ASM Handbook, Volume 17, *Nondestructive Evaluation of Materials*
Aquil Ahmad and Leonard J. Bond, editors
DOI: 10.31399/asm.hb.v17.a0006438

Guide to Nondestructive Evaluation Techniques

Samuel W. Glass III, Pacific Northwest National Laboratory

NONDESTRUCTIVE EVALUATION (NDE), alternately known as nondestructive examination (NDE), commonly used within the nuclear industry; nondestructive testing (NDT), commonly used in the aerospace and petrochemical industry; and nondestructive inspection (NDI), commonly used within the military; is as the name implies test and assessment technologies designed to detect and characterize components without damaging or compromising the ability of the component to perform its design function. The most common application is detection of flaws caused either by manufacturing anomalies, service or environmental stresses, or natural material aging. More generally, applications may include estimation of mechanical and material properties, stress/strain, and dynamic response behavior. Postmanufacturing inspections frequently enjoy unfettered access to the component or part that may allow more thorough examinations, while in-service examinations after components have been in use may impose severe access limitations for in situ testing. If flaws are detected, they must then be evaluated relative to their size, location, growth rate, growth mechanism, and likelihood that they are or will become a liability to the function or performance of a component. NDE comprises a large family of specific test disciplines including visual inspection, dimensional metrology, ultrasound, radiography, penetrant test, magnetic particle test, leak test, eddy current tests, potential drop tests, flash and vibro thermography, shearography, acoustic emissions, and many other methods. This article provides discussion on general NDE science and considerations for specific technique selection.

NDE is typically reserved for safety-critical enterprises because any NDE program adds cost to the component. Alternatives to 100% NDE programs include (1) fix or replace after failure; (2) destructive batch sample examination (DE) where if the sample passes, then all items within that batch and with similar manufacturing process parameters are deemed to be acceptable; or (3) sample NDE where random or worst-case/highest susceptibility components are targeted for inspection and serve as harbingers of the condition of similar untested components. Selection of NDE techniques are based not only on the efficacy of any particular method but also on the associated costs. In some cases, safety-sensitive industries such as nuclear or aerospace require NDE of safety-critical parts, but the NDE cannot increase the component cost beyond what the market will bear.

Flaw Detection and Evaluation and Probability of Detection

Before an NDE method is chosen, the inspection objective should be clearly defined. Considerations include:

- The reason(s) for performing the NDE (failure prevention, failure risk minimization, performance enhancement, end-of-life prediction, quality control, etc.)
- The type(s) of flaws or material characteristics of interest (fatigue cracks, stress-corrosion cracks, creep, pitting, erosion, embrittlement, wear, planer cracks/voids, transverse/axial cracks, color variation, density, resonant frequency, etc.)
- The size and orientation of rejectable/reportable flaws
- The anticipated locations of flaws of interest (surface, volumetric, welds, heat-affected zones, high stress points, areas subject to wear, etc.)
- Characteristics of the material being evaluated (hardness, toughness, density, strength, etc.)
- The size, environmental, temporal, and spatial accessibility of the test component.

Typical inspection techniques examine some metric or signal against a threshold level. The signal may be a simple voltage or indicator level or it may be a complex multidimensional image; however, the basic concept of flaw detection is that the signal exceeds a detection threshold. Quantification of this aspect of NDE is known as the probability of detection (POD) (Ref 1). The detection threshold must be set sufficiently low to ensure detection of significant flaws yet not so low that noise or normal variations in the signal frequently exceed the threshold thereby increasing the probability of false alarms (PFA) (Fig. 1). Realistic expectations for any NDE technique appreciates that there is some threshold of flaw size below which it is unlikely for the inspection to detect the degradation. Conversely, there is some signal level at which the chance of flaw detection is very high. NDE methods are designed to allow detection thresholds to be set to maximize the POD while minimizing the PFA. Factors affecting the POD include the signal sensitivity to degradations of interest, the size or severity of degradation that constitutes a critical flaw, instrument and material noise, and human factors. Quantification of all of these factors is possible in cases where a significant number of inspections have been performed and where the population of actual flaws is known and may be compared to the NDE detection record. For some critical examinations, the influence of human factors may be minimized by independent analysis where the data is presented to multiple independent reviews. Any indication disposition discrepancy is subject to an additional review. This type of analysis is common in critical nuclear NDE programs such as steam generator tube inspection (Ref 2).

Certification and Qualification

The efficacy of NDE tests depends on the capabilities of practitioners. Many NDE contracts demand proof of personnel capabilities through certification programs. *Qualification* is the compliance with requirements of certification. These requirements typically include education and training, experience, and demonstration of knowledge evidenced by passing an independently administered examination. *Certification* is the written testimony by the certification body that personnel have met the requirements. Certifications generally fall into two categories—employer certification (as with ASNT) and central body certification (as with ASME and BINDT/PCN). Under employer certification programs, the authority for the certification rests with the employer and the certification ends concurrent with an employee's severance from his employer. Under central body certification programs, the certification remains valid independent of employers. Globally there are numerous formal certification programs (Table 1).

Common practice among most of the certification standards or guidelines is to grade the level of qualification into Levels I, II, and III. Level I individuals are qualified to perform calibrations and evaluations to determine acceptance or rejection in accordance with written instructions and to record those results. Level II individuals are qualified to set up and calibrate equipment and to interpret results with respect to applicable codes, standards, and specifications. Level IIs should also have a working knowledge of the materials that are being inspected as well as the failure modes to which they are susceptible. Level III individuals are capable of developing, qualifying, and approving procedures; establishing and approving techniques; interpreting codes, standards, specifications, and procedures, as well as designating the particular NDT methods, techniques, and procedures to be used. NDT method selection requires the Level III to also have general familiarity with other appropriate NDT methods as well as the method(s) in which he/she is certified. In addition, Level IIIs should be capable of training and examining NDT Level I and II personnel for certification.

In addition to qualification and certification of personnel to perform NDE, each specific examination procedure should be demonstrated to be able to detect the target flaw or anomaly for which the test is intended. In the ASME Code, qualification is performed at one of three levels of rigor. For a low rigor qualification, only a technical justification report explaining why the inspection method will detect flaws of interest is required. An intermediate rigor qualification also requires a successful performance demonstration on a limited number of test specimens in order to achieve an acceptable POD and PFA score. A high rigor qualification requires a technical justification report plus a successful performance demonstration on blind samples. The procedure qualification technical justification may be accomplished by demonstration of performance on representative and worst-case flaws. Frequently, the justification that the target volume is covered and that the representative flaws envelop the flaws of concern throughout the area of interest is addressed by physics-based models of the part and the interaction of the inspection technique with that part. In the simplest form, a reference standard is prepared with a flaw or series of flaws representative of typical flaws of interest and particularly including flaws near the threshold of detection. The NDE method is applied to this reference standard to demonstrate the efficacy of the method and of the practitioner to detect and, if required, to locate and size the reference standard flaws. Depending on the level of rigor and the specific application expectation, the inspection procedure can be demonstrated on open samples where the inspector has prior knowledge of the flaw geometry. In these cases, the inspection documentation must be clearly explained to an examiner competent in the method that the inspection procedure is adequate to reliably detect, locate, and if necessary size the flaws of interest. Alternatively, the reference standard may contain blind flaws whose details are not known by the inspection personnel. It is left to the lead inspection personnel to correctly detect, locate, and size the flaws in the reference standard (Ref 3). Evaluating this kind of test does not necessarily require the examiner to have intimate understanding of the NDE method or analysis procedure as he/she may simply grade performance based on the correlation between the inspection results and the actual flaws of the blind sample. Following qualification of the inspection procedure, a separate step is normally performed to qualify personnel to correctly interpret and apply the procedure. Personnel qualifications are normally performed on blind test samples or simply on recorded data where there is no a priori knowledge of what flaws may be present, what are the flaw sizes, or where the flaws are.

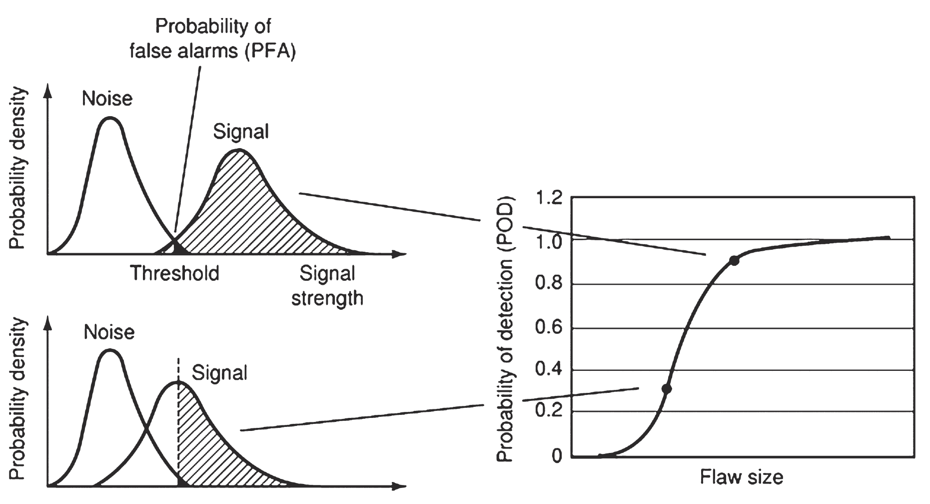


Fig. 1 Probability of detection (POD) concepts and the probability of false alarm (PFA) are determined by fractions of signal and noise distributions above a threshold. Signal distribution generally shifts to higher levels as flaw size increases, leading to the sigmoidal POD curve.

Codes and Standards

There are numerous codes and standards for inspection. Examples are shown in Table 2. They range from being quite specific to rather general in nature. Information can include details on calibration standards, inspection frequency, guidance on how to perform inspections, applicability, mandatory and nonmandatory practice, tips on where to focus inspections to align with likely areas of damage or degradation, and a number of other aspects of inspection. NDE personnel should be familiar with codes and standards related to their industry and consider the code and standard guidance when applicable. As a practical matter, if an inspection is performed in accordance with some standard, it may be more efficient to cite the standard for some details of how the inspection was performed. Such a standard reference may also add credibility to the examination.

Codes and standards do not address all aspects of NDE and cannot replace education, experience, and the use of engineering judgment. Moreover, new methods are continually being introduced to improve NDE accuracy, reduce schedule, or reduce cost compared with traditional methods. Some examples of new methods that are immerging in the industry as of this writing (2017) are:

- *Wave field analysis* where an ultrasound emitter generates an acoustic wave and the behavior of the wave is mapped over the

Table 1 Nondestructive evaluation method and personnel qualification standards and guidelines (not comprehensive)

Standard/guideline	Ref
ASNT SNT-TC-1A	3, 4
ASME ANDE	5
API 1104	6
BINDT/PCN	7
AIA NAS 410	8
European Nations EN473/ISO 9712 (ENIQ)	9

part surface using scanning contact transducers, water or air immersion transducers, or laser velocimetry sensors.

- *Guided wave ultrasound* using piezoelectric, EMAT, or magnetostrictive sensors. This is a growing inspection approach because of advancement of techniques for focusing, steering, temperature, and dispersion curve compensation, and other advances that are leading to increased confidence in this technology.
- *Nonlinear ultrasound* exploits nonlinear waveform distortion of the primary excitation frequency that typically manifests in a first harmonic or 2× the primary frequency. Filtering for this harmonic response can enhance sensitivity to some conditions of interest including some types of medical imaging, nuclear reactor hydrogen embrittlement, and titanium diffusion bonds.
- *Various thermographic techniques* taking advantage of higher resolution temperature discrimination and new ways to induce thermal gradients including sonic, flash heating lamps, inductively coupled eddy currents, and lasers.
- *Visual image processing algorithms* designed to detect subtle differences between a reference part image and an inspection object or between images of a part taken after a time interval looking for indications of change.

The primary guiding principle of any inspection whether it is well addressed by industry codes and standards or a completely new method is that the NDE-specific procedure must be demonstrated to be able to detect and, if required, size any defects of interest. Demonstration of the efficacy of any NDE method typically follows a rigorous logical process whose steps include:

- Definition of the minimum target size and overall range of flaw, defect, material characteristic anomalies, or measurement property of interest. Typically this is performed in conjunction with responsible component designers who understand the component design stresses plus fracture mechanics experts who can help assess likely failure mechanisms, stress concentration factors, expected flaw growth rates, strength design margins, frequency of inspection, and so forth.
- Identification of the material to be inspected plus the geometry of interest and any preferential flaw location if applicable.
- Producing an actual and/or analytical representation of the component to be inspected with a range of flaws, defects, or anomalies preferably including examples near the minimum target size.
- Examining the produced sample(s) and either justifying the detectability to a referee who is competent in the NDE discipline to corroborate claims of detectability or by demonstrating successful detection of flaws of interest through blind test samples.
- If flaw sizing (usually wall thickness, crack length, crack width, or flaw volume) is also of interest, the accuracy of measurements is usually determined by a regression analysis of estimated or measured versus true flaw sizes. Errors are typically characterized as the root mean square error (RMSE) of the NDE flaw size estimate (Fig. 2), which is conveniently in the same units as the measurement.

NDE is a mature engineering science where inspection methods have been developed based on a number of well-understood physics principles. An overview of these methods with their applications, limitations, and advantages are shown in Table 3. This is intended only as a noncomprehensive overview. Most of these technologies are more fully addressed within other articles in this Volume.

Table 2 Examples of nondestructive evaluation codes, standards, and industry guidelines (not comprehensive)

Codes, standards, and industry guidelines	Ref
ASME Boiler and Pressure Vessel Code, Section XI: Rules for Inservice Inspection of Nuclear Power Plant Components	11
ASME Boiler and Pressure Vessel Code, Section V: Nondestructive Examination	12
ASME Boiler and Pressure Vessel Code, Section III: Subsection NB Class 1 Components – Rules for Construction of Nuclear Facility Components	13
API 1104, Welding of Pipelines and Related Facilities	7
ACI 349.3-02, Evaluation of Existing Nuclear Safety Related Concrete Structures	14
ACI 228, Nondestructive Test Methods for Evaluation of Concrete in Structures	15
International Atomic Energy Agency Guidebook on Nondestructive Testing of Concrete Structures	16
AC 43.13-1B, Acceptable Methods, Techniques, and Practices – Aircraft Inspection and Repair	17
MIL-HDBK-6870B, Nondestructive Inspection Program Requirements for Aircraft and Missile Materials and Parts, 2012	18
ASTM E164, Standard Practice for Contact Ultrasonic Testing of Weldments DoD Adopted	19
ASTM E213, Standard Practice for Ultrasonic Testing of Metal Pipe and Tubing	20
ASTM E376, Standard Practice for Measuring Coating Thickness by Magnetic-Field or Eddy-Current (Electromagnetic) Testing Methods	21
ASTM E498/E498M-11, Standard Practice for Leaks Using the Mass Spectrometer Leak Detector or Residual Gas Analyzer in the Tracer Probe Mode	22
ASTM E1742/E1742M, Standard Practice for Radiographic Examination	23
ASTM E2033, Standard Practice for Computed Radiology (Photostimulable Luminescence Method)	24

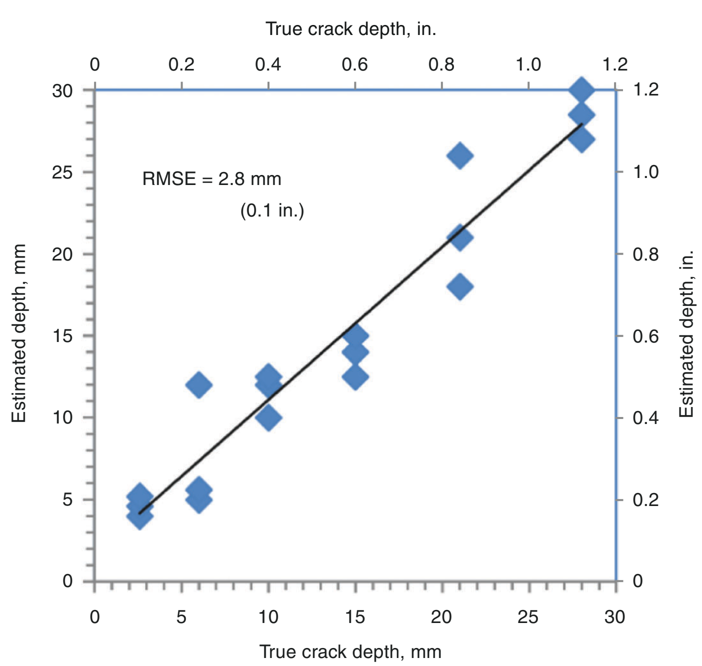


Fig. 2 Phased array UT depth sizing regression (in mm) for dissimilar metal weld test blocks. Source: Ref 25

REFERENCES

1. R.B. Thompson, Using Physical Models of the Testing Process in the Determination of Probability of Detection. *Materials Evaluation*, Vol 69 (No. 7), 2001, p 861–865
2. "Steam Generator Management Program: Pressurized Water Reactor Steam Generator Examination Guidelines: Revision 8," TR-3002007572, presented at EPRI 2016 (Albufeira, Algarve, Portugal), Electric Power Research Institute, October 2016

Table 3 Overview of various nondestructive evaluation methods and categories of tests (not comprehensive)

Method	Principle of operation	Application	Limitations	Advantages	Materials
Visual VT—direct viewing, borescopes, video, magnifying glass, speckle metrology	Visual observation of the test object surface to evaluate dimensions, color, and the presence of surface discontinuities indicative of defects	Postmanufacturing inspections, in-service inspections looking for dimensional anomalies, color variation, cracks, pits, etc.	Only sensitive to surface flaws	Low cost, intuitive. May be enhanced with PT, MT, IR, UV, laser reference lines, video, and magnifying lenses	All
Thermal/IR—includes passive thermal inspections, flash thermography, and laser flying spot/line technology	Surface temperatures are imaged by point sensors, IR cameras, or coatings that change appearance with temperature. Analysis is based on absolute and change of temperature following a stimulus.	Locating near-surface defects that alter heat transfer through object	Requires precise control of environment, heat input, and high-resolution instrument. IR systems sensitive to surface emissivity	Rapid examination of large areas. Can be adapted to production inspections	Circuit board solder joints, solar cells, heat transfer equipment, metals, composites, concrete
Sonic IR/vibro-thermography, inductive IR	Acoustic or inductive energy is applied to part and cracks heat up more than surrounding regions.	Detection of surface-breaking cracks	Acoustic energy is mode and shape sensitive so nonuniform in heating effect. Also no crack depth information	Same as thermal/IR above	Mostly metals and composites
Liquid penetrant PT	Liquid penetrant fluid preferentially collects in crevices and attracts dyes that accentuate flaw visibility.	Locating fabrication discontinuities, stress cracks	Will not find subsurface defects	Ease of application. Improvement over simple visual inspection	Most metals and composites with nonporous surfaces
Magnetic particle MT	Magnetic particles are attracted to breaks in magnetic lines of force.	Near-surface flaws sensitive to magnetization. Includes blow holes, laps, and cracks	Not applicable to nonmagnetic metals or material	Can detect flaws up to ¼ inch below surface under good conditions	Iron and steel, nickel, cobalt
Ultrasonic UT—phased array, guided wave, wave-field, EMAT, magnetostrictive (MS) piezoelectric	High-frequency vibrations are introduced into sample. Waves are reflected or scattered by discontinuities. In many cases, UT may be used to detect, size, and image flaws.	Cracks, laps, thickness/wall thinning, pitting, erosion/corrosion, voids, delamination. Works best on parts with parallel sides	Sensitivity is reduced by rough-surface parts, odd-shaped pieces are difficult to analyze, and a skilled operator is required. Most exams require fluid coupling.	Can be applied either in through-transmission two sides or reflection mode single side. Good sensitivity in acoustically clear material	All metals and alloys, sintered carbides, glass and ceramics, rubber, structural plastics, concrete
Electromagnetic ET—ACFM, remote field ET, flux leakage, pulsed ET PEC	Measures impedance of coil close to conductive material. Coil impedance changes in relationship with cracks, inclusions, etc., affecting induced eddy currents in the material.	Material identification, crack detection/sizing to 1–2 mm, erosion, corrosion, pitting, wall thinning on thin-wall material, surface profiling	Absolute measure difficult. Normally for qualitative comparison with reference sample. Part edges complicate results, and a skilled operator is required.	No fluid coupling required. Can adapt to high-speed production lines	Metals only
Potential drop (PD)	Measures electrical resistance between/among probes contacting material surface	Online monitoring of material degradation between electrodes	Not suitable for scanning due to irregular contact resistance with surface	High sensitivity to small resistance change. Suitable for online monitoring	Mostly metals but limited use for concretes and composites
Barkhausen noise	Changes in magnetic flux from stress applied to magnetic material	Online monitoring of magnetic material	Not suitable for nonmagnetic material	Online monitoring, can sense stress without cracking	Magnetic metals steel, nickel, iron, chrome
Penetrating radiation RT, computed tomography (CT), digital x-ray, neutron RT, x-ray diffraction	Penetrating rays x-ray, γ, or neutron passing through or reflecting from test object cast shadows or patterns on film or digital imaging plates	Manufacturing, weld inspection, finding objects in closed containments, metrology of enclosed objects, thickness	Hazardous radiation operation, not sensitive to defects less than 1–2% thickness of total metal, complex shapes are difficult to analyze	Permits visual analysis of buried defects or components in assembly. Also possible to measure near-surface strain	Metals, foods, films, nonmetals, composites, assemblies
Acoustic emission (AE)	Multiple distributed sensors detect and triangulate AE stress wave source in response to mechanical or thermal stress.	Corrosion, stress-corrosion cracking, weld cracking, creep and fatigue cracking	Sensitive to noise and vibration; identifies location of defect rather than type of defect	Allows the whole volume of the structure to be inspected nonintrusively in a single loading operation	Aircraft, bridges, welds, metal forming, composite and metal pressure vessels, piping
Replica two-part silicon compound (now more widely used than acetate tape techniques)	Two-part liquid silicon-based compound is applied to surface, allowed to set, then removed for caliper, optical, or laser measurements.	Crack widths down to $\approx$100 μm	Difficult to apply on ceiling or wall-oriented surfaces	Simple and intuitive	Metals, composites, virtually any nonporous surface
Laser/dimensional metrology photogrammetry, laser total station, laser tracker, coordinate measurement machines	Either angle/angle stereo image or angle/time-of-flight allows measurement of surface profile of the object.	Reverse engineering, dimensional metrology, crack and pit detection	Limited to visible surfaces	Intuitive, can be applied over wide scale from micrometers to kilometers	Any solid structures including concrete, aircraft, nuclear components, etc.
Vibration analysis	Accelerometers, velocimeters, and/or displacement sensors measure local displacements. Analysis includes amplitude, frequency, and differential measurements	Changes may indicate weakening or failed structural components, alignment issues, bearing wear, shaft crack, imbalance.	Indirect indication of degradation or failure	May provide advance warning of degradation prior to catastrophic failure	Piping systems, rotating machinery
Leak detection—acoustic, pressure drop, or gas sniffing	Most certification programs deal with helium or other gas sniffing using mass spectrometer or other instrument.	Pipes, vessels, heat exchanger tubes, tanks	Some methods such as pressure drop only confirm leak presence/absence. Other techniques help locate leak.	…	Any gas or liquid container, tank, vessel, or pipe including metal, concrete, and composite
Microwave, millimeter wave, terahertz, ground penetrating radar	Electromagnetic radiation from ~10 GHz to 10 THz used in either through-transmission or reflectance mode	Airport body scanners, space shuttle foam inspection, in situ NDE of concrete rebar, moisture detection in composite, paint thickness	Cannot penetrate conductive metals	High-quality imaging can be performed at distance. No ionizing radiation so harmless to people at common energy levels	Paper, plastics, concrete, ceramics, fabrics, wood, paint

VT, visual test; PT, liquid penetrant test; MT, magnetic particle test; IR, infrared; UV, ultraviolet; UT, ultrasonic testing; EMAT, electromagnetic acoustic transducer; ET, electromagnetic testing; ACFM, alternating current field measurement; PEC, pulsed eddy current; RT, radiographic testing

3. "Boiler and Pressure Vessel Code, Rules for Inservice Inspection of Nuclear Power Plant Components; Mandatory Appendix VIII, Performance Demonstration for Ultrasonic Examination Systems, ASME, 2015
4. "Recommended Practice No. SNT-TC-1A: Personnel Qualification and Certification in Nondestructive Testing," American Society for Nondestructive Testing, 2006
5. "Standard for Qualification and Certification of Nondestructive Testing Personnel," ANSI/ASNT CP-189, American Society for Nondestructive Testing, 2016
6. "Nondestructive Examination and Quality Control Central Qualification and Certification Program," ANDE-1, ASME, 2015
7. "Welding of Pipelines and Related Facilities," API STD 1104, American Petroleum Institute, 2014
8. "General Requirements for Qualification and PCN Certification of NDT Personnel," The British Institute of Non-Destructive Testing, 2016
9. "NAS Certification & Qualification of Nondestructive Test Personnel, Rev. 1," AIA/NAS NAS410, Aerospace Industries Association of America, 2014
10. "*Nondestructive Testing — Qualification and Certification of NDT Personnel — Fourth Edition*, ISO 9712, International Organization for Standardization, 2012
11. "Boiler and Pressure Vessel Code, Section XI: Rules for Inservice Inspection of Nuclear Power Plant Components," ASME, 2015
12. "Boiler and Pressure Vessel Code, Section V: Nondestructive Examination," ASME, 2015
13. "Boiler and Pressure Vessel Code, Section III: Rules for Construction of Nuclear Facility Components." ASME, 2015
14. "Evaluation of Existing Nuclear Safety Related Concrete Structures," ACI 349.3R, American Concrete Institute, 2002
15. "Nondestructive Test Methods for Evaluation of Concrete in Structures," ACI 228.2R-13, American Concrete Institute, 2013
16. *Guidebook on Nondestructive Testing of Concrete Structures*, Training Course, Series No. 17, International Atomic Energy Agency, Vienna, Austria, 2002
17. *Aircraft Inspection and Repair*, FAA-AC43.13-1B, Skyhorse Publishing Inc., New York, 2009
18. "Nondestructive Inspection Program Requirements for Aircraft and Missile Materials and Parts," MIL HDBK 6870B, Department of Defense, Air Force Research Laboratories, 2012
19. "Standard Practice for Contact Ultrasonic Testing of Weldments," ASTM E164-13, ASTM International, 2013
20. "Standard Practice for Ultrasonic Testing of Metal Pipe and Tubing," ASTM E213-14, ASTM International, 2014
21. "Standard Practice for Measuring Coating Thickness by Magnetic-Field or Eddy-Current (Electromagnetic) Testing Methods," ASTM E376-11, ASTM International, 2011
22. "Standard Practice for Leaks Using the Mass Spectrometer Leak Detector or Residual Gas Analyzer in the Tracer Probe Mode," ASTM E498/E498M-11, ASTM International, 2011
23. "Standard Practice for Radiographic Examination," ASTM E1742/E1742M-12, ASTM International, 2012
24. "Standard Practice for Computed Radiology (Photostimulable Luminescence Method)," ASTM E2033-99 (R 2013), ASTM International, 2013
25. B.G. Braatz, P.G. Heasler, R.M. Meyer, "PARENT Quick Blind Round-Robin Test Report," PNNL-22677, Pacific Northwest National Laboratory, September 2014

ASM Handbook, Volume 17, *Nondestructive Evaluation of Materials*
Aquil Ahmad and Leonard J. Bond, editors
DOI: 10.31399/asm.hb.v17.a0006463

Probability of Detection

Eric Lindgren, United States Air Force AFRL/RXCA
David Forsyth, TRI Austin
John Aldrin, Computational Tools
Floyd Spencer, sfHire

NONDESTRUCTIVE EVALUATION (NDE) methods described in this Volume are used extensively for the inspection of multiple items to detect defects either during or after manufacturing. This ensures there are no defects that could compromise the intended performance or, after a period of use, no defect has formed that can compromise the integrity of that part. As such, the use of NDE methods is identified as a method to help ensure safety. The consequences of not performing these inspections correctly can be severe, including the loss of life.

Probability of detection (POD) was established to assess the performance of an NDE-based inspection. Probability of detection is a method used to determine the capability of an inspection as a function of defect type and defect size. This section is intended to provide an overview of the concept of POD, why it is needed, the history behind the development of POD, how POD assessments are performed, how modeling and simulation can be integrated into the execution of a POD assessment, several cases to illustrate how studies have been performed to date, and some thoughts on future areas for research and development in the area of POD. Several of these areas represent an evolution of capability or are new since the previous edition of this Volume and other handbooks were published, such as in-depth testing of model-assisted methods.

With this information, the reader should gain an introduction to the theory and practice of POD as well as an understanding of the components required to perform a POD assessment. Because the determination of POD is a statistical concept and therefore is a bit different than standard engineering, it is not uncommon for misunderstandings to evolve in the application, performance, or interpretation of POD assessments. Thus, to help provide additional clarity, several case studies in the use of POD, the analysis tools that are common in POD determination, and the interpretation of the outcome are included in this article. However, a strong word of caution if the reader plans to embark on a POD study: It is not a simplified exercise in which data are collected, placed into a formula, and an outcome is calculated. As every engineer is aware, using a formula without properly understanding the assumptions when performing the calculation can lead to an incorrect answer. As an example, statistical concepts require a minimum number of samples to provide a meaningful answer. However, a mathematical formula does not care if the number of samples meets the required number to satisfy the assumption. Therefore, if the reader is not familiar with the statistical concepts being presented, they are strongly encouraged to seek input from someone trained in statistical concepts before initiating the journey of planning and executing a POD assessment.

As with any summary, the information presented here provides an introduction to the overall scope of POD. A list of references provides additional resources for the reader. An excellent supplement to the material presented here is the Military Handbook (MIL-HDBK) 1823, Revision A, titled *Nondestructive Evaluation System Reliability Assessment*, and published in April 2009. This document is written to address the theory and practice of planning and performing a POD assessment, and it is perceived to be the latest and most comprehensive summary available to the general public on this topic.

Motivation

A typical POD assessment is performed to determine the overall capability of an NDE-based inspection process when it is performed as intended in the environment in which the inspection would be performed. Note that there are several descriptive aspects of this statement that will be expanded upon within this section. This includes the determination of the relationship between detection capability and the defect size, commonly referred to as the POD curve. This curve is commonly plotted as detection capability in terms of percentage of detection as a function of defect size. This curve can be used as one of multiple inputs to calculate risk of failure or other related safety factors when NDE-based inspections are used as a component of risk management.

It is important to note what a POD assessment can and cannot provide. It can provide a metric of how well an inspection procedure can detect a defect of interest. However, this assessment should mimic the actual inspection environment as closely as possible, including access, temperature, and instructions for the inspector as they complete the inspection. When completed in this fashion, the assessment can provide a metric of the capability of the inspection and identify areas of possible improvement if the outcome of the assessment does not meet the requirements of the inspection. In addition, a POD assessment can be configured to provide insight on how one instrument compares to another when all of the other parameters that influence an inspection are kept constant. However, it is not only the instrument itself that can be evaluated this way. Any factor that affects an inspection can be evaluated under controlled conditions to determine how that factor influences the capability of the inspection.

One such parameter that is frequently cited is human factors. As a short editorial on this subject, it is important to recall that a properly designed POD study will identify any aspect of the inspection process—whether it be the equipment, the written procedure, the environment in which the inspection is being performed, the training required for the inspector, or the access/complexity of the part to be inspected—that can affect the outcome of the inspection. If an issue is detected with any of these factors, it is possible to determine the impact of this factor on the overall capability of the inspection and, ideally, design improvements to minimize the effect of these factors. However, a POD assessment will not determine impact of an inspector who does not perform the inspection as intended. This can be due to a number of issues, including fatigue, personal stressors, or any other factor that can affect their ability to be motivated to perform the task as defined in the written procedures. This aspect of inspection performance is a study of the psychology and physical condition of an inspector, and it becomes the focus of management of the workforce performing the inspections.

With these considerations in mind, the intent is to communicate the theory and practice of performing POD studies that will prepare anyone intending to perform such a study to address the scope and processes that lead to the best possible outcome to meet the intent of the assessment. This includes an appreciation for the flexibility that must be retained as a study is being developed and as the analysis of the results is being performed to ensure that the accomplishment of the POD study avoids the pitfalls that can occur when its statistical nature is not appreciated. One of the stalwarts in NDE reliability, Mr. Ward Rummel, has been quoted frequently as saying, “POD is not a recipe,” communicating that there are no step-by-step instructions that will ensure the assessment will always meet the intent of the POD study. With the information provided in this article, the NDE community will be given greater insight to appreciate the nature of POD determination.

Background and History

The evolution of measurements to help to assure safety of systems has a long history in its development and application. However, the ability to quantify the capability to perform an inspection has been a challenge. This section provides an overview of the history of how this need has developed and the preliminary efforts to address this need. It is followed by additional detail on the development of the mathematical measurement process to measure POD. These sections provide a comprehensive background on the current state of POD assessments.

Requirements for NDT Performance Measurement

In 1900, Wilbur Wright wrote, “I am constructing my machine to sustain about five times my weight and I am testing every piece. I think there is no possible chance of it breaking while in the air” (Ref 1). The first fatal passenger airplane crash occurred on September 17, 1908. Orville Wright was the pilot; he survived, but his passenger, Lt. Thomas Selfridge, did not. As noted in the *New York Herald* on September 18, 1908, Orville Wright had been warned to perform careful inspections of his aircraft to minimize the danger (Fig. 1). Witnesses of the accident suggest that a propeller failed. The propellers had just been installed and were never used before the flight.

A review of conference proceedings and journals related to NDT shows that the idea of the reliability of NDT has been discussed for some time. In 1965, the Advisory Committee on Reactor Safeguards (ACRS) wrote a letter (dated November 24, 1965) to the Atomic Energy Commission (AEC), which spurred a number of developments in NDT in the nuclear power industry (Ref 2). Although it is not directly stated, the letter hints at the close relationship between NDT, fracture mechanics, and lifing:

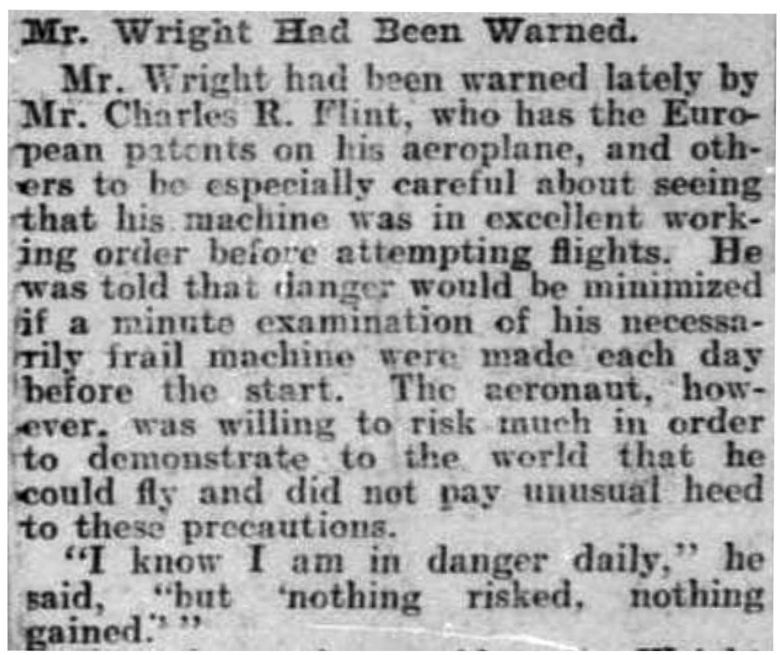
> “(The ACRS) suggests that the industry and the AEC give still further attention to methods and details of stress analysis, to the development and implementation of improved methods of inspection during fabrication and vessel service life, and to the improvement of means for evaluating the factors that may affect the nil ductility transition temperature and the propagation of defects during vessel life.”

Both the National Aeronautics and Space Administration (NASA) and the United States Air Force (USAF) were moving along a similar path at this time. The motivation was the use of philosophies of damage tolerance in design and maintenance: It would be assumed that parts contained discontinuities when they left manufacturing, and these discontinuities could grow as cracks under the expected operation of the part. The other technology advance that enabled this approach was the maturing of fracture mechanics approaches that could be used to model crack growth in metals. The first known publication to make this key connection was by Packman et al. in 1968 (Ref 3). The abstract of this report is as follows:

> “This report describes the work conducted on a program designed to investigate the potential applicability of a combined fracture mechanics-nondestructive inspection procedure as a design approach for aircraft structures. The program consisted of three phases, conducted simultaneously: (1) a literature survey to determine if sufficient fracture toughness information exists to determine a statistically valid value of K_{Ic}; (2) a test program to determine the minimum size of a crack that can be detected by each of four NDT methods: x-ray, magnetic particle, penetrant, and ultrasonics; and (3) a test program to determine if fracture mechanics, when combined with defect size as determined by NDT, can accurately predict the failure load of selected structures.”

NASA would adopt an approach to fracture mechanics and damage tolerance for the Space Shuttle program. The USAF adopted fracture mechanics and damage tolerance in response to accidents in F-111 and C-5 aircraft that were virtually new (Ref 4). The USAF released MIL-STD-1530 on September 1, 1972, to revise their Aircraft Structural Integrity Program (ASIP) to include damage tolerance. MIL-A-83444 was released on July 2, 1974, detailing the requirements for damage-tolerance for airplane safety of flight structure. The nuclear industry also adopted a fracture-mechanics-based approach for risk management, implemented in the ASME Boiler and Pressure Vessel Code.

A simple version of the role of inspection in damage tolerance is illustrated in Fig. 2. Assuming a severe initial cracklike discontinuity



Mr. Wright Had Been Warned.

Mr. Wright had been warned lately by Mr. Charles R. Flint, who has the European patents on his aeroplane, and others to be especially careful about seeing that his machine was in excellent working order before attempting flights. He was told that danger would be minimized if a minute examination of his necessarily frail machine were made each day before the start. The aeronaut, however, was willing to risk much in order to demonstrate to the world that he could fly and did not pay unusual heed to these precautions.

“I know I am in danger daily,” he said, “but ‘nothing risked, nothing gained.’ ”

Fig. 1 Excerpt from the *New York Herald*, September 18, 1908

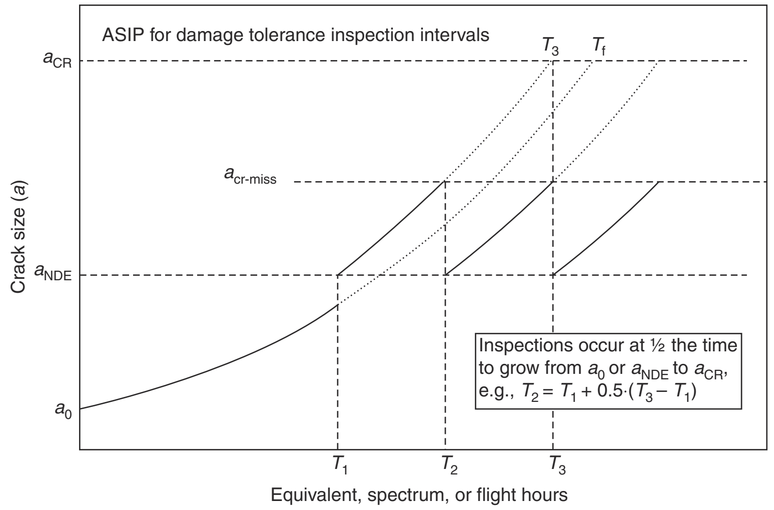


Fig. 2 Illustration of the role of nondestructive testing in a deterministic approach to damage tolerance for structural integrity. ASIP, Aircraft Structural Integrity Program. Source: Ref 5

(often called the rogue defect) (denoted as a_0), the amount of usage for this discontinuity to grow a crack to some critical size (denoted as a_{CR}) is estimated. Inspections are scheduled based on the threshold of NDE capability (denoted as a_{NDE}) to have one or more opportunities in this usage interval to detect the crack and repair or replace the part before failure.

The assumption of the initial discontinuity size a_0 varies depending on the regulatory authority. Aerospace authorities have considered this a property of manufacturing and assigned numbers based on historical evidence and engineering judgment. In some other industries, a final preservice inspection is considered the determining factor for the initial discontinuity size a_0; that is, a_0 is a function of the POD of the preservice inspection.

The approach to damage tolerance illustrated in Fig. 2 shows a single crack size defining a threshold above which all cracks would be found (a_{NDE}). Despite the nondestructive testing (NDT) community defining POD as a continuous function of crack size, the approaches to damage tolerance described previously assumed a single crack size above which all cracks were detected, a_{NDE}. In the 1974 publication of MIL-A-83444, the value of POD to be used for this threshold of detection was chosen to be the 95% lower confidence bound at the POD of 90%. This often is referred to as $a_{90/95}$ or simply the 90/95 value. This value has become the de facto single POD value used to characterize NDT capability.

The acceptance of the 90/95 criterion as a threshold for inspection capability seems to have been due to a confluence of factors:

- A threshold was needed for deterministic fracture mechanics approaches. The USAF and NASA approaches to damage tolerance were such that a crack growth analysis was started from a rogue defect, and the time to crack instability was calculated from this starting point. Inspections are used to detect the crack before instability. To reduce risk, one would prefer 100% POD at crack sizes greater than the rogue defect size. Because 100% POD is not likely at any size of crack, a lesser threshold had to be chosen.
- Binomial statistics were in use to calculate POD and confidence bounds. Using binomial statistics to demonstrate high POD numbers requires many tests. Twenty-nine successes out of 29 trials are required to say one has 95% confidence that the POD is at least 90% at the crack size in question. Many more trials are required to increase this.
- 90/95 was consistent with the USAF MIL-HDBK-5B basis materials allowables and thus familiar to the people working with the approaches to damage tolerance.

The time to initial inspection and subsequent inspections interval shown in Fig. 2 also may vary, depending on the relevant regulatory authority. Many aerospace authorities accept an initial inspection time of half the time to a_{CR} and a subsequent reinspection time of half the time for a crack to grow from a_{NDE} to a_{CR}.

Estimation of NDE Capability

The previous section described the evolution of engineering approaches to structural integrity that require a single detectable crack size estimate of NDE. However, it was recognized very early on that POD is a function of crack size. The first plots of POD as a function of crack size (often called POD curves) were constructed using moving averages or averaging the response of all cracks in an interval and manually fitting a curve through these points (Ref 3, 6–8). The POD at a specific crack size a, denoted POD_a, can be estimated from a series of inspections of cracks of size a as:

$$\text{POD}_a = \frac{n_d}{n} \quad \text{(Eq 1)}$$

where POD_a is the probability of detection at the crack size a, and n_d is the number of cracks of size a detected out of n, the total number of cracks of size a in the trial.

A standard or guideline on how to perform an experiment to estimate a POD curve was not available immediately. Within the American Society for Nondestructive Testing (ASNT), an effort was initiated by W.H. Lewis to develop a recommended practice document. Although the final version was finished in 1976, publication did not occur until 1982 (Ref 9). This recommended practice is based on binomial methods. In the meantime, *Nondestructive Inspection and Quality Control*, Volume 11 of *Metals Handbook*, 8th ed., 1976, included an article from Packman et al. (1976) (Ref 10) describing a number of possible binomial-based methods for POD estimation. It is noted that the ASM International document was developed in concert with the ASNT authors.

The event that spurred an updated statistical approach to POD was the analysis of a large USAF study of the capability of inspectors/inspections being performed at USAF depots in the mid-1970s, widely known as the “Have Cracks, Will Travel” study (Ref 11). In this study, many inspectors inspected each specimen, so it was possible to plot the mean POD for each crack and fit a continuous POD curve through these points. It was noted in the analysis of these data that cracks of the same size were not detected equally: In addition to the variability in a repeated measurement on a single crack, there was significant variability in the response of different cracks of the same size.

Based on their analysis of the “Have Cracks, Will Travel” data, Berens and Hovey (Ref 12) proposed a probabilistic description of POD, where:

- The POD is more than a function of just crack size.
- At any particular size a, the POD of a large number of cracks of size a are distributed approximately as a normal distribution.
- The variance or spread in the distribution is not a function of the crack size.

This definition of POD is illustrated in Fig. 3.

It should be noted that the “Have Cracks, Will Travel” data were recorded only in terms of hit and miss. The previous approach was documented in an American Society for Testing and Materials Special Technical Publication (Ref 13). It is noted therein that these POD models provide better estimates of POD than the previous binomial-based estimates. In particular, the POD estimates using these

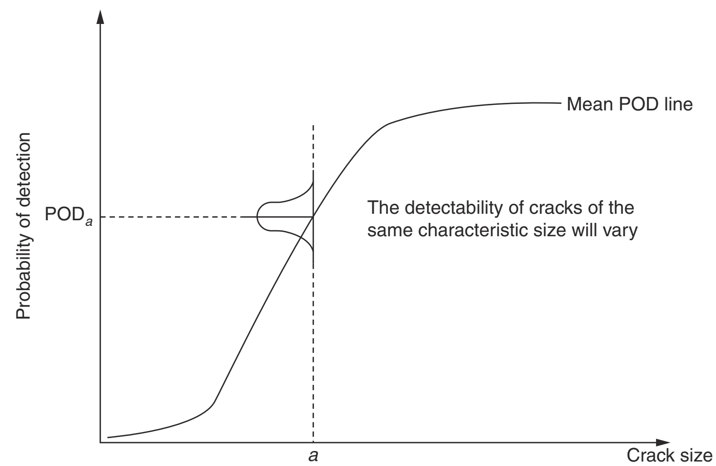


Fig. 3 Illustration of a probabilistic concept of probability of detection (POD). At any crack size, different cracks of the same size will have varying detectability.

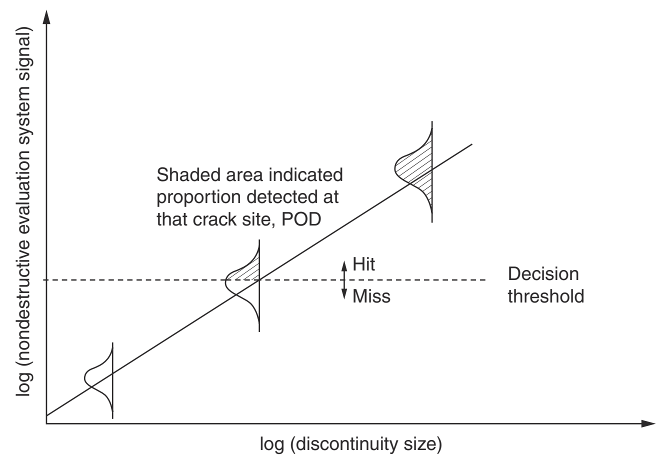


Fig. 4 Illustration of the linear relationship observed between nondestructive testing signal magnitude and crack size in automated eddy current bolt hole inspections by the United States Air Force. POD, probability of detection

analysis methods converge with significantly less data.

Further evolution of POD was driven by the USAF Engine Structural Integrity Program (ENSIP). The ENSIP community could not obtain acceptable component lives and inspection intervals with the use of large rogue defect sizes accepted by the ASIP community. Automated eddy-current testing (ET) inspection systems were developed to attempt to improve POD over manual systems. This allowed the collection of ET signal magnitudes and corresponding crack sizes (known as "*a*-hat," or $\hat{a}$, versus *a* data). The review of the data collected by these systems showed that the data generally were linear on a log-log scale, with variance normally distributed around the mean, independent of crack size (Fig. 4).

These types of data are well suited to regression analysis, and this approach to POD estimation was documented in *Nondestructive Evaluation and Quality Control*, Volume 17 of *Metals Handbook*, 9th ed., 1989 (Ref 14). Up to this time, the USAF did not have an internal document defining the acceptable method for determining the a_{NDE} crack size to be used in support of ASIP or ENSIP, although the previously mentioned works were widely used. The ENSIP community supported the development of MIL-HDBK-1823 to document the $\hat{a}$-versus-a approach. This work was first published as NATO AGARD-LS-190 in 1992. In this reference and in MIL-HDBK-1823, the methods described in Ref 14 are enhanced. In particular, the latest edition of MIL-HDBK-1823 contains improved solutions for regression models and confidence bounds.

All of the models referred to so far require two parameters, which define the location (i.e., the value of the crack size at 50% POD) and the slope of the POD curve. Both log-odds and log-normal start at 0 for crack size 0 and are asymptotic to 1 for large crack sizes. There are at least two reasons to consider further changes to the models:

- It has been noted in the literature that these models are sensitive to situations of large false call rates and/or detections at small crack sizes. In both cases, raising the POD at small crack sizes has the corollary of reducing it at large crack sizes, despite any experimental data otherwise (Ref 15–17).
- It often has been argued that POD should not go to unity; that is, there always is some chance a crack will be missed by the system. This has been demonstrated by certain large POD experiments (Ref 17) and reviews of inspection performance in aircraft maintenance depots (Ref 18).

In response to these issues, an extension of the two-parameter models was proposed (Ref 19), adding two additional parameters that allow the POD curve to start at values greater than 0 and to terminate asymptotically at values less than 1. This can be written as:

$$\mathrm{POD}(a) = \mathrm{p_h} + (1 - (\mathrm{p_m} + \mathrm{p_h})) \cdot \mathrm{F}(a;\, \mu,\, \sigma) \qquad \text{(Eq 2)}$$

where POD(a) is the mean probability of detection at the discontinuity size a, $\mathrm{p_h}$ is the false call probability, $\mathrm{p_m}$ is the probability of missing a discontinuity independent of discontinuity size, and $\mathrm{F}(a;\, \mu,\, \sigma)$ is the two-parameter distribution used to fit the data, that is, log normal.

Maximum likelihood estimators can also be used to find estimates of $\mathrm{p_m}$ and $\mathrm{p_h}$ from Eq 2. If the magnitudes of either of these parameters are much greater than 0, it may indicate that the model of Eq 2 does not adequately fit the data.

In application, it has been found that the lower asymptote, as one logically would expect, fits data with moderate false call rates better than a lower asymptote of 0. Certain data sets have shown to be a better fit by an upper asymptote of less than 1, although this in general indicates problems of process control rather than inspection capability. The four-parameter model has not been widely used or included in standard documents to date, despite its demonstrated success in fitting data from large POD experiments.

Confidence bounds have been calculated on POD estimates since the early damage-tolerance standards were written. The POD curves calculated by the methods of the aerospace, nuclear, and petrochemical industries all yield the mean POD as a function of crack size. The confidence bound provides a statistical estimate of the accuracy of the mean POD calculation and is highly influenced by the number of samples used in the POD study. The USAF MIL-HDBK-1823 provides methods for confidence bound calculation for two-parameter POD estimates.

While the quality of the mean POD estimate is of interest, it also is of interest to know what variability in POD results can be expected in a given situation. Prediction bounds, tolerance bounds, and quantiles are accepted statistical techniques that can be used to answer questions such as, "What is the worst performance I can expect from an inspection that belongs to the population that has this mean POD curve?" (Ref 20, 21).

Methods

This section describes the methods by which POD is determined. This includes greater detail on the experimental process to acquire the needed data, the mathematical methods to obtain a POD curve, and techniques to assess the uncertainty in the POD curve as it is obtained from a limited data set. In addition, the concept of model-assisted POD (MAPOD) is introduced, with additional details and representative examples of MAPOD included later in this article.

POD Experiments

The standard reference document for conducting a POD experiment is the United States Department of Defense MIL-HDBK-1823. The difficult questions of experiment design usually are based on resource limitations: What is the best experiment that can be performed with available time and funding?

The fundamental requirement of the experimental plan is to understand what variables affect the POD for the situation at hand. These variables must be appropriately fixed or controlled to represent the target application, if possible. For example, the instrument used for an inspection may be fixed to be a certain type. If an important variable cannot be controlled, then it must be represented in the experiment (for an obvious example, the use of a set of inspectors taken from the population of inspectors).

Most of the literature on this subject is concerned with the defects used in a POD study. Usually, the manufacture of controlled test specimens is 50% or more of the total cost of a POD study. How the defects are manufactured is important, and how many are included is important. Fatigue cracks are a common inspection target, and it has been found in numerous studies that the variability between fatigue cracks is a key driver of POD (Ref 12, 19). How these defects are manufactured also can affect the NDT response (Ref 22), and this must be considered carefully. The choices always should mimic the expected conditions of the actual inspection subjects.

If it is too difficult to manufacture a large set of naturally occurring defects, the transfer function approach (see the section "Transfer Function Example" in this article) can be used. An example could be the development of fatigue cracks in large or expensive structures. Simple structures and simpler defect representations, such as electrodischarge-machined (EDM) slots, could be used to approximate fatigue crack response, with the caveat that some transfer function on the NDT signal response is required to approximate the difference.

Another key experiment choice is the independent variable. That is, what is the POD measured against? This typically is a measurement of crack size (length or depth). It could be either, or it could be some other measurement (i.e., area). An important consideration for this choice is understanding the intended use of the POD curve. If the POD is needed to support an engineering assessment of structural integrity, then the structural engineer typically will have a measurement of interest for the POD. Another important consideration is the NDT-defect interaction itself. Sometimes, the POD is not a strong function of a simple crack dimension. There are examples of data in the literature that show this effect for certain penetrant and eddy-current testing situations (Ref 23, 24).

Test procedures must be fixed. It is not unusual for undocumented "tribal knowledge" to exist in organizations; this must be either explicitly expressed in techniques or eliminated. Interpretation of signals and images for defect detection can be complicated. While criteria may be expressed in a simple screen height metric, inspectors also are doing pattern recognition to look at more complex signal shapes. This simply is human nature. Interpretation of images is easy for humans to do but difficult to control and to express as a quantitative NDT response analogous to a percent screen height. The method of image evaluation and detection criteria must be fixed. If there is no simple way to express the detection as a single variable, then hit/miss POD analysis can be used.

While false calls are not an explicit part of a POD estimation, they must be included to properly understand the performance of a technique. Otherwise, POD can be easily improved by reducing decision thresholds. In some applications, false call rates are obvious to define; for example, in the case of detection of cracks at a fastener hole, a false call rate in terms of false calls per uncracked hole is an obvious choice. If the inspection is over an area, such as the inspection of a large structure for local corrosion damage, the definition of false call rate could be per unit area. In either case, there must be significant nondefect opportunities for inspection in the POD experiment. The acceptable false call rate is a decision based on risk and cost trade-offs. False call rates above approximately 5% cause problems with simple two-parameter POD curve fits and must be addressed using three- or four-parameter POD curve fits (Ref 19).

POD Fitting

Mathematical Basis for POD. The basic model behind the derivation of a POD curve is captured in Eq 3, where a function of the signal strength, S, is related to a function of the defect dimension, a, in an average sense, and the departures from the average are represented by the random variable, ε:

$$g(S) = f(a) + \varepsilon \qquad \text{(Eq 3)}$$

A detection of the defect occurs if the signal exceeds a threshold, T. Assuming that the function $g(S)$ is monotonically increasing, then $S \geq T$ if and only if $g(S) \geq g(T)$. Therefore:

$$\begin{aligned}\mathrm{POD}(a) &= \mathrm{Prob}\,(g(S) \geq g(T)) \\ &= \mathrm{Prob}\,(\varepsilon \geq g(T) - f(a)) \qquad \text{(Eq 4)}\end{aligned}$$

The form of Eq 4 demonstrates that the probability of detection is ultimately derived from the distribution of the random variable that characterizes signal behavior about a mean.

Defect-to-defect variation constitutes the major source for the variation of signals. That is, two distinct defects of the same size will produce different signals due to defect morphology differences, material differences, and so on. A second source of variation is that of the inspection itself. That is, reinspection of the same defect using the same nondestructive inspection (NDI) equipment and setup will produce a different signal response. These two sources of variation can be noted explicitly by the expression $\varepsilon = \varepsilon_f + \varepsilon_r$, where ε_f captures the defect-to-defect variation, and ε_r captures the inspection variation.

Parameterizations of POD. The general formulation of signal relationship to defect dimension shown in Eq 3 is made more concrete with the choices of the functions $g(S)$ and $f(a)$ and the specifying of the distributional form for ε. For the purposes of estimating POD curves (discussed in the section "Nonparametric Estimate of POD" in this article), it is important that the residual random variable, ε, has a distributional form that is independent of the defect dimension, a. Thus, the functional forms for $g(S)$ and $f(a)$ generally are made so the mean of ε is 0 and the variation of ε is a constant over the range of defect dimensions.

The most commonly used functional forms for Eq 3 are captured in the following equation:

$$S^{\delta} = \beta_0 + \beta_1 \cdot a^{\lambda} + \varepsilon \qquad \text{(Eq 5)}$$

In Eq 5, either $\delta = 0$ or $\lambda = 0$ would seem to remove the signal or the defect size from the equation. However, because the limit of the linear transformation $(S^{\delta} - 1)/\delta$ is $\ln(S)$ as δ goes to 0, $\delta = 0$ is taken to be equivalent to replacing S^{δ} with $\ln(S)$. The same argument holds for replacing a^{λ} with $\ln(a)$ when $\lambda = 0$.

Consider $\delta \geq 0$; then a detection being made when $S \geq T$ is equivalent to the POD curve:

$$\mathrm{POD}(a) = \mathrm{Prob}\,(\varepsilon \geq T^{\delta} - \beta_0 - \beta_1 \cdot a^{\lambda}) \qquad \text{(Eq 6a)}$$

Taking the distribution of ε to be Gaussian with mean 0 and variance σ^2 yields:

$$\text{POD}(a) = \text{Prob}\left(\frac{\varepsilon}{\sigma} \geq \frac{T^{\delta} - \beta_0 - \beta_1 \cdot a^{\lambda}}{\sigma}\right)$$
$$= \phi\left(\beta'_0 + \beta'_1 \cdot a^{\lambda}\right) = \phi\left(\frac{a^{\lambda} - \mu}{\tau}\right) \quad \text{(Eq 6b)}$$

where $\phi(\cdot)$ is the standard normal distribution function, $\beta'_0 = \left(\beta_0 - T^{\delta}\right)/\sigma$, $\beta'_1 = \beta_1/\sigma$, $\mu = (T^{\delta} - \beta_0)/\beta_1$, and $\tau = \sigma/\beta_1$. The last two equalities in Eq 6(b) are based on the fact that ε/σ is a standard normal random variable, and therefore, the probability of being greater than or equal to x is equal to the probability of being less than or equal to $-x$.

The last two equalities in Eq 6(b) emphasize that only two parameters are needed to define the POD curve. Both forms appear in the statistical literature associated with POD estimation. The POD function as derived in the aforementioned equation can be estimated by estimating the unknown parameters in the regression Eq 5 when signal data are available. The functional form with only two unknown parameters also can be estimated from hit/miss data (hit/miss data being in the form of known defects and associated sizes and whether each of the defects was or was not detected in an inspection).

Statistical Estimation of POD Curves. Estimation of POD curves can be made from two data types: signal data consisting of the data pairs (a_i,S_i) or hit/miss data consisting of the data pairs (a_i,X_i), where i indexes the defects, a_i is the defect size, and $X_i = 0$ or 1 according to whether defect i was not detected or detected. Estimation of a POD curve when signal data are available follows directly from statistical linear regression analysis if the signal and defect size powers, δ and λ, are known.

The functional form for the POD curve given in Eq 6(b) arises from a probit analysis of the binary data (a_i,X_i). In both data cases, the POD curve is estimated by first estimating the unspecified parameters of the POD function. Using a given POD model, the likelihood of the observed data is maximized with respect to the defining parameters. The values of the parameters that maximize the likelihood function are referred to as maximum likelihood estimates (MLEs).

For the binary case where the data are given as the ordered pairs (a_i,X_i), $i = 1, 2, \ldots, n$, let p_i be the probability of a hit; then the likelihood for the individual data point is p_i if it was a hit or $(1 - p_i)$ if it was a miss. This can be written as $L_i = p_i^{x_i} \cdot (1 - p_i)^{(1-x_i)}$. The total likelihood for all the data is the product of the individual likelihoods and is given by:

$$L = \prod_{i=1}^{n} \left((p_i)^{x_i}(1 - p_i)^{(1-x_i)}\right)$$

Maximizing the likelihood function is equivalent to maximizing the log of the likelihood given by:

$$LL = \ln(L) = \sum_{i=1}^{n} \left(x_i \ln(p_i) + (1 - x_i)\ln(1 - p_i)\right)$$
$$= \sum_{\text{hits}} \ln(p_i) + \sum_{\text{misses}} \ln(1 - p_i) \quad \text{(Eq 7)}$$

where p_i is the individual defect probability given by Eq 6(b), and the maximization is with respect to the unknown parameters of the equation form used.

For the response-based model, the overall likelihood equation is the product of the likelihoods associated with the individually observed response signals. Let the data be in the form of $\{(a_i, s_i), i = 1, \ldots, n\}$, that is, a size and a signal response. From the model of Eq 5, the probability that signal response falls in the interval $[l, u]$ is given by:

$$L = \text{Prob}(l_i \leq S_i \leq u_i) = \text{Prob}\left(l_i^{\delta} \leq S_i^{\delta} \leq u_i^{\delta}\right)$$
$$= \phi\left(\frac{u_i^{\delta} - (\beta_0 + \beta_1 \cdot a_i^{\lambda})}{\sigma}\right)$$
$$- \phi\left(\frac{l_i^{\delta} - (\beta_0 + \beta_1 \cdot a^{\lambda})}{\sigma}\right) \quad \text{(Eq 8)}$$

The likelihood of Eq 8 is written as if the signal corresponding to the ith defect of size a_i occurred within an interval. This form is easily translated into the three cases that most often arise from NDI, specifically:

1. A signal was not recorded because it fell below an inspection threshold, S_{thr}, as may occur when data are gathered only for detected defects.
2. The signal was saturated with respect to the recording device and therefore only known to be larger than the saturation level, S_{sat}.
3. A specific signal level is known.

In case 1, $u_i = S_{\text{thr}}$, $l_i = 0$, and the likelihood is $\phi\left(\frac{S_{\text{thr}}^{\delta} - (\beta_0 + \beta_1 \cdot a_i^{\lambda})}{\sigma}\right) - \phi\left(\frac{-(\beta_0 + \beta_1 \cdot a_i^{\lambda})}{\sigma}\right)$ when $\delta > 0$ and $\phi\left(\frac{S_{\text{thr}}^{\delta} - (\beta_0 + \beta_1 \cdot a_i^{\lambda})}{\sigma}\right)$ when $\delta = 0$, that is when the log transform for the signal is used.

In case 2, $l_i = S_{\text{sat}}$, $u_i = \infty$, and the likelihood is $1 - \phi\left(\frac{S_{\text{sat}}^{\delta} - (\beta_0 + \beta_1 \cdot a_i^{\lambda})}{\sigma}\right) = \phi\left(\frac{\beta_0 + \beta_1 \cdot a_i^{\lambda} - S_{\text{sat}}^{\delta}}{\sigma}\right)$.

In case 3, the difference of the upper and lower limits can be taken as the resolution of the signal recording, Δ. That is, $l_i = s_i - \Delta/2$, $u_i = s_i + \Delta/2$. In the limit, as Δ becomes small, the likelihood is proportional to the derivative of Φ. In the case of the standard normal distribution, this would be:

$$\frac{1}{\sigma}\varphi\left(\frac{s_i^{\delta} - (\beta_0 + \beta_1 \cdot a_i^{\lambda})}{\sigma}\right)$$

where $\varphi(x)$ is the standard normal density function.

As with the binary data, the maximization of the likelihood function is facilitated by maximizing the log likelihood, which converts the products of the individual likelihoods into sums of the individual log likelihoods. Finding the parameter values that maximize the function can be done numerically.

The estimated POD curve in both the binary data case and the signal response (regression case) is obtained by substituting the MLEs into the appropriate functions. Thus, for the binary case, $\widehat{\text{POD}(a)} = \phi\left(\widehat{\beta'_0} + \widehat{\beta'_1} \cdot a^{\lambda}\right)$ where $\widehat{\beta'_0}, \widehat{\beta'_1}$ are the values that maximize the log-likelihood equation. For the regression case, $\widehat{\text{POD}(a)} = \phi\left(\frac{\left(\widehat{\beta_0} + \widehat{\beta_1} \cdot a_i^{\lambda}\right) - T^{\delta}}{\widehat{\sigma}}\right)$ which is dependent not only on the fitted parameters $\widehat{\beta_0}, \widehat{\beta_1}, \widehat{\sigma}$ that maximize the log-likelihood equation but also on the signal threshold, T. This allows the estimation of future POD curves that would result with changing thresholds.

The regression nature of signal versus defect size leads naturally to the use of the standard normal distribution function as a POD model. However, it should be noted that other-than-Gaussian assumptions can be made about the error term. For example, using a logistic distribution for the error term gives another form of the POD function that often is used to model binary data.

Nonparametric Estimate of POD. There is much flexibility in the parameterizations of POD curves, as previously discussed, but the question remains as to whether any specific parametric model adequately captures the behavior of a set of POD data. One way to address this issue is to fit POD behavior making the minimum possible assumptions. This is done with hit/miss data by making only the assumption that the POD function is a continuous monotonically nondecreasing function. It is shown in Ref 25 that, with only this assumption, an intuitive grouping algorithm of the data leads to a maximum likelihood estimate of POD at the points at which defects exist in the data set.

The grouping of the data in the given algorithm is intuitive in the sense that if the data do not reflect an increase but the model does not allow for a decrease, then the best fit would be to assume that the values are equal at two defect sizes. This combination of data can be applied sequentially until POD estimates are monotonically nondecreasing.

Comparing a nonparametric estimate of the POD to any given parametric fit provides the basis of assessing where, in the region of defect sizes, the data behavior may depart substantially from what would be predicted by the parametric model.

Generalizations of Basic POD Curve. Using a logarithm transform ($\lambda = 0$) for the defect size in Eq 5 and 6(a, b) ensures that the POD functions will go from 0 at $a = 0$ to 1 as a becomes arbitrarily large. However, in any given NDE implementation, the inspection process may have issues that would imply that the probability does not have to go to 0, nor does it have to be arbitrarily close to 1 for large defects. A natural way to account for this possibility is to have the POD model accommodate additional inspection issues that lead to hit or misses