{a} = B_0 + B_1 a + \delta \tag{3.2}$$

and δ is normally distributed with constant standard deviation, σ_δ, then the resulting POD(a) function is a cumulative normal distribution function. Monotonic transformations of $\hat{a}$ or a can be analyzed in this framework. In fact, the model has been shown to fit a large number of cases in which a logarithmic transformation of both a and $\hat{a}$ was applied.

As an example consider the formulation of the $\hat{a}$ versus a analysis that has been used exclusively in the evaluation of the RFC/ENSIP automated eddy current inspection system. The relation between $\hat{a}$ and a is expressed in terms of the natural logarithms of $\hat{a}$ and a.

$$ln\,\hat{a} = B_0 + B_1 \bullet ln\,a + \delta \tag{3.3}$$

where δ is Normal (0, σ_δ). For a decision threshold of $\hat{a}_{dec}$,

$$POD(a) = \Phi\left[\frac{ln\,a - \mu}{\sigma}\right] \tag{3.4}$$

where $\Phi(z)$ is the cumulative standard normal distribution function and

$$\mu = \frac{ln\,\hat{a}_{dec} - B_0}{B_1} \tag{3.5}$$

$$\sigma = \sigma_\delta / B_1 \tag{3.6}$$

The calculation is illustrated in Figure 3.1.5. The parameters of the $\hat{a}$ versus a model (B_0, B_1, and σ_δ) are estimated from the data of the demonstration specimens. The probability density function of the ln $\hat{a}$ values for a 13 mil crack depth is illustrated in the figure. The decision threshold in the example is set at $\hat{a}_{dec}$ = 165. The POD for a randomly selected 13 mil crack would be the proportion of all 13 mil cracks that would have an $\hat{a}$ value greater than 165, i.e. the area under the curve above 165. In this example, the decision threshold was selected so that POD(13) = 0.90. The estimate of the POD(a) function and its 95 percent confidence bound for the decision threshold of 165 counts is presented in Figure 3.1.6. It might be noted that when all cracks have a recorded response between the signal minimum and maximum, the maximum likelihood estimates are identical with those obtained from a standard regression (least squares) analysis. However, when crack response is below the signal minimum or above the maximum (saturation level of the recorder), more sophisticated calculations are required to obtain parameter estimates and the confidence bound. For complete details of the maximum likelihood calculations and more discussion of the $\hat{a}$ versus a analysis, see MIL-HDBK-1823, Berens [1988], and Berens [2000].

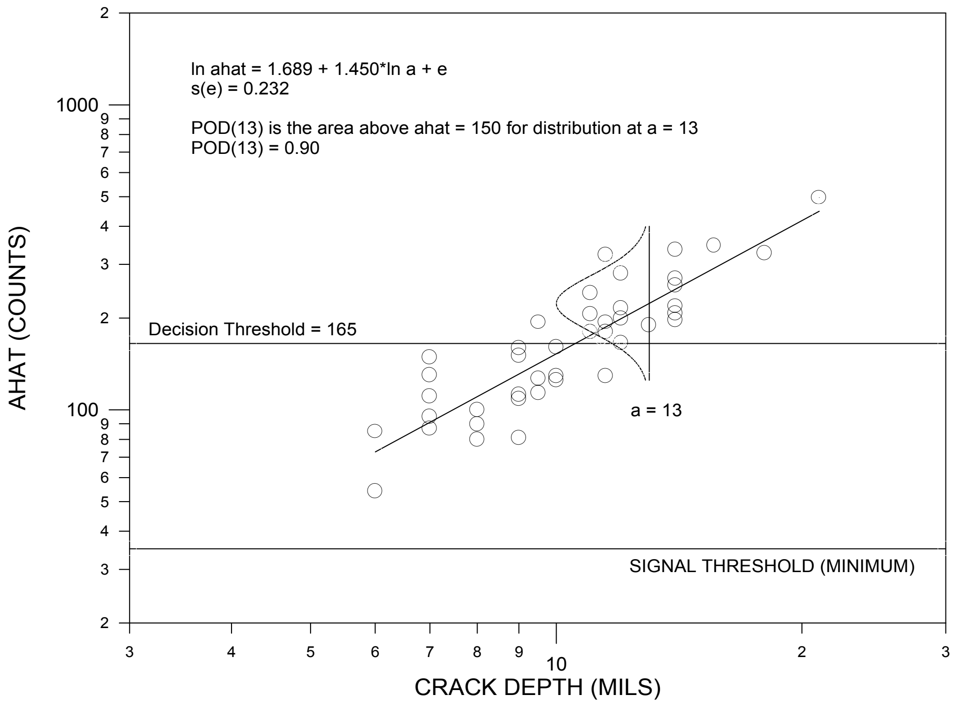


Figure 3.1.5. Example POD(a) Calculation from $\hat{a}$ versus a Data

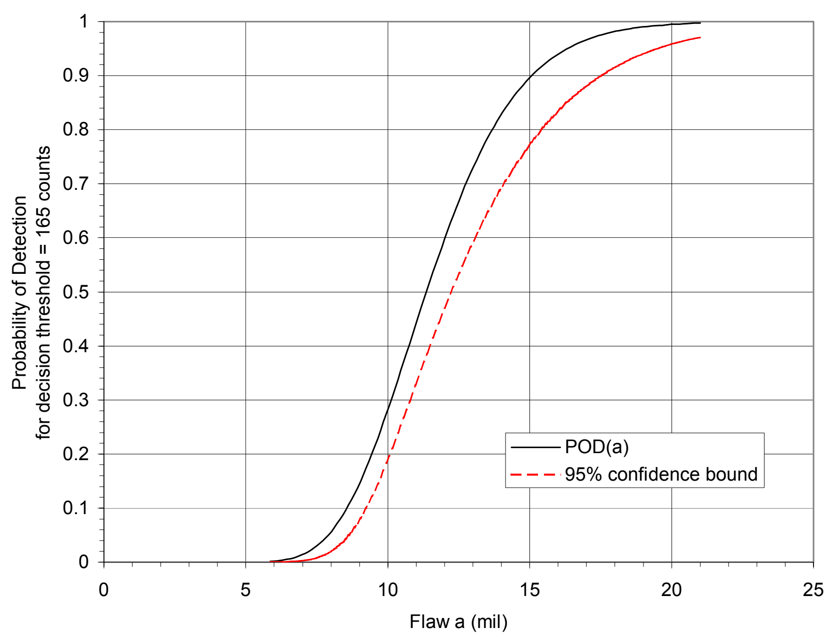


Figure 3.1.6. POD(a) Function with 95 Percent Confidence Bound for an Example $\hat{a}$ versus a Analysis

The preceding formulation of the $\hat{a}$ versus a model is based on three assumptions:

a) the mean of the log responses, ln $\hat{a}$, is linearly related to log crack size, ln a;

b) the differences of individual ln $\hat{a}$ values from the mean response have a normal distribution; and,

c) the standard deviation of the residuals, σ_δ, is constant for all a.

These assumptions can be tested using the results of the data from the demonstration. When the assumptions are not acceptable, current practice is to restrict the analysis to a range of crack sizes for which the assumptions are acceptable.

These assumptions can be easily checked and statistical tests for all three assumptions are built into the standard analysis of the POD Version 3 computer program of MIL-HDBK-1823.

If the ln $\hat{a}$ versus ln a relation is not linear, it may be possible to use other transformations of either the signal response or the crack size. If the three assumptions are reasonably valid for other transformations of the data, the above analysis can be applied using the different transformation. The inverse transformation of the results provides the answers in the correct units. Data sets have been observed in which no transformation was required and the fit was made directly to $\hat{a}$ versus a data (i.e. without the logarithmic transform). Other data sets have been analyzed in which the three assumptions were acceptable when the analysis was performed in terms of ln $\hat{a}$ versus $1/a$. It should be noted that extreme caution must be exercised when extrapolating the results beyond the range of crack sizes in the data. The POD Version 3 computer program has been designed to perform the POD analyses using transformations other than the logarithmic. The logarithmic transform of both crack size and inspection response is the default transform.

Hit/Miss Analysis

The results of an inspection system are often recorded only as a decision as to the presence (hit, find, or pass) or absence (miss, no find or fail) of a crack. The available data from the capability demonstration of such inspections comprise data pairs of crack size and the inspection result. The parameters of a POD(a) model for such data can be estimated using maximum likelihood as follows:

Let a_i represent the size of the i^{th} crack and Z_i represent the result of the inspection: $Z_i = 1$ if the flaw was found (hit) and $Z_i = 0$ if the flaw was not found (miss). Assume POD(a_i) is the equation relating probability of detection to flaw size for the inspection. The likelihood of obtaining a specific set of (a_i, Z_i) results when inspecting the specimens is

$$L(\theta) = \prod \left[POD(I)\right]_i^Z \left[1 - POD(a_i)\right]_i^{1-Z} \tag{3.7}$$

where $\theta = (\theta_1, \theta_2, \ldots, \theta_k)$ is a vector of the parameters of the POD(a) function. Values of $\theta_1, \theta_2, \ldots, \theta_k$ are determined to maximize L(θ). For typical POD(a) models, it is more convenient to perform the analyses in terms of logarithms.

$$\ln L(\theta) = \sum Z_i \ln POD(a_i) + \sum (1 - Z_I) \ln[1 - POD(a_i)] \tag{3.8}$$

The maximum likelihood estimates are given by the solution of the k simultaneous equations:

$$\frac{\delta \ln(\theta)}{\delta \theta_i} = 0, \quad i = 1, \ldots, k \tag{3.9}$$

In general, an iterative solution will be required to solve Equations 3.9.

Any monotone increasing function between zero and one can be used for POD(a). However, an early study of data with multiple inspections per crack [Berens & Hovey, 1981] indicated that the log odds or, equivalently, the cumulative lognormal models were more generally applicable than the others investigated. Further, the assumptions leading to a cumulative log normal model for the POD(a) function for $\hat{a}$ versus a data have often been verified for eddy current data. The log odds and cumulative lognormal models are equivalent in a practical sense in that the maximum difference in POD(a) between the two for fixed location and scale parameters is about 0.02 which is well within the scatter from repeated determinations of a POD(a) capability.

POD Version 3, the computer program recommended by MIL-HDBK-1823, is based on a cumulative normal equation but allows transformations of the crack size. The default transform of POD Version 3 is the natural logarithm transform so that the program will fit the cumulative lognormal equation by default. However, the program also provides a solution based on the log odds equation. Other models for the POD(a) function may be appropriate but, if preferred, would require a different computer implementation.

Repeating Equation 3.4, the cumulative log normal equation for the POD(a) functions is:

$$POD(a) = \Phi\left(\frac{\ln a - \mu}{\sigma}\right) \tag{3.10}$$

where $\Phi(z)$ is the standard normal cumulative distribution function. The log odds model for the POD(a) function is:

$$POD(a) = \left\{1 + exp - \left[\left(\frac{\pi}{\sqrt{3}}\right)\frac{(\ln a - \mu)}{\sigma}\right]\right\}^{-1} \tag{3.11}$$

Equation 3.10 or 3.11 is substituted in Equations 3.7 through 3.9 for POD(a). $\hat{\mu}$ and $\hat{\sigma}$ are determined so as to maximize $L(\mu, \sigma)$, the likelihood of obtaining the observed inspection results. Note that POD(μ) = 0.5 for both models. σ is a scale parameter that determines the degree of steepness of the POD(a) function. A negative value of σ is not contradictory but, for a negative σ, the POD(a) function will decrease with increasing a.

There are occasions when Equations 3.9 do not converge. No solution will be obtained if the sizes of found cracks do not overlap with the sizes of missed cracks. Little information is obtained from cracks that are so large they are always found or so small they are always missed. More overlap is needed for the cumulative lognormal model than for the log odds model. It is also possible to obtain negative estimates of σ from erratic data sets. Results of this nature are due to the wrong range of crack sizes in the demonstration or to an inspection process that is not under proper control. When the crack sizes in the specimens are not in the range of increase of the POD(a) function, the effective sample size is smaller and the effect is reflected in larger standard deviations of the sampling distributions of the parameter estimates and, thus, wider confidence bounds.

Damage tolerance analyses are driven by the single crack size characterization of inspection capability for which there is a high probability of detection. Typically, the one number characterization of the capability of the NDE system is expressed in terms of the crack length for which there is 90 percent probability of detection, a_{90}. But a_{90} can only be estimated from a demonstration experiment and there is there is sampling uncertainty in the estimate. To cover this variability, an upper confidence bound can be placed on the best estimate of a_{90}. The use of an upper 95 percent confidence bound, the $a_{90/95}$ crack size has become the de facto standard for this characterization of NDE capability. The use of $a_{90/95}$ is intended to be conservative from the viewpoint of damage tolerance analyses.

In the hit/miss analysis of POD Version 3 a single value of POD(a), say 0.90, is selected and an upper confidence bound, say 95 percent is calculated for the POD value. This procedure is known as a point by point confidence bound. These are valid confidence bounds for any one POD value but not for the entire POD(a) curve.

The confidence bounds for the estimates of a_{90} are calculated using the asymptotic normality properties of the maximum likelihood estimates [Berens, 2000]. Figure 3.1.7 presents an example of a fit to hit/miss data from a semi-automated, directed eddy current inspection.

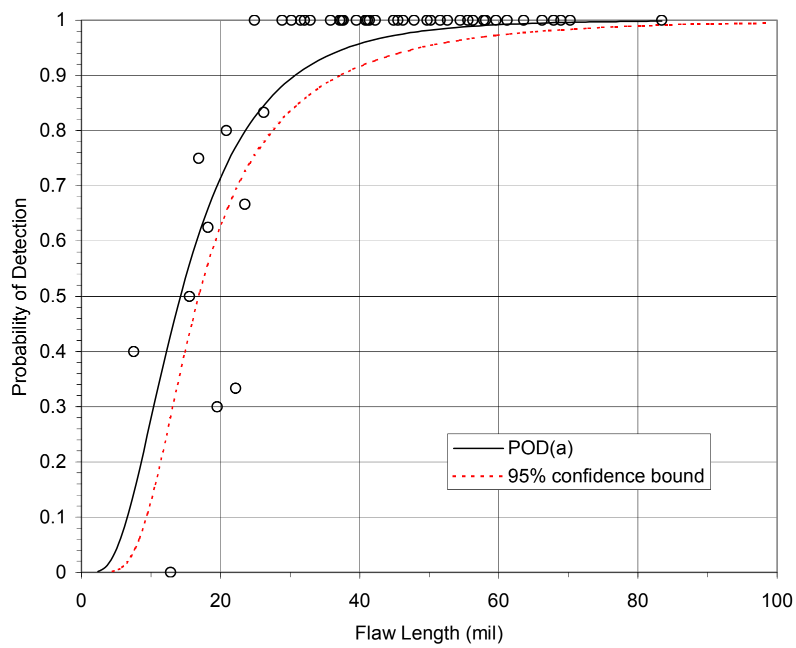


Figure 3.1.7. Example POD(a) for a Semi-Automated, Directed Eddy Current Inspection

Binomial Analysis for Cracks of Fixed Size

Because of the individual physical differences between cracks, cracks of the same size will have different detection probabilities for a given NDI system. However, a single POD for all cracks of that size can be postulated in terms of the probability of detecting a randomly selected crack from the population of all cracks of the given size. In this formalism, the proportion detected in a random sample of the cracks is an estimate of POD for that size and binomial distribution theory can be used to calculate a lower confidence bound on the estimate. Given a sample of inspection results from cracks of a target size, say a_{NDI}, the inspection system is considered adequate if the lower confidence bound on the proportion of detected cracks exceeds the desired POD value.

The theory of the binomial analysis is as follows. Given independent inspection results from specimens containing n cracks of size a_{NDI}, the target reliably detected crack size. Assume that r

of the cracks are detected. If POD is the true (but unknown) probability of detection for the population of cracks, the number of detections is modeled by the binomial distribution. The probability of r detections in n independent inspections of cracks of size a_{NDI} is:

$$P(r) = \frac{n!}{(n-r)!} POD^r (1-POD)^{n-r} \tag{3.12}$$

The unbiased, maximum likelihood estimate of POD is

$$\overline{POD} = r / n \tag{3.13}$$

The 100(1-γ) percent lower confidence bound, POD_{CL}, on the estimate of POD is obtained as the solution to the equation:

$$\gamma = \sum_{i=r}^{n} \frac{n!}{(n-1)!\,i!} POD_{CL}^{\ i} (l - POD_{CL})^{n-i} \tag{3.14}$$

The interpretation of POD_{CL} as a lower confidence bound is as follows. If the demonstration was completely and independently repeated a large number of times, 100(1-γ) percent of the calculated lower bounds would be less than the true value of POD. There is 100(1-γ) percent confidence that POD_{CL} from a single demonstration will be less than the true value.

Solutions to Equation 3.14 are tabulated in Natrella [1963] for 90, 95, and 99 percent confidence limits and selected sample sizes. General solutions expressed in terms of the incomplete beta function and the normal approximation to the binomial distribution can be found in many statistical references, for example, Mood [1950]. Minimum values of n and r which yield predefined values of POD_{CL} and confidence level, 100(1-γ), are often quoted. Selected values can be found in Packman, et al. [1976].

For example, consider a demonstration that there is 95 percent confidence that at least 90 percent of all cracks of size a_{NDI} will be detected by a given inspection system. To achieve the desired level of confidence and POD would require results as given in Table 3.1.2.

Table 3.1.2. Minimum Number of Detections Require to Conclude that POD > 0.90 with 95 Percent Confidence

Number of Cracks of Size a_{NDI}	Number of Cracks Detected
29	29
46	45
61	59
75	72
89	85
103	98

If there were 28 cracks in the demonstration and all 28 were detected, the lower 95 percent confidence bound on the estimate of POD would be 0.899. If less than 28 were detected, the lower confidence bound would be even lower. Since the minimum number of specimens that can yield a 90 percent POD at 95 percent confidence is 29, this approach to capability demonstration has been referred to as the "29 out of 29" method.

There are several objections to the use of this approach to quantifying inspection capability:

1) This demonstration approach to capability provides only minimal and reasonably gross POD information for the single crack size used for the inspections. Steep POD(*a*) functions are generally considered superior to flat POD(*a*) functions and a single crack size capability demonstration provides no information regarding POD(*a*) steepness.
2) Passing or failing the demonstration provides no discrimination of degree of detectability at the high POD levels. For example, consider the 29 finds out of 29 cracks criterion for demonstrating the 90/95 capability. If the true POD is less than 0.9, there is up to a 5 percent chance that the demonstration will conclude that the true POD is 0.9 or greater. Conversely, if the true POD is 0.995, there is a 15 percent chance that at least one crack out of 29 will be missed and the demonstration will fail to conclude that there is 95 percent confidence that the POD is greater than 0.9. At POD = 0.976, there is about a fifty-fifty chance of concluding the POD is greater than 0.9. POD(*a*) tends to be relatively flat above 0.9 and there could easily be a very large crack size difference between, say, a 0.9 capability and 0.995 capability. Even when crack detection is absolutely certain for the given size, only a 90/95 capability can be claimed after the demonstration.
3) When attempting to demonstrate a 90/95 capability and one crack out of 29 is missed, the demonstration must be repeated with at least additional 17 cracks. Since demonstrations are planned with the expectation of meeting the criteria, the need for additional specimens can create significant problems.

For these reasons, quantifying inspection capability in terms of the entire POD(*a*) function has evolved as the preferred method [MIL-HDBK-1823].

It might be noted that attempts have been made to use a binomial approach to the analysis of demonstration data comprising a range of crack sizes [Yee, et al., 1976]. These approaches have been generally abandoned but a Bayesian approach to such analyses is being considered [Bruce, 1998].

3.1.3 NDI Capability Evaluation for Corrosion

The impact of corrosion on the sustainment costs of an aging fleet is significant, particularly for transport aircraft. The presence of corrosion indicates a failure of the corrosion protection system and necessitates some sort of action in the maintenance plan. Regardless of the corrosion control maintenance strategy, NDI plays an important role in its implementation and the need exists to quantify the corrosion detection capability of the inspection system.

Several types of corrosion are typically found in aging airframes – uniform, pitting, intergranular, exfoliation, crevice (uniform and pitting), and stress corrosion cracking. Adaptations of the standard NDI methods discussed in Subsection 3.1.1 can be used to detect the various types of corrosion damage and new inspection methods are evolving. Although there is a need to quantify the corrosion detection capability of an NDI system, at present there is no commonly accepted procedure for doing so. Approaches to characterizing corrosion detection capability can be found in Alcott [1994], Howard and Mitchell [1995], Roach [1997], Komorowski, et al. [1998], and Hoppe, et al. [2000]. This subsection discusses two of the major problems in quantifying NDI capability for detecting corrosion and describes the method of Hoppe, et al. [2000].

3.1.3.1 Corrosion Metrics

When characterizing the NDI capability for detecting cracks, the natural metric for measuring crack damage was the linear crack dimension used in damage tolerance analyses. The selection of the appropriate metric for corrosion damage, however, is not immediately apparent. There are different types of corrosion damage and different metrics can be used to quantify the damages. For example, in hidden corrosion in lap joints and doublers on fuselage structures there are several possible metrics: thickness loss, pit depth and/or frequency, surface roughness, and joint pillowing. When inspecting for intergranular and exfoliation corrosion, around fasteners, useful metrics might be the maximum radial distance that the corrosion extends from the fastener hole or the corrosion area about each fastener. In some sense, each metric plays a role in the effect that the corrosion defect has on the structure. Consequently, it is important to consider all of the metrics for a given application. Each corrosion type must be considered separately, but the important aspect of the metric is that it measures corrosion severity. Ideally, the metric should be based on an "effect of defects" study; however, in practice the important metrics are generally known, and, in order to keep the assessment focused, it becomes necessary to select only one metric at a time for detection assessment. If it is absolutely essential to include an evaluation of more than one metric, then multiple evaluations must be performed (one evaluation per metric).

There is a necessary relation between the corrosion metric and the NDI technique. Obviously, the NDI technique must be responsive to changes in the corrosion damage metric. For example, in inspecting for hidden corrosion in lap joints, eddy current is responsive to thickness loss but may not be sufficiently responsive to pit depth. If pit depth is a critical parameter, a different NDI technique would be needed.

3.1.3.2 Corrosion Specimen Selection and Design

In the case of a crack detection assessment, representative cracks can be grown quite successfully in the laboratory. Since methods of corrosion growth are not well established, most notably for hidden corrosion, at present it is necessary to include real aircraft pieces with real corrosion in the specimen sets to be used in NDI capability demonstrations. Finding specimens with appropriate levels of corrosion is not a trivial problem. Potential specimens can be obtained from obsolete aircraft and from depots. While such specimens may contain real corrosion, they are not necessarily representative for a particular application. Further, a "good" NDI system for detecting hidden corrosion would be needed to select the specimens with varying degrees of corrosion damage. On the other hand, this situation does not eliminate the need for engineered and manufactured specimens. These specimens provide a level of control not available with the aircraft specimens. The type, location, and size of the defect (as measured by the chosen metric) can be controlled. The particulars of the engineered specimens must be determined from the specific metric chosen and the application. For thickness loss between layers, engineered specimens might include machined out areas of various depths and lateral dimensions. Experiment objectives also impact specimen designs. For example, a spatial resolution test would require a specially designed and manufactured specimen.

3.1.3.3 Example of Evaluating the Capability of an Eddy Current Inspection to Detect Hidden Corrosion in Lap Joints

The following example presents the results of an evaluation of an eddy current inspection for corrosion damage in C/KC-135 lap joints taken from Hoppe, et al. [2000]. For the example, the corrosion damage metric was taken to be thickness loss as thickness loss is an important criteria in judging severity of corrosion damage and eddy current is sensitive to thickness loss in the top layer of the lap joint.

Both real and engineered specimens were used for the capability demonstration. Several pieces from C/KC-135 and Boeing 707 fuselages were acquired. The specimens represented areas of interest on the aircraft and were expected to contain representative amounts of crevice corrosion. The specimens included fuselage lap joint and doubler sections that were anticipated to contain corrosion, as determined by disassembly of adjacent pieces of the skin. The specimens also included areas of little or no corrosion. The specimens that were selected incorporated the type, material, size and spacing of fasteners, thickness and lay-up of skins, presence of substructure, and specimen curvature variability that were expected to be experienced in typical aircraft inspections.

An engineered specimen was designed and manufactured for measuring the spatial resolution of the eddy current system. Spatial resolution of the system was necessary in to order to ascertain inspection regions of complete independence of the eddy current response. This specimen was constructed with several sets of lines of different widths machined in to the back surface of the front layer of an assembly of aluminum layers.

Specimens of a skin configuration were inspected using the eddy current system. NDI responses were recorded at independent sites within each specimen producing an inspection output profile of the specimen. Because thickness loss due to corrosion is variable within a specimen, the responses at the independent sites represent different samples of response at different thickness losses. The process is illustrated in Figure 3.1.8. The eddy current output at a point, *P*, in a response image is a function of the corrosion in a small region (or cell), *C*, on the specimen. The set of non-overlapping cells represents the collection of independent inspection opportunities from which probability of detection as a function of thickness loss can be calculated.

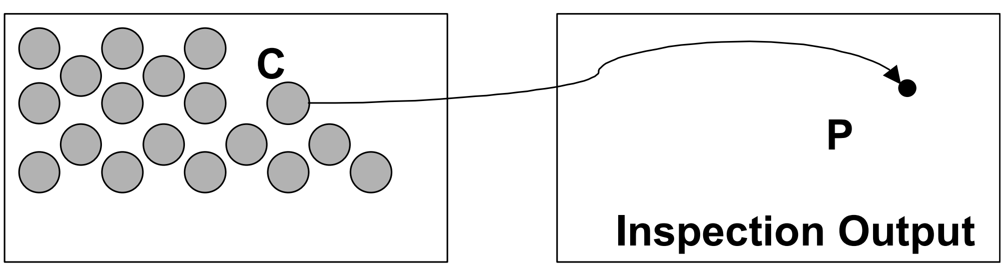


Figure 3.1.8. Schematic Diagram of Specimen and Inspection Output Images

After completion of the inspection of a specimen, the actual corrosion profile of the specimen was determined. The specimens were carefully disassembled by drilling out the fasteners and prying apart the layers. Corrosion products were chemically removed using a high concentration nitric acid exposure protocol. Measurement of remaining thickness was accomplished using calibrated topographic radiography. The inspection system output images and actual thickness loss profiles were registered to specimen features, such as fasteners and lap joint edges, in order to relate measured to actual thickness loss across each specimen.

Data pairs of real and EC measured thickness loss were generated for the independent inspection cells. The data pairs are plotted analogously to the $\hat{a}$ versus a plot of crack detection POD estimation. Figure 3.1.9 is an example of thickness loss versus EC response for one of the structural configurations. The scatter of the EC responses about the mean trend determines the POD as a function of thickness loss. Figure 3.1.10 shows the POD function for a threshold chosen to yield 90 percent detection for a 10 percent thickness loss. Also shown in Figure 3.1.10 is the 95 percent confidence bound on the POD function.

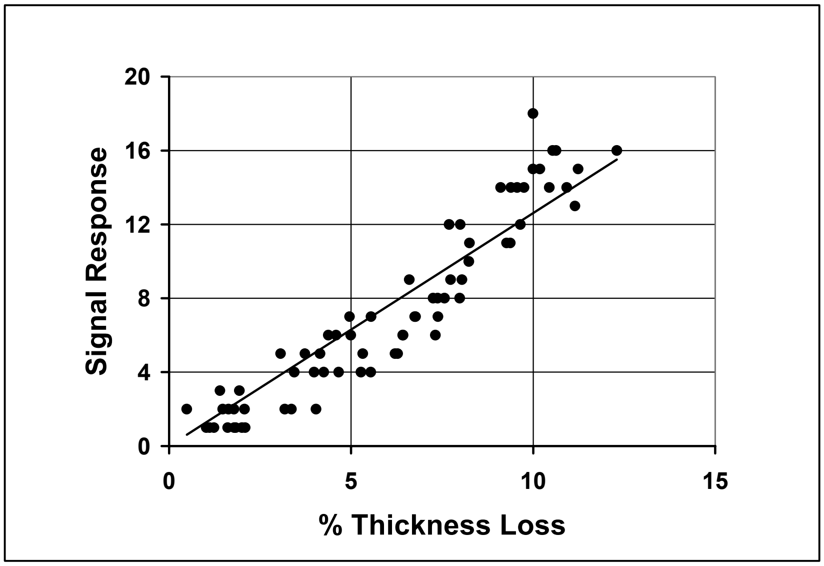


Figure 3.1.9. Example Eddy Current Response for Cells of Different Thickness Loss

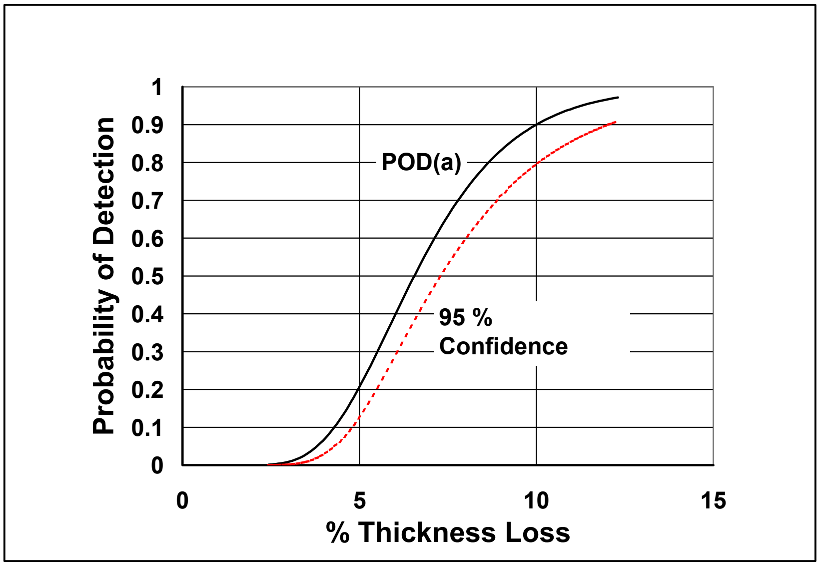


Figure 3.1.10. POD versus Percent Thickness Loss with Response Detection Threshold Set to Yield POD of 90 percent at 10 percent Loss