# Guide for Qualification of Nondestructive Evaluation Procedures

## Engineering Directorate

## Structural Engineering Division

## September 2018

National Aeronautics and
Space Administration

**Lyndon B. Johnson Space Center**
Houston, Texas

Public Release Statement:

This document has been reviewed for Proprietary, SBU, and Export Control (ITAR/EAR) and has been determined to be nonsensitive. It has been released to the public via the NASA Scientific and Technical Information (STI) Process DAA 59601.

# Guide for Qualification of Nondestructive Evaluation Procedures

Prepared by: Ajay M. Koshti
Ajay M. Koshti,
Materials and Processes Branch/ES4
9/04/2018
Date

Reviewed by:
David M. Stanley,
Materials and Processes Branch/ES4
9-6-18
Date

Reviewed by:
John Alred,
Materials and Processes Branch/ES4
9/05/2018
Date

Approved by:
Rachel Kamenetzky,
Chief, Materials and Processes Branch/ES4
09/06/18
Date

| REVISIONS | | |
|---|---|---|
| Revision | CHANGES | DATE |
| New | Original version | Sept. 3, 2018 |
| | | |
| | | |
| | | |
| | | |

# CONTENTS

# 1.0 SCOPE

This document provides guidelines for qualification of NDE procedures by demonstration. The qualification provides minimum requirements for some NDE procedures that meet conditions stated within. The qualification covers those applications where NDE flaw detectability is given as a flaw size. The document is intended comply with NASA-STD-5009 guidelines. It also provides guidelines for NDE procedure qualification methods not covered in NASA-STD-5009 and appropriate approval is required before using these approaches. The guide is not intended to cover qualification of every kind of NDE procedure.

# 2.0 APPLICABILITY

This guideline is applicable to POD demonstration test for qualification of the Special NDE procedures. This guideline is also applicable to limited validation of flaw detectability size in Custom NDE procedures. The guideline also covers a case where both POD demonstration and limited validation are used.

# 3.0 USAGE

## 3.1 General

There are three choices when the desired reliably detectable flaw size for the selected NDE method is smaller than that provided in Table 1 of NASA-STD-5009[6].

1. First choice is to qualify NDE procedure as Special NDE by using POD methods defined in NASA-STD-5009.
2. Second choice is to qualify NDE procedure as a Custom NDE with hybrid of POD and limited validation, if approved by responsible Fracture Control Board (FCB) or the engineering authority.
3. Third choice is to qualify NDE procedure as a Custom NDE with limited validation, if approved by responsible Fracture Control Board (FCB) or the engineering authority.

This guide is intended for use by NASA JSC NDE demonstration administrators and NDE engineers, but it may be used by others as approved by the engineering authority. The result of such demonstration test normally results is calculation of $a_{90/95}$ (or $a_{90/95\_min}$) flaw size, a Custom NDE with limited validation flaw size $a_{lv}$ or just supporting evidence that the NDE procedure is capable of detecting flaws used in the test.

Flaws with $a_{90/95}$ flaw size are detected with directly or statistically calculated 90% POD and 95% confidence and a low POF such as $\leq 0.1\%$. POD methods shall meet NASA-STD-5009 POD requirements. Flaws with $a_{lv}$ flaw size are detected such that $a_{lv} > a_{90/95}$ and low POF such as $\leq 0.1\%$ with high confidence (such as $\geq 95\%$) based on margins assumed. Custom NDE with limited validation shall follow this guide.

The guide is applicable to NDE procedures in most methods including but not limited to eddy current testing, ultrasonic testing, liquid dye penetrant testing, magnetic particle testing, and x-ray radiography testing. For some NDE procedures, a flaw size parameter instead of actual flaw size dimension, may be more appropriate as signal response used in the flaw call has linear correlation with the flaw parameter. The flaw size may be given as flaw length, depth, area, or flaw parameter. For dye penetrant testing, surface crack length is normally taken as the flaw size. A typical flaw size parameter used in x-ray radiography testing is crack depth-to-part thickness ratio. A typical flaw size parameter used in flash thermography testing is delamination width-to-delamination depth ratio. The guide assumes that signal response-to-flaw size correlation is approximately linear and noise is constant in region of the target flaw size. The signal response may be measured as full screen height (FSH) or dB in ultrasonic testing; vertical, horizontal or vector voltage in eddy current testing, gray value in digital radiography, film density in film radiography etc. This guide is not applicable, if signal response is not monotonic with the flaw size or noise is not constant with flaw size; and if conventional POD model per MIL-HDBK-1823[1,2] is not applicable in the neighborhood of target flaw size.

## 3.2 Custom NDE with Limited Validation Applicability

In some circumstances, Custom NDE with limited validation may be appropriate. Custom NDE is not defined in NASA-STD-5009. Custom NDE with limited validation is not intended to reduce number of flaws below 34 which is a minimum number needed in a Binomial point estimate POD method[3,4,7]. Therefore, Custom NDE with limited validation is not considered to be a substitute for NASA Binomial point estimate demonstration. In order to verify NDE procedure and operator skill including repeatability, it is recommended that the demonstration test has a minimum of 34 flaws.

Use of Custom NDE with Limited validation is recommended where full POD validation is either impractical, cost prohibitive or the POD approach is not provided in NASA-STD-5009. Use of Custom NDE in place of Special or Standard NDE requires approval from responsible FCB or the engineering authority. When approved, Custom NDE with Limited validation of flaw size may be used for damage tolerance safe life analysis of FC parts.

Custom NDE with Limited validation situations arise when,

a. NDE procedure qualification has **insufficient number of flaws** per flaw detectability type for completing POD validation i.e. flaws of correct size and quantity to provide calculation of POD/Conf. of ≥ 90/95 for the required flaw size are not available. Although, the limited validation also requires careful choice of flaw sizes and certain minimum quantity of flaws.
b. There is a difference in NDE signal response between real and artificial flaws, and artificial flaws are used for calibration; and **transfer function** between real and artificial flaws is used in validation.
c. There are **many flaw detectability types** due to varying part geometry, wall thickness, and surface contour of part; and due to various types, orientations and locations of flaws. Requiring 34 flaws for Binomial point estimate for each extreme or worst flaw detectability type may result in more than 100 flaws making demonstration test less practical for manual testing.

### 3.3 Custom NDE with Limited Validation Basis and Rationale

Custom NDE essentially uses a simpler data analysis and rules of thumb. These rules of thumb such as use of signal-to-noise ratio (SNR) and contrast-to-noise ratio (CNR) and largest missed flaw size, have been in practice in approximate ways for a long time. The limited validation method develops these rules further by introducing decision threshold-to-noise ratio (TNR) and smallest detected flaw size. $a_{\mathrm{lv}}$ is designed to be larger than $a_{90/95}$. Therefore, there is a tread off between validating a larger flaw size by using Custom NDE with limited validation with cost savings and smaller POD flaw size with higher cost, if POD demonstration on real flaws on a real part specimen is a practical alternative.

One of the limited validation methods is based on POD Model. For example, limited validation of applications that use artificial flaws for calibration and use signal response transfer function, is based on curve fit POD model. Some approaches are based on comparing merit ratios to with the same from the POD model and are validated using Monte Carlo sampling. For example, Hit-miss, signal response; and 1D, 2D, 3D data analysis is validated based on Monte Carlo sampling. This guide provides recommended requirements, procedures, and comparative technical analysis of limited validation with POD approaches. List of reference documents are provided that support POD, Binomial point estimate and limited validation in Section 15.0.

See Appendix A and Appendix B for comparative information on NDE procedure qualifications. Qualifications of specific NDE methods are not covered except for x-ray radiography testing. See Appendix G for recommended x-ray radiography nondestructive evaluation requirements.

## 4.0 NDE DATA TYPES

### 4.1 Types of NDE Data

There are two types of NDE data. These are,

4.1.1 Signal Response Data ($\hat{a}$)
  a. Single hit for a single flaw (real time data, A-scan display, example: eddy current impedance plane display)
  b. Multiple hits for a single flaw (1D, 2D or 3D representation of data)

4.1.2 Hit-miss Data (binary 1 or 0)
  a. Single hit

## 4.2 NDE Data Descriptions

4.2.1 Hit-miss Data

a. Normally single hit applications
b. For hit-miss data, no signal response is measured on flaw indication.
c. Uses minimum flaw size decision threshold, flaw length-to-width aspect ratio and visible characteristics are used for making flaw detection call. This may include clustering of two or more indications within edge to edge distance between indications less than diameter of the largest indication.
d. If call is true, it is a hit with value of 1. If there is no detection call for a flaw, then it is a miss with value of 0.
e. Hit-miss data POD analysis is used.
f. It also uses data on false positives i.e. flaw calls where there is no flaw. A POF is determined using false positive calls and opportunities for false positive calls.
g. Examples: Dye penetrant testing, magnetic particle testing, and film radiography

4.2.2 Signal Response - Single Hit

a. Real time data. 2D representation of data is not used.
b. Signal response and background noise are measured.
c. *a-hat* (signal response) versus "*a*" (flaw size) POD analysis is used.
d. Includes minimum signal decision threshold for making a reject call.
e. Noise, signal response and decision threshold are used to determine POF.
f. Examples: Real time testing using ultrasonic A-scan and eddy current impedance display.

4.2.3 Signal Response - Multiple Hit

a. 1D, 2D and 3D representation of signal response data.
b. Current NASA-STD-5009 POD approaches do not account for effect of 1D, 2D and 3D representation of data. MIL-HDBK-1823 does not provide requirements for multiple hit data.
c. When multiple pixels in a cluster or a line provide high contrast for a detected flaw in 1D or 2D representation of data, it provides higher POD/confidence over single hit detection.
d. Examples: Ultrasonic C-scan, eddy current C-scan, x-ray computed tomography 2D images.

# 5.0 FLAW SPECIMENS

## 5.1 General

There are various types and kinds of flaws and various kinds of specimens. NDE procedure qualification requires that flaw size, location, orientation is known for all flaws in the test specimens. There are two kinds of flaws: real (or natural) and artificial. Real flaws are produced by replicating the flaw growth process of a naturally occurring flaw, or are produced by a controlled process that provides similar flaw morphology, signal response and therefore, flaw detectability to that of real flaws.

For demonstration specimens, fatigue cracks are accepted as representing the worst cracklike flaws from flaw detectability point of view. Therefore, demonstration of detectability of fatigue cracks covers all cracklike flaws. The fatigue cracks in demonstration specimens typically have ~0.00005" wide opening at the surface. These cracks are produced by using a fatigue crack growth process. The process uses an estimated maximum tensile stress of 50-70% of yield stress of the material. Minimum tensile stress in the fatigue crack growth process is less than or equal to 10% of maximum tensile stress. Reversed bending should be avoided during the fatigue crack growth process. A tensile pull or cantilever bending set-up is typically used. The cyclic frequency is ~20 Hz. If maximum load is more than the recommended percentage of yield stress, it may produce cracks with greater opening which may not bracket the worst cracklike flaws that need to be detected reliably. Cracklike flaws provide linear indications with length-to-width ratio of greater than or equal to 3. Lack of fusion (LF), hot or cold cracks are some of the examples of cracklike flaws. Primary set of flaws used in demonstration testing is assigned number 1 (i.e. Set 1) here.

For demonstration purposes, extreme of worst flaw detectability types are chosen. For example the fatigue crack specimens, described above, are considered to have worst flaw detectability for cracklike flaws in dye penetrant testing. These specimens define a flaw type domain as cracklike flaws with length and depth greater than or equal to that has been validated.

Consider an example of ultrasonic shear wave inspection for surface crack detection in thickness that varies between 0.1” to 0.25”. Here, the two extreme flaw types are: 1. flaw in 0.1” thick material and 2. flaw in 0.25” thick material. These two extreme flaw types define the flaw type domain for which the flaw detectability needs to be demonstrated.

Artificial flaws are used in place of real flaws, provided correlation (regression) of signal response with the real flaws or transfer function is established in appropriate simple geometry specimens such as plate or cylinder specimens. In this situation, the flaw detectability should be assessed using artificial flaw specimens that use real part geometry, material and surface finish. Electro-discharge machined (EDM) flaws, side drilled holes (SDHs), and flat bottom holes (FBHs) are examples of artificial flaws. If artificial flaws, in addition to the primary demonstration set are used for demonstration testing, the flaw set is called secondary set and is assigned number 2 (i.e. Set 2).

Two sets of flaws in simple geometry specimen are needed for establishing transfer function or correlation between signal responses of real and artificial flaws. One set is made from real flaws and is assigned number 3A (i.e. Set 3A). The second set is for artificial flaws with sizes approximately same as that in set 3A and flaws providing responses same as smallest real flaws to be detected .This set is assigned number 3B (i.e. (Set 3B). These two subsets together are assigned number 3. This set is not used for operator certification testing but is used in the NDE procedure qualification.

There are various kinds of NDE procedure standardization (or calibration) reference standards and inspection quality aids. NDE reference standards have simulated flaws in representative part geometry. NDE inspection quality aids are used to demonstrate certain level of inspection sensitivity. Some specimens are made from flat plates. Some specimens are made from sections of real part or from the entire part itself. These are called representative quality indicators (RQIs). It is difficult, if not impractical, to produce embedded real flaws of known size. Therefore, a part may be sectioned to generate a cutout specimen. Real or artificial flaws are introduced in the cutout specimen and the cutout is reassembled with the remaining part to make a representative part-like test specimen or an RQI. In x-ray radiography, inspection quality aids called image quality indicator (IQI) that simulate change in part thickness when placed over the part are used for indicating process quality by demonstrating thickness sensitivity and resolution. In magnetic particle testing quantitative quality indicator (QQI) shims are used to create a surface layer with cracklike flaw for demonstrating flaw detectability. There are other designs flaw detection sensitivity indicators tailored for the NDE application. Some of these examples are phantoms, line pair gages used in x-ray radiography. The calibration reference standards and inspection quality aids are assigned set number 4.

## 5.2 Dye Penetrant POD Specimens

Flaws in the POD specimens should have same or higher degree of difficulty in detection compared to that in NDE flaw detection application. In many situations, it is not practical to make specimens with flaws of concern with known size. For example, in order to qualify a dye penetrate procedure for detecting very small weld cracks, ideally the specimens should have very small weld cracks with known size. In reality, the POD specimens have fatigue cracks in wrought machined plates. The fatigue cracks are harder to detect than the weld and heat affected zone (HAZ) surface cracks and are considered to be an appropriate substitution for all surface cracklike flaws in welds. Normally, surface fatigue crack specimens are used to qualify dye penetrant, magnetic particle, x-ray, ultrasonic and eddy current procedures. However, other types of flaws may be appropriate depending upon the application.

Crack specimens used in dye penetrant qualification need special care. The specimens usually have a surface finish or roughness Ra of 63 µinch or better. The specimens are made from a titanium alloy or an alloy that is same or similar to the part that would be inspected using the qualified procedure. Specimen Cracks are produced by fatigue crack growth process. The process produces thumbnail cracks with fixed length-to-depth ratio. Length of the crack is monitored using a telescopic microscope while the specimen is in the fatigue loading set-up and under load. A number of cracks are destructively tested to assess the flaw length-to-depth ratio. These specimens are etched to a depth of ~0.0004 inch (for Titanium). The cracks are cleaned using solvent wipe and ultrasonic cleaning in ~140°F deionized water for half an hour and dried in hot air circulating oven at ~ 140°F. All crack lengths are accurately measured either in scanning electron microscope (SEM) or by using high magnification microscope (> 300X). The cracks are photo documented. The cracks have approximately 0.0002” wide opening at surface but are much tighter in gap between crack faces below surface. There should be no smeared metal across the crack opening. There should be no particulate material inside the crack opening. Each specimen is stored in a separate paper envelope.

The specimens are subjected to a written fluorescent dye penetrant procedure (penetrant sensitivity 3 or 4) and degree of difficulty in detection of the cracks is noted for each crack. The degree of difficulty is given as easy (1), moderate (2) and difficult (3). The procedure may be repeated using a different qualified operator up to two times. All cracks meeting above conditions and easy-to-moderate degree of difficulty qualify to belong in the POD demonstration set. 29 cracks within 10% of the target flaw size plus 5 larger (by ~20-30 %) cracks are chosen for demonstration of detectability.

The specimens are cleaned with approved solvents after each dye penetrant inspection followed by ultrasonic cleaning and drying. Test administrator is responsible for verifying quality of test specimens before providing the specimens for the demonstration test.

After about five inspection cycles, the specimens should be inspected under high magnification microscope or SEM and photo documented. After repeated use, at least 80% of length of each crack should be free of smear and particulate matter in the crack opening. The crack opening should be of the order of 0.0002" for the entire length. The specimens should not show any evidence of corrosion and surface damage such as scratches. The specimens are considered to be reusable when results of these quality checks are affirmative.

After every 15 dye penetrant procedures, in addition to above cleaning process and microscope inspection, the specimens should be tested using the same written dye penetrant procedure and the degree of difficulty should be noted and compared with earlier records degree of difficulty.

### 5.3 X-ray Crack Specimens

These specimens are manufactured using the same fatigue crack growth process used for the dye penetrant crack specimens except the surface is not etched. The surface should have a machined finish of Ra 63 µinch. Typically, 0.1" plate specimens have been used. For part thicknesses under 0.070" recommend a specimen thickness to be same or lower by 20%.

## 6.0 LIST OF NDE QUALIFICATION DATA ANALYSIS METHODS

a. See Appendix C and Appendix D for comparative information on data analysis methods.

### 6.1 POD Qualifications Covered in NASA-STD-5009

6.1.1 Signal Response versus Flaw Size or *a-hat* versus "*a*" Analysis (single hit)
   a. MIL-HDBK-1823 Maximum Likelihood Estimation (MLE) with General Linear Model (GLM)

6.1.2 Hit-miss Analysis
   a. Binomial Point Estimate (NASA preferred method)
   b. MIL-HDBK-1823 Maximum Likelihood Estimation (MLE) with General Linear Model (GLM)

### 6.2 POD Qualifications Not Covered in NASA-STD-5009

6.2.1 Signal Response versus Flaw Size or *a-hat* versus "*a*" Analysis (single hit)
   a. Physics Model Assisted POD (MAPOD)
   b. Curve Fit/Surface Fit POD. Example: Matlab curve fit

### 6.3 Custom NDE with Limited Validation (Not Covered in NASA-STD-5009)

6.3.1 Signal Response Analysis
   6.3.1.1 Single Hit Signal Response Analysis
   a. Uses minimum 6 flaws for each flaw detectability type (worst or extremes).
   b. Uses margin factor to manage risk and conditions on merit ratios to assess confidence qualitatively.
   c. Confidence on recommended merit ratios is validated by using Monte Carlo simulation of assumed model.
   d. Limited validation uses engineering analysis instead of POD.

▪

6.3.1.2 Multiple Hit Signal Response Analysis

a. Used for analyzing flaw indication data in 1D, 2D or 3D format.
b. Uses minimum 6 flaws for each flaw detectability type (worst or extremes).
c. Uses conditions on merit ratios and resolution to assess confidence.
d. Validated by modeling.
e. Limited validation uses engineering analysis instead of POD.

6.3.1.3 Transfer Function between Real and Artificial Flaws

a. Uses curve fit for transfer function.
b. Uses minimum 10 real flaws and minimum 10 artificial flaws for each flaw type domain.

6.3.2 Hit-miss Analysis

6.3.2.1 Single Hit Hit-miss Analysis

a. Uses minimum 16 flaws in a distribution of flaw sizes.
b. Confidence on recommended number of flaws is validated by using Monte Carlo simulation of assumed model.

### 6.3 Delta Qualification (Not covered in NASA-STD-5009)

a. May be used to fill gaps or shortcomings in the above qualifications or in justification of Standard NDE. Delta qualification may use limited validation.

## 7.0 CASES FOR VALIDATION OF NDE FLAW DETECTABILITY SIZE

a. Some NDE flaw detectability size validation cases are described below. Only NDE procedures that are capable of meeting MIL-HDBK-1823 POD requirements are covered by the following validation cases. Some cases may be hybrid of these cases and may not be covered here.
b. See Appendix B for a tabulated list of the validation cases provided below.

### 7.1 ST-1: Standard NDE, Delta Qualification, Single Hit, Hit-miss or Signal Response Data, Real Flaw

a. Delta qualification is used to tailor the qualification to assess similarity between the NDE procedure and Standard NDE procedure to verify the standard NDE flaw size.
b. Limited validation is not required for the delta qualification. Although, some of the limited validation conditions and approaches may be tailored for delta qualification.

### 7.2 SP-1: Special NDE, POD Qualification, Single Hit, Hit-miss or Signal Response Data, Real Flaws

a. Most ideal.
b. Typically uses fatigue crack flat plate specimens which likely differ from actual part geometry and it is assumed that POD results are applicable to inspection of actual hardware based on qualitative similarity analysis between the inspection procedure applied to real part and the inspection procedure used on the demonstration specimens.
c. Use MIL-HDBK-1823 or Binomial point estimate method.

### 7.3 SP-2: Special NDE, POD Qualification, Single Hit, Signal Response Data, Artificial Flaws

a. This validation is used if it is impractical to make real flaws of known size in quantity needed for POD but it is practical make artificial flaws of known size and quantity.
b. Usually artificial flaws are assumed to be as good as or worse than the real flaws and no transfer function between real and artificial flaws is needed. Example: POD qualification using programmed delaminations and disbonds in composites.
c. Use MIL-HDBK-1823 or Binomial point estimate method.

## 7.4 CU-1A: Custom NDE, Hybrid of POD and Limited Validation, Single Hit, Signal Response Data, Real and Artificial Flaws

a. NDE POD validation on real flaws in simple geometry specimens and limited validation on artificial flaws in part geometry specimens.
b. This validation is used if it is impractical to make real flaws of known size and in a quantity needed for POD in specimens with part geometry and material. A signal response transfer function between real and artificial flaws is necessary. Assumes that artificial flaws can be fabricated in part geometry specimens.
c. Use MIL-HDBK-1823 or Binomial point estimate method as a guideline for POD on simple geometry specimens. Use signal response transfer function correlation to determine equivalent artificial flaw size. Use limited validation requirements for artificial flaws. Compare the limited validation artificial flaw data and POD data using merit ratios. Comparable merit ratios between the real and equivalent artificial flaws and ratios meeting minimum requirements indicate that the validation is successful for respective flaw sizes. Choose most conservative (i.e. larger) of the two flaw sizes. If the conservative flaw size is from the validation of artificial flaw detection, then use the real flaw equivalent of the artificial flaw as the qualified flaw size.
d. This qualification is considered Custom NDE with limited validation even though POD level of testing is performed. The limited validation testing results may indicate that POD validated flaw size is qualified and NDE procedure classification will be equivalent to Special NDE. If the limited validation testing results do not support the POD flaw size, then the classification will be Custom NDE.
e. See Appendix E for more information.

## 7.5 CU-1B: Custom NDE, Limited Validation, Single Hit, Signal Response Data, Real Flaws

a. This validation can be used if it is impractical to make real flaws of known size in a quantity needed for POD analysis. It uses limited number of real flaws in real part geometry and materials. Artificial flaws are used for calibration.
b. If merit ratios meet the minimum requirements then the validation is successful.
c. See Appendix E for more information.

## 7.6 CU-1C: Custom NDE, Limited Validation, Single Hit, Signal Response Data, Real and Artificial Flaws

a. This validation is used if it is impractical to make real flaws of known size in a quantity needed for POD analysis.
b. Uses transfer function between signal responses of real and artificial flaws in representative simple geometry specimens.
c. Artificial flaws in part geometry and material specimens are used in validation testing.
d. Some artificial flaws are used in calibration.
e. If merit ratios meet the minimum requirements then the validation is successful.
f. See Appendix E for more information.

## 7.7 CU-1D: Custom NDE, Limited Validation, Multiple Hit, Signal Response Data, Real Flaws

g. This validation can be used if it is impractical to make real flaws of known size in a quantity needed for POD analysis. It uses limited number of real flaws in real part geometry and materials. Artificial flaws are used for calibration.
h. If resolution and merit ratios meet the minimum requirements then the validation is successful.
i. See Appendix E for more information.

## 7.8 CU-1E: Custom NDE, Limited Validation, Multiple Hit, Signal Response Data, Artificial Flaws

a. This validation is used if it is impractical to make real flaws of known size in a quantity needed for POD analysis.
b. Uses transfer function between signal responses of real and artificial flaws in representative simple geometry specimens.
c. Artificial flaws in part geometry and material specimens are used in validation testing.

d. Some artificial flaws are used in calibration.
e. If resolution and merit ratios meet the minimum requirements then the validation is successful.
f. See Appendix E for more information.

### 7.9 CU-2: Custom NDE, Limited Validation, Single Hit, Hit-miss, Real Flaws

a. This validation is used if it is impractical to make real flaws of known size in quantity needed for POD but it is practical make smaller number of real flaws of known size and quantity.
b. See Appendix F for more information.

### 7.10 CU-3: Custom NDE, Process Control Validation, Single Hit or Multiple Hit, Hit-miss or Signal Response, Real or Artificial Flaws

a. If NDE flaw detectability size cannot determined by POD or engineering analysis and NDE technique meets industry standard practices
b. Usually the technique sensitivity implies a flaw detectability size but confidence in the flaw detectability size is usually not quantified.
c. Should use limited validations principles.

## 8.0 GENERAL GUIDELINES FOR VALIDATING NDE FLAW SIZE

In order to develop an NDE procedure validation plan, first the target flaw size (which is usually the initial critical flaw size), its location (IML surface, OML surface, embedded, edge, corner, fillet), orientation, type (cracklike, void, disbond, IP), needs to be determined. Depending upon above factors, many flaw detectability sizes may be expected. Therefore, there may be many flaw detectability types. However, these flaw detectability types may be combined using similarity as flaw detectability groups so that flaw detectability within a group or domain is expected to be monotonic with the major variable of the group. Consider an example of flaw detection where circumferential cracklike flaws for a part with thickness ranging from 0.1” to 0.25” are to be detected using ultrasonic shear wave testing. These flaw types of varying part thickness can be grouped with group or domain ranging from low thickness to high thickness. The flaw detectability domain can be represented by two (or more) extreme flaw detectability types i.e. one for lowest thickness and one for highest thickness in the validation demonstration test.

All flaw type domains and their representative extreme and/or worse cases of flaw detectability types should be identified. Each (extreme or worst) flaw detectability type should be validated though demonstration testing using NDE procedure that would be used in inspection of actual part. NDE procedure shall be optimized and qualified based on test data generated by the procedure developer prior to the operator certification testing. Some trial-and-error work may be needed to optimize the NDE procedure to detect the target size flaw before flaw sizes for flaw sets are chosen.

According to NASA-STD-5009, each flaw detectability type should have a complete POD demonstration. For Binomial point estimate POD demonstration, minimum 34 flaws are needed for each flaw detectability type identified for the demonstration. For MIL-HDBK-1823 *a-hat* versus “*a*” analysis, minimum 40 flaws are needed and for MIL-HDBK-1823 hit-miss analysis minimum 60 flaws are needed.

If limited validation with signal response analysis is used, minimum 6 flaws are needed for each flaw detectability type, but minimum 34 flaws are needed to demonstrate operator skill, procedure repeatability and acceptable POF. If the NDE procedure is required to detect six or more extreme or worst flaw detectability types, then the validation will naturally have more than 34 flaws. 34 flaws with required number of inspection opportunities would be sufficient for demonstrating operator skill, procedure repeatability and acceptable POF.

If limited validation with hit-miss analysis is used, minimum 16 flaws are recommended for each extreme or worst flaw detectability type but minimum 34 flaws are needed to demonstrate operator skill, procedure repeatability and acceptable POF. If the NDE procedure is required to detect 2 or more flaw detectability types, then the validation will naturally have 32 flaws needing two additional 2 flaws. 34 flaws with required number of inspection opportunities would be sufficient for demonstrating operator skill, procedure repeatability and acceptable POF based on NASA practice of Binomial point estimate demonstration used since early 1980's.

All analyses covered in this guide assume that scatter in the signal response data, hit-miss data; and noise is more or less uniform with flaw size. Analyses assume standard distribution for the scatter. Data used in flaw size analysis is carefully chosen.

It is assumed that, for MIL-HDBK-1823 POD and limited validation, the demonstration data used is more or less evenly distributed in size. The target flaw size is in the center 40% of the flaw size range used for hit-miss type of data. For signal response analysis, flaw size range should omit saturated values on upper end and omit low signal values close to noise on lower end. This data filtering should be based on the chosen signal response threshold levels, i.e. upper signal response threshold and lower signal response threshold. Demonstration data should also omit outliers. For curve fit methods, it is advised to use optimal range of flaw size such that the signal response increases monotonically with flaw size within the chosen range. A minimum correlation coefficient $R^2$ of 0.8 is recommended.

Only Binomial point estimate method, MIL-HDBK-1823 *a-hat* versus *"a"* and hit-miss analysis methods are covered by NASA-STD-5009. All other analysis methods including POD and non-POD methods that use transfer function, curve-fit POD method, physics model based fit POD method, and Custom NDE with limited validation methods are not covered by NASA-STD-5009 and should be authorized by the engineering authority.

NDE process qualification has two phases. The first phase is for NDE procedure development and optimization. This phase is done by the NDE procedure developer who is normally an NDE professional, preferably degreed engineer with knowledge and experience in the NDE method. During development, the developer develops the NDE procedure, and optimizes the procedure and provides process control requirements for the NDE procedure i.e. technique calibration and quality indicator requirements, x-ray cone angle requirements, CNR, TNR, CTR requirements etc. These requirements should be adequate to control the NDE procedure to provide repeatability in flaw detectability by certified operators. The developer chooses an appropriate NDE qualification approach and prepares the qualification plan. The developer designs the flaw sets for the procedure development, and optimization as well as designs and validates flaw sets for operator certification testing. With the help of operators, if needed, the developer checks out the NDE procedure on the certification demonstration set in a blind manner. This testing should be similar to phase 2 testing, except depending on results, a revision to the development testing and procedure may be needed, Once the development testing is completed, a report is prepared with information on development objective, plan, specimens, testing results, conclusions, written procedure and information on approach for operator certification testing, other information needed in certification testing such as flaw maps, blank data forms and expected results. At the conclusion of development phase, the NDE procedure should have inherent flaw detectability better than the target flaw size so that a trained NAS410 level 2 operator has a reasonable chance of passing the demonstration test. The development phase is done once for a given procedure. If any changes are made in equipment, materials, set-up or part inspected, these changes should be evaluated for impact on flaw detectability of the procedure. If the impact is likely to be adverse on flaw detectability, as assessed by responsible NDE professional, a delta qualification or requalification may be needed. If phase 1 testing is successful indicating positive results in terms of flaw detection and corresponding process control conditions, the NDE procedure has potential to be qualified.

In second phase, operator certification testing is conducted for Special NDE and for limited validation. The testing assumes that the NDE procedure has been developed and checked out to provide adequate process control and flaw detectability so that there is a reasonable chance for the operator to pass the demonstration test. If a POD flaw size is assigned, then there shall be sufficient number of flaws of the required flaw detectability type in Binomial point estimate method. Similarly, there shall be sufficient number of flaws in a recommended flaw size range for MIL-HDBK-1823 POD method. See Section 9.0. At the successful conclusion of second phase, the Special NDE or Custom NDE with limited validation is qualified. The qualification is specific to procedure, NDE equipment and operator.

Application of qualified NDE procedure requires similarity analysis i.e. assessing similarity between the qualified NDE procedure and the actual inspection procedure for which the validated flaw size is assumed. Such analysis may be based upon evaluating similarity between the part under inspection and specimens used in validation for material type, alloy, surface finish, geometry, material thickness, NDE technique, operator access to inspection area etc. The similarity analysis may include comparison of signal responses on same thickness/geometry sections and measurement of noise. If responsible NDE level 3 evaluates that the part inspection is dissimilar to the qualified NDE procedure, then either a delta qualification to the current NDE procedure qualification or a new NDE procedure qualification by demonstration may be needed.

# 9.0 SELECTED POD VALIDATION DATA ANALYSIS METHODS

See Appendix C and Appendix D for comparative information on data analysis methods.

## 9.1 Binomial Point Estimate Hit-Miss POD Method

NASA JSC predominantly uses Binomial point estimate hit-miss POD method. The Binomial statistics allows use of a smaller number of flaws compared to the MIL-HDBK-1823 methods. The minimum number of flaws needed is 29 for the target size flaw. In addition, 5 larger (by 20-30%) flaws are needed to demonstrate that flaw detection becomes easier with larger flaw size. Therefore, minimum 34 total flaws per flaw detectability type are needed in Binomial point estimate method. The analysis, however does not calculate the actual $a_{90/95}$, but rather a flaw size $a_{90/95\_point\_estimate} > a_{90/95}$. The Binomial point estimate flaw size has POD of 90% and confidence of minimum 95% and may be denoted as $a_{90/95\_min}$. Flaw detectability size is a characteristic of a combined effect of NDE procedure (including equipment), NDE operator and part inspected. The demonstration chooses a flaw size that is greater than the unknown $a_{90/95}$ by trial-and-error. If the target flaw size is smaller or close to $a_{90/95}$, the demonstration is highly likely to fail. The point estimate flaw size is always larger than the MIL-HDBK-1823 $a_{90/95}$. The 29 flaws used are of the target $a_{90/95\_target}$ size or smaller with a small tolerance on the flaw size. The tolerance is typically about ±10%. Although, it is not necessary that the tolerance should be smaller than 10%. Larger tolerance is likely to reduce chances of passing the demonstration. If a procedure can detect a flaw with size $a$ with POD of 90%, and there are 29 successful detections out of 29 trials, then the confidence can be calculated using the Binomial statistics as follows.

Confidence = $1 - (0.90)^{29} = 95\%$.

The qualified flaw size is average of the 29 flaws detected. No larger flaws should be missed. Flaws smaller than the smallest in the 29 flaw set are allowed to be missed. Instead of 29 flaws, demonstration can use 46 flaws. Minimum 45 out of 46 flaws must be detected to pass the demonstration. No larger flaws shall be missed. Flaws that are smaller than the smallest in the 46 flaw set are allowed to be missed. In this case, the qualified flaw size is average of the 46 flaws assuming only one flaw is missed. Else if all 46 flaws are detected, calculate average of the smallest 29 flaws detected with same conditions, i.e. no larger flaws are missed. 0.1% POF is recommended. MIL-HDBK-1823 recommends that the demonstration has minimum 1000 false call opportunities (inspection opportunity in unflawed locations) for verifying POF of less than 0.1% and minimum 100 false call opportunities for verifying POF of less than 1%. The handbook also recommends that there should be 3 times the number of unflawed locations. Unflawed locations are needed to avoid guessing and to calculate the POF. Unflawed locations need not be only in unflawed specimens. One can use Clopper-Pearson Binomial distribution and determine the unflawed inspection opportunities needed to calculate the POF. Inspection opportunity is defined loosely as an inspection area that the operator can examine at a time without manipulating part or changing line-of-sight. Therefore, if flaw can exist in any location of the specimen, multiple inspection opportunities exist on the same demonstration specimen. Inspection opportunity area (for example 1" x 1") can be defined. Inspection area of all demonstration specimens can be measured. Number of inspection opportunities is calculated as available inspection area divided by the single inspection opportunity area. The inspection opportunities are estimated so that the demonstration can claim existence of these inspection opportunities. Essentially the demonstration does not allow any unexplained false calls. There should be minimum 3 blank specimens to provide necessary unflawed and flawed inspection opportunities and prevent guessing.

## 9.2 *a-hat* versus *"a"* MIL-HDBK-1823 POD Methods

Typically minimum 40 flaws are recommended for each flaw detectability type. Flaws sizes shall be evenly distributed. The smallest flaws must give net signal-to-noise ratio greater than 2.5. The largest flaw response shall not have saturated signal response. Condition on the range are given by following equations and inequalities,

$$C_{rel} = \hat{a}_{50} - \hat{a}_{decision}. \tag{1}$$

$$2.5\sigma < \hat{a}_{50} - 3C_{rel} < \hat{a}_{range} < \hat{a}_{50} + 3C_{rel} < \hat{a}_{suaturated} \tag{2}$$

where,

$a_{range}$ = range of flaw sizes for demonstration flaws,

$\hat{a}_{range}$ = range of response range,

$C_{rel}$ = relative contrast,

$\hat{a}_{50}$ = mean signal response from target flaw,
$\hat{a}_{decision}$ = decision threshold, and
$\sigma$ = standard deviation of noise.

User may follow MIL-HDBK-1823 guidelines on flaw distribution and analysis. Minimum 40 noise measurement data points should be taken. A minimum correlation coefficient $R^2$ of 0.8 is recommended.

### 9.3 Curve Fit and Physics Model Based POD Methods (NOT Covered in MIL-HDBK-1823 and NASA-STD-5009)

Same requirements as above are applicable. Ref. [8] and [11] provide some information on curve fit method. Curve fit methods may also include transfer function. Physics based model may be used as a curve fit equation.

### 9.4 Hybrid of POD and Limited Validtion (NOT Covered in MIL-HDBK-1823 and NASA-STD-5009)

The hybrid approach uses POD demonstration using real flaws in specimens that have simpler geometry than actual part. For example, using cracks in flat plates to demonstrate ultrasonic shear wave technique on a pressure vessel wall with spherical surface can be a hybrid approach. Here, we get a POD flaw size for the NDE technique for the flat plate geometry. But the real part is not flat and effect of part geometry needs to be verified. Therefore, in addition to the simple geometry POD demonstration, part geometry limited validation on equivalent artificial flaws is under taken. Here, requirements of POD, transfer function and limited validation apply.

### 9.5 Hit-Miss MIL-HDBK-1823 POD Method

Minimum 60 flaws are recommended for each flaw detectability type. Flaws sizes shall be evenly distributed. Approximately, 40-45% of data points shall span from smallest detected flaw to largest missed flaw. The remaining data points shall be equally distributed to the left of or smaller than smallest detected flaw and to the right of or larger than largest missed flaw. This assumes that the flaw sizes are sorted from smallest i.e. from left side to largest i.e. to right side. User may follow the MIL-HDBK-1823 guidelines on flaw distribution and analysis. Minimum 1000 unflawed inspection opportunities are recommended to demonstrate POF of less than 0.1%. Minimum 100 unflawed inspection opportunities are recommended to demonstrate POF of less than 1%.

## 10.0 CUSTOM NDE WITH LIMITED VALIDATION DATA ANALYSIS METHODS

See Appendix C and Appendix D for comparative information on data analysis methods.

### 10.1 Signal Response versus Flaw Size Dataset[9,10,11]

Limited validation uses 3 flaws of target size and 3 flaws of sub-target size for each flaw type for demonstration. These could be real flaws or artificial flaws. Each demonstration shall have more than 34 flaws to verify procedure repeatability and operator skill.

In order to pass the demonstration, each demonstration flaw shall meet the merit ratio conditions. Thus, use of less number of flaws is compensated by more stringent merit ratio requirements on all demonstration flaws. If merit ratio requirements cannot be met, a larger target flaw size can be chosen and the validation testing exercise can be repeated until the ratio conditions are met.

### 10.2 Custom NDE with Limited Validation: Hit-miss versus Flaw Size Dataset[8]

When devising this NDE procedure validation it is important to choose the flaw size range correctly as given in Ref. [8].

### 10.3 Comparative Information on Validation Cases, Validation Approchess

See Appendix A for general comparison of NDE procedure qualifications.
See Appendix B Table for NDE procedure qualification cases.
See Appendix C for comparison of NDE data analysis methods by data type.
See Appendix D for comparison of data analysis methods by the analysis methods.
These tables are provided to help in choosing the right validation method for the NDE procedure considered for validation.

## 11.0 DEMONSTRATION ADMINISTRATOR

Demonstration administrator has been tasked to maintain integrity of the operator certification demonstration test. Many duties have been assigned to the administrator. The administrator does following duties.

a. Ensures that the flaw specimens to be used for demonstration are clean, dry, have required quality and are appropriate for the test.
b. Maintains control over flaw maps of the specimens and does not share them with operators or their agents preventing guessing and cheating.
c. Provides the blank flaw maps to the operator for recording the location of indications.
d. Provides oversight during the demonstration so that the operator cannot undermine demonstration test.
e. Performs quality audit of inspection facility that includes checking calibration on equipment, checks level II or III NDE certification records and obtains copy of certification and written procedures.
f. Provides oversight to determine that the written procedure is followed.
g. Grades the demonstration results.
h. Provides debriefing on the observations of demonstration and process quality checks to the operator and his supervisor.
i. Prepares a memo addressed to NASA cognizant engineer with conclusion of the demonstration and recommendation for or against the certification the NDE procedure and operator.
j. Maintains all records of the demonstration is a separate folder for a minimum 5 years.

## 12.0 DEMONSTRATION

### 12.1 Dye Penetrant Demonstration Guidelines for Qualifying Crack Size of 0.050” or Longer

Flat fatigue crack specimens used in this demonstrations typically have dimensions of 4” x 6” x 0.2”. Specimen surface finish is Ra 63 µinch or better. Normally a set of 29 cracks within 10% of the target flaw size are chosen for primary trials. In order to select the 29 flaws of the primary set, all flaw sizes in the demonstration set should be sorted in increasing order of size. A block of contiguous 29 flaws are located so that the average flaw size is smaller than or equal to the target flaw size. These flaws can be used as the main demonstration cracks.

All main demonstration cracks, 29 in all, must be detected to pass the demonstration. The specimen containing the main demonstration cracks may have smaller and larger cracks. All larger cracks are also considered to be part of the demonstration and must also be detected to pass the demonstration. The smaller cracks are not omitted from the demonstration. Non-detection of the smaller cracks does not affect the demonstration of capability at the target flaw size. Typically 10-20 specimens may be part of the demonstration with 60% flaws in the main flaw set and 25% larger than the largest in the main flaw set and 15% smaller than the smallest in the main flaw set.

Minimum 3 blank specimens should be included in the dye penetrant demonstration. Operator taking the demonstration test, should not have prior knowledge of flaw locations, indications or absence of the same. A set of practice specimens can be used prior to the demonstration test for training purposes. The operator may practice the NDE procedure on the training specimens. A successful demonstration would require detection of all main flaws and all larger flaws. In addition, there should be no unexplained false calls. Alternately, the reliably detectable flaw size, $a_{90/95\_min}$ can be calculated as average of the contiguous block of smallest 29 flaws that have been detected, provided there are no missed flaws of higher size.

The demonstration qualifies dye penetrant procedure for detection of surface cracks in most metals. A minimum surface finish of Ra 63 µinch is assumed. An unobstructed access to the part surface, so that penetrant process steps of cleaning, drying, penetrant application, excess penetrant removal or washing, developer application and indication evaluation under black light are not obstructed or impeded; and can be performed with direct line-of-sight angle greater than or equal to 45 degree with part surface.

### 12.2 Dye Penetrant Demonstration Guidelines for Qualifying Crack Size of Smaller Than 0.050”

This penetrant demonstration must account for part geometry, part material, surface finish and inspection rate. The demonstration specimens should be made from a similar alloy and with a similar or worse surface finish. Due to smaller flaw size, the flaw detectability is more sensitive to processing access to part surface compared to that of 0.050” flaw size demonstration. Therefore, the penetrant process application is assessed for above factors first. The demonstration should simulate the access conditions of the dye penetrant application. The demonstration should simulate the worst part geometry for dye penetrant inspection. In inspecting a metallic cylinder, the interior surface is likely to be more difficult to inspect compared to the external surface. Therefore, the worst penetrant inspection geometry is the interior surface. The flat specimens should be arranged inside a cylinder with diameter and length same as the part to be inspected. In addition, all conditions of the demonstration for target flaw size greater than or equal to 0.050 inch apply.

The inspection rate is defined as the area that can be inspected in a single run. It can be assessed by demonstrating the dye penetrant procedure on the actual part and on the specimens that are arranged in the geometric configuration covering about the same area. Typically, the indications are evaluated between 10 to 60 minutes after application of the developer. The indication evaluation period limits the area that can be inspected in one run of the procedure so that sufficient time is provided to evaluate the developed surface.

### 12.3 X-ray Radiography Testing Demonstration[13]

The demonstration should use the same x-ray technique that would be used for actual part inspection. If double wall x-ray is used on the real part, then the demonstration should use double wall x-ray including any other materials such as the composite overwrap in COPV. If demonstration test is intended to qualify detection of cracklike flaws, then the cracks in the specimens should be located at the extreme x-ray angle to be qualified. X-ray angle with the normal to specimen surface should be accurately measured. Use of laser beam alignment devices, alignment levels and tape measure are recommended to achieve the desired x-ray angle with surface normal to the part. Follow guidelines in Ref [13] in Appendix G for validation of x-ray techniques.

## 13.0 PROCESS QUALITY SURVEY

ASTM process quality checks should be verified at the time of operator certification demonstration test by the administrator. Operator certification of NAS410[5] level 2 or higher in the relevant method should be verified at the time of demonstration test.

## 14.0 QUALIFICATION REPORT AND LETTER

### 14.1 Qualification Report

Qualification report should include documentation on:

a. Inspection objective: reject criteria, associated $a_{90/95}$ size, if any
b. Part: geometry, flaw types (worst or extremes), flaw locations, flaw orientation, flaw detection size requirements
c. Technique: description (i.e. ultrasonic pulse echo, contact, immersion)
d. Calibration and process quality standards including calibration reference standards, IQI, RQI, Phantoms, gages: flaw/reflector types, flaw location, flaw orientation, flaw sizes and tracking information
e. Calibration: procedure, reject level, noise, real flaw to calibration flaw relationship/correlation/transfer function, reject level or decision threshold-to-noise ratio
f. Transducer, probe, camera and equipment: characteristics, focusing, frequency, aperture, wedge, make, model, version etc.
g. Scan plan/technique: directions, index, data processing/analysis, data scan formats, scan plan
h. Written procedure: need to assess if the procedure is adequate to its stated objective
i. Statement regarding verification of ASTM process quality checks

j. Operator certification: level 2 certification in NDE method,
k. Validation testing report
    1. Validation testing specimen information: Description, flaw types, sizes, orientation, location, quantity, blanks
    2. Procedure optimization and qualification testing data. This is generated only once for a given procedure.
    3. Operator certification demonstration test results: Account of demonstration test (operator names, administrator name, facility, dates) and results (flaws detected, flaws missed, false calls),
    4. Analysis and conclusions: Analysis of POD, POF, limited validation or tailored validation, analysis of operator and procedure repeatability and conclusions regarding success of the validation testing and flaw size qualified. Any restrictions or constrains on procedure, personnel, equipment and facility in applicability of the validation
    5. Validation test report shall be signed and dated by the author.
    6. Qualification letter and certificate.

## 14.2 Certification Letter and Certificate

A qualification letter is issued by the NASA cognizant engineer based on recommendation by the demonstration administrator. The qualification letter should state,
a. Name of the NDE operator certified,
b. NDE procedure name and number with revision,
c. Date of the demonstration test and effectivity of the certification (expires usually 3 years from the date of demonstration test),
d. Any restrictions or constraints on the procedure, personnel, equipment and facility,
e. State the $a_{90/95}$ flaw size as having 90% POD with minimum 95% confidence, or
f. State $a_{lv}$ flaw size as having high confidence that the flaw size exceeds the flaw size with 90% POD with 95% confidence but data is insufficient for POD calculation.
g. If a wall certificate is issued, the certificate shall mention the certification letter number on the certificate, name of the NDE operator certified, NDE procedure name and number with revision.
h. The certificate shall state the flaw size and validation type Special (POD) or Custom (limited validation) and effectivity of the certification, name and signature of authorizing person.

## 15.0 REFERENCES

[1] MIL-HDBK-1823A, “Nondestructive Evaluation System Reliability Assessment,” Department of Defense, USA, (2009).

[2] Annis, Charles, “mh1823 POD Software V5.2.1,” Statistical Engineering, http://www.statisticalengineering.com/mh1823/, (2016).

[3] ASTM E 2862-12, “Standard Practice for Probability of Detection for Hit/Miss Data,” ASTM International, (2012).

[4] Rummel Ward D., “Recommended Practice for Demonstration of Nondestructive Evaluation (NDE) Reliability on Aircraft Production Parts,” Materials Evaluation, Issue 40 pp 922, (August 1988).

[5] NAS410, “NAS CERTIFICATION & QUALIFICATION OF NONDESTRUCTIVE TEST PERSONNEL.”

[6] NASA-STD-5009, “NONDESTRUCTIVE EVALUATION REQUIREMENTS FOR FRACTURE-CRITICAL METALLIC COMPONENTS.”

[7] Koshti, A. M., Optimizing probability of detection point estimate demonstration”, Nondestructive Characterization and Monitoring of Advanced Materials, Aerospace, and Civil Infrastructure, Proc. of SPIE Vol. 10169, 2017.

[8] Koshti, Ajay. M., “NDE flaw estimation using smaller number of hit-miss data-points”, NASA L. B. Johnson Space Center, (DAA TN56089 Approved May 9, 2018, Unpublished 2018)

[9] Koshti, Ajay. M., “Assessing reliability of NDE flaw detection using smaller number of demonstration data points”, NASA L. B. Johnson Space Center, (DAA TN56083 Approved May 9, 2018, Unpublished 2018)

[10] Koshti, Ajay. M., “NDE flaw detectability validation using smaller number of signal response data-points”, NASA L. B. Johnson Space Center, (DAA TN56087 Approved May 9, 2018, Unpublished 2018)

[11] Koshti, Ajay. M., “Reliably detectable flaw size for NDE methods that use calibration”, Nondestructive Characterization and Monitoring of Advanced Materials, Aerospace, and Civil Infrastructure 2017, Proc. of SPIE Vol. 10169, (March 2017).

[12] Koshti, Ajay. M., “Eddy current crack detection capability assessment approach using crack specimens with differing electrical conductivity,” Proc. SPIE, 10599, Nondestructive Characterization and Monitoring of Advanced Materials, Aerospace, Civil Infrastructure, and Transportation XII, 105991M, (March 2018)

[13] Koshti, Ajay. M., Memo to Brian Mayeaux, ES4-17-065, “Recommended X-ray Nondestructive Evaluation Requirements”, NASA L. B. Johnson Space Center, (Dec., 8, 2017).

# 16.0 DEFINITIONS AND ACRONYMS

## 16.1 Definitions

| | |
|---|---|
| Artificial Flaw | A manufactured flaw that is used to simulate a real flaw for NDE qualification testing, calibration or quality aids. EDM notch and flat bottom holes are examples of artificial flaws. |
| Binomial Point Estimate Analysis | Uses Binomial distribution for calculating confidence associated with POD. The analysis may include Clopper–Pearson Binomial method given in ASTM E2862. Analysis provides POD and POF. |
| Calibration | Refers to process of instrument standardization using NDE reference standard to set the technique sensitivity and decision threshold in signal response NDE methods. |
| Certification | Process of qualifying personnel resulting in a letter or certificate of completion of the qualification to stated objectives. For qualification of NDE procedure, the certification should be specific to written procedure, operator, equipment and facility. |
| Curve/Surface Fit POD Analysis | The analysis uses signal response versus flaw size data (one or two parameters) and maps the correlation as either a curve or a surface and determines 90/95 POD/Conf. Analysis provides POD and POF. |
| Custom NDE | NDE other than NASA-STD-5009 Standard or Special NDE is termed as Custom NDE. Custom NDE does not have a POD $a_{90/95}$ flaw size associated with it. It may or may not have limited validation flaw size, $a_{lv}$ associated with it. |
| Demonstration Test | A controlled blind test where operator taking the test does not have prior knowledge of location of the flaws within test specimens. A demonstration test is designed such that guessing the flaw calls is likely to fail the demonstration. The demonstration test uses a written NDE procedure where process controls are used to ensure quality and repeatability of the procedure. Outcome of the test forms the basis for qualification of the NDE procedure as well as certification of the NDE operator. |
| Demonstration Administrator | NDE professional tasked with administering the operator certification demonstration test and is entrusted to ensure blind nature of the test. |
| Discontinuity | An intentional or unintentional interruption in the physical structure or configuration of a material or component that may be detectable by nondestructive testing. Discontinuities are not necessarily rejectable. |
| Engineering Authority | The term is used in a limited scope here and it pertains to a board or office responsible to approve or disapprove deviations or variances to contractual requirements affecting NDE implementation. |
| Equivalent Artificial Flaws | An artificial flaw that provides same average signal response as the real flaw. For example, in ultrasonic shear wave inspection technique, an EDM surface notch can be an equivalent artificial flaw for a surface crack of certain size. |
| Flaw | Same as discontinuity. |
| Flaw Detectability | Same as flaw detection sensitivity. |

| | |
|---|---|
| Flaw Detection Sensitivity | Flaw detection sensitivity refers to smallest flaw detectability size. The term may be used qualitatively for comparing NDE procedures. |
| Flaw Detectability Size | A flaw size that NDE procedure is capable of detecting or expected to detect reliably. It is assumed that the procedure can also detect flaws with size larger than the flaw detectability size reliably. |
| Flaw Type | NDE description of flaw i.e. crack, void, pore. |
| Flaw Kind | Real and artificial are the two kinds of flaws. |
| Flaw Detectability Type | A set of flaw descriptors, including flaw type (i.e. crack, delamination, pore), location (i.e. surface, embedded, ID, OD), orientation (i.e. longitudinal, circumferential), clocking with respect to part geometry, angle with a geometrical feature, and local part thickness etc. define a flaw detectability type such that any change in any of the above factors may result in change in flaw detectability. If flaw detectability size is needed, it shall be demonstrated for extreme or worst cases of each flaw detectability type. Flaw size range used in MIL-HDBK-1823 is established for a single flaw detectability type. |
| Flaw Type Domain | A group of similar flaw detectability types so that flaw detectability within the group (envelope or range) is assumed to be monotonic with the major variable of the group other than flaw size i.e. local section thickness. For example, circumferential OD surface flaw from thickness ranging from 0.1" to 0.25" may belong to same flaw detectability domain for ultrasonic shear wave inspection, although POD flaw size may be different for flaws in 0.1" and 0.25" material thickness. |
| Flaw Type Equivalency | A term used to describe basis for use of artificial flaw in place of a real flaw in NDE procedure qualification. The equivalency may be based on signal response transfer function data obtained by comparing signal responses from one type of real flaws with the same type of artificial flaws. |
| Fracture Critical | Part classification which assumes that fracture or failure of the part resulting from the occurrence of a crack will result in a catastrophic hazard. Fracture critical components will be identified as such on the engineering drawing. Refer to NASA-STD-5019 for more details and related part classifications. |
| Hit-miss Data | There are four kinds of data. 1. True positive or hit data, i.e. flaw is detected where there is flaw. 2. False negative or miss data, i.e. there is a no-flaw call where there is flaw. A hit is assigned a value of 1 and a miss is assigned a value of 0 in POD and limited validation analysis. 3. True negative data i.e. there is no-flaw call when there is no flaw. 4. False positive data i.e. flaw call when there is no flaw. True negatives (unflawed), and hits plus misses (flawed) together constitute total inspection opportunities. False positive calls are counted separately and analyzed for POF as a proportion of unflawed inspection opportunities. Hit-miss data and false positive data are analyzed separately. |
| Hit-miss POD Analysis | This analysis is based on hit-miss data using MIL-HDBK-1823 or Binomial point estimate analysis providing POD and POF. |
| Indication | Evidence of a discontinuity that requires interpretation to determine its significance. |
| Inspection | Same as nondestructive evaluation or testing performed for part acceptance. |
| Inspection Sensitivity | A term used to indicate a quality level of inspection with implied flaw detection sensitivity. Same as NDE technique sensitivity. |
| Limited Validation | NDE procedure validation that provides a flaw detectability size, $a_{\mathrm{lv}}$ with high confidence that $a_{\mathrm{lv}} > a_{90/95}$ and has low POF such as < 0.1% or other approved value. Limited validation does not provide $a_{90/95}$ flaw size and uses limited number of flaws for demonstration of flaw detectability. Limited validation methods, although based on POD models, are not POD methods and are intended to provide evidence that there is a high confidence in $a_{\mathrm{lv}} > a_{90/95}$. |

| | |
|---|---|
| Merit Ratios | CNR, TNR and CTR are the three merit ratios. |
| Multiple Hit | When a single flaw is detected multiple times due to signal or contrast coming from different part of the flaw. For example, if a radiography technique has a demonstrated (i.e. real) resolution of 0.030” and a pore of diameter 0.1” is detected., then the x-ray image of the pore has multiple real resolution size pixels and is considered to be a case of multiple hit. A cluster i.e. 1D, 2D or 3D array of single hits on a single discontinuity constitutes multiple hits. |
| NDE Classification | Standard, Special and Custom are the three classes of NDE procedures. |
| NDE Criteria for Reliable Flaw Detection | POD/Confidence $\geq$ 90/95 for the targeted flaw size with POF $\leq$ 0.1% (or other approved value i.e. 1%). FCB or engineering authority may approve use of Custom NDE with limited validation in place of POD validation. |
| NDE Method | NDE testing fields such as radiographic testing, ultrasonic testing, eddy current testing are called NDE methods. Typically NAS 410 certification of personnel is provided specific to a method. |
| NDE Procedure | NDE process from start to finish. Documented NDE procedure is also called a written procedure. |
| NDE Procedure Developer | An NDE professional or engineer responsible for developing and implementing NDE procedure. |
| NDE Procedure Qualification | Is a process of verifying flaw detectability or technique sensitivity by demonstration of NDE procedure by detecting a set of real and/or artificial flaws or features. A qualified procedure may have flaw detectability given as validated flaw size based on assessment of similarity between the procedures used for validation testing and part inspection. |
| NDE Reference Standards | Physical test pieces that have known size simulated flaws in representative part geometry and are used in technique calibration and verification of technique sensitivity. |
| NDE Technique | NDE technique captures necessary conditions of the NDE procedure such that when the technique is repeated at different times or by a different qualified operators, results are repeatable. |
| NDE Set-up | Same as NDE technique. |
| Noise | Noise is measured either as peak values or as standard deviation, $\sigma$. It is measured as signal response in the neighborhood of flaw indication. Peak value noise shall be expressed as percentile i.e. 99 percentile net noise, $n_{net99}$. $n_{net99} \approx 2.36\ \sigma$. |
| Noise Ratio | Ratio of standard deviation of signal response scatter in fitted signal response to standard deviation of noise. |
| Operator | Person who performs NDE procedure for part acceptance. Same as NDE operator or inspector. |
| Physics Model Based POD Analysis | Uses a model that predicts signal response or contrast for a given flaw size. The analysis may use the physics model for the curve/surface fit in the analysis. |
| POD | Probability of detection. Usually it is given as POD number (i.e. 90%) with confidence of 95%. |
| POD/Conf. 90/95 | The point where the 95% lower confidence bound on the POD vs. flaw size curve crosses 90% POD or 90% POD with 95% lower confidence bound. |
| POF | Probability of false positive or false calls. Usually it is given as POF number without confidence. |

| | |
|---|---|
| Qualification | A process of verifying and documenting NDE procedure to its stated objective. Also includes operator qualification and may include operator certification. |
| Quality Aids/Indicators | Physical devices with simulated flaws, or resolution or contrast features used either concurrently or separately to verify technique sensitivity. |
| Reliably Detectable Flaw Size | A flaw size that has been demonstrated or justified by analysis to meet NDE criteria for reliable flaw detection. |
| Reliable Flaw Detection | Reliable flaw detection requires that a flaw size is detected with 90 POD with minimum 95% confidence or there is high confidence that the detected flaw size is greater than $a_{90/95}$. In addition, the POF shall be less than a certain value such as 0.1%. |
| Real Flaw Type | A type of flaw that is expected in a real part. For example, a crack would be a real flaw type. |
| Reject Call | Same as flaw detection call. |
| Single Hit | A flaw is detected due to a single measurement or single channel detection. For example, an imaging technique where target flaw size is comparable to the real resolution, or detection is based on single scan line with the portion of flaw interrogated is about same as the real resolution. Dye penetrant flaw detection is considered to be single hit. |
| Real Resolution | Smallest artificial flaw width that is barely detectable when these flaws are placed at a spacing equal to their width. Modulation transfer function (MTF) of 0.2 is considered to be the barely detectable threshold for measuring the resolution. |
| Standard NDE | NDE procedure that assumes $a_{90/95}$ flaw size. See Table 1 of NASA-STD-5009. Assumes that operator is certified as NAS410 level 2 or 3 in the NDE method. |
| Special NDE | A fracture control term denoting nondestructive inspection personnel, procedures, and equipment with a demonstrated reliability of POD/Conf. 90/95 to detect flaws smaller than those normally detected by Standard NDE procedures. |
| Signal Response POD Analysis | POD analysis that uses comparison of flaw signal response with decision threshold to make a flaw call. MIL-HDBK-1823 *a-hat versus "a"*, physics model based POD and curve fit POD are examples of signal response POD Analysis. |
| Target Flaw Size | Chosen flaw size for NDE procedure qualification using the NDE demonstration test. |
| Sub-target Flaw Size | A flaw size that is smaller than the target flaw size and is used in the NDE demonstration test. Typically sub-target flaw size provides at least 20% lower signal response than that compared to the target flaw size. |
| Transfer Functions | If artificial flaws are used in real part specimens to assess detectability of real flaws, correlation (regression) between the signal responses of real and artificial flaws and their data scatter needs to be accounted. Separate sets of real flaw and same type of artificial flaw correlation specimens are used to obtain the correlation information. The specimens are used to measure the respective signal responses and estimate the above quantities and use in flaw size $a_{90/95}$ or $a_{\mathrm{lv}}$ assessment for a given technique. The correlation specimens usually have simple geometry so that the real and artificial flaws can be manufactured. The correlation is called transfer function. Transfer function is used in estimation of reliably detectable flaw size or the decision threshold when artificial flaws are used. |
| Validated Flaw Size | A flaw size validated by POD methods (i.e. flaw size $a_{90/95}$) or using limited validation (i.e. flaw size $a_{\mathrm{lv}}$). |

## 16.2 Acronyms

| | |
|---|---|
| CNR | Contrast-to-noise Ratio |
| COPV | Composite Overwrapped Pressure Vessel |
| CTR | Contrast to decision Threshold Ratio |
| EDM | Electro discharge Machined |
| FC | Fracture Critical |
| FBH | Flat Bottom Hole |
| FCB | Fracture Control Board |
| FSH | Full Screen Height |
| GLM | General Linear Model |
| ID | Inner Diameter |
| IML | Inner Mold Line |
| IP | Incomplete Penetration |
| IQI | Image Quality Indicator |
| LF | Lack of Fusion |
| MAPOD | Model Assisted POD |
| MLE | Maximum Likelihood Estimation |
| MTF | Modulation Transfer Function |
| NDE | Nondestructive Evaluation |
| OD | Outer Diameter |
| OML | Outer Mold Line |
| SDH | Side Drilled Hole |
| SNR | Signal-to-Noise Ratio |
| SEM | Scanning Electron Microscopy |
| TNR | Decision Threshold-to-Noise Ratio |
| POD | Probability of Detection |
| POD/Conf. | Probability of Detection/Confidence |
| POF | Probability of False positive |
| QQI | Quantitative Quality Indicator |
| RQI | Representative Quality indicators |

## 16.3 Units

| | |
|---|---|
| dB | Decibel |

# APPENDIX A: GENERAL COMPARISON OF NDE PROCEDURE QUALIFICATIONS

| **Comparison Item** | **Standard NDE (POD assumed)** | **Special NDE (POD demonstrated)** | **Custom NDE with Limited Validation (non-POD)** |
|---|---|---|---|
| NASA-STD-5009 | Covered in NASA-STD-5009.<br>Qualified flaw sizes are given in Table 1 of NASA-STD-5009. | Covered in NASA-STD-5009. No flaw sizes provided in NASA-STD-5009. | Not covered in NASA-STD-5009. |
| Methods | Five Methods Only: Ultrasonic Testing,<br>x-ray Film Radiography,<br>Eddy Current Testing,<br>Dye Penetrant Testing and<br>Magnetic Particle Testing | Any method that meets requirements. | Any method that meets requirements. |
| Number of Flaws Needed | None, | Requires known size real flaws in large number.<br>Assumes that POD increases with flaw size.<br>No transfer function is allowed as results cannot certify 90/95 POD/Confidence. | Can use transfer function between real and artificial flaws for signal response analysis.<br>Needs less number of real and artificial flaws.<br>Assumes that POD increases with flaw size. |
| Rationale | POD/Confidence of 90/95 for a standard NDE flaw sizes are assumed to be true based on Shuttle program test data.<br>Based on qualitative assessment or assumption of similarity to the original test data. | POD/Confidence of 90/95 for a Special NDE flaw sizes ($a_{90/95}$) are assured through POD demonstration. | Confidence that limited validation flaw size is greater than the unknown $a_{90/95}$ for the validated method is based on validation of the limited validation model and meeting NDE data conditions derived from the validated model. |
| Margin to $a_{90/95}$ | Assumes large margin between qualified flaw size and unknown $a_{90/95}$ | No margin if MIL-HDBK-1823 GLM method is used.<br>Unknown margin if Binomial point estimate method is used. | Margin is managed to provide high confidence that qualified flaw size is larger than unknown $a_{90/95}$. |
| Probability of False Positive (POF) | POF is not defined but is assumed to be very low ≤ 1%. Since there is no demonstration, it is not assessed. | POF requirement is not defined by NASA, but is assumed to be very low ≤ 1%. Point estimate demonstration does not allow any unexplained false calls. POF is normally not calculated in Binomial point estimate demonstration. Inspection opportunities are counted in hit-miss POF analysis. But defining area of inspection opportunity is a subjective assessment. | POF is incorporated in the signal response model.<br>POF can be calculated same way as in hit-miss analysis.<br>POF used in the model is low i.e. ≤ 0.1% or ≤ 1%. |

## COMPARISON OF NDE PROCEDURE QUALIFICATIONS- GENERAL (CONTINUED)

| **Comparison Item** | **Standard NDE (POD assumed)** | **Special NDE (POD demonstrated)** | **Custom NDE with Limited Validation (non-POD)** |
|---|---|---|---|
| Validation/Technique Control | Meet requirements related to signal-to-noise ratio (SNR), contrast-to-noise ratio (CNR), radiation angle, image resolution etc. given in NASA-STD-5009 | Specific requirements for the qualified procedure only. Requirements state limitations of qualification and requirements to prevent change in the set-up and material/part conditions. Requirements and constraints are stated in the certification letter. | Similar to standard NDE but more advanced controls. Meets conditions on Signal-to-noise ratio (SNR), contrast-to-noise ratio (CNR), decision threshold-to-noise ratio (TNR), image resolution, pixel count etc. to provide high confidence that $a_{lv} > a_{90/95}$ and POF ≤ 1% |
| Qualification Goal | Not applicable.<br>Qualified requirements are given in NASA-STD-5009. | Qualifies a procedure for single point in requirements typically.<br>Hit-miss and Binomial point estimate NDE analysis is not suitable for qualifying quantitative NDE data merit conditions.<br>Signal response analysis has not been used to qualify quantitative NDE data merit conditions as this analysis is not part of MIL-HDBK-1823. | Qualifies a procedure.<br>For signal response analysis, qualifies requirements on procedure for flaw type domain. More flaw types can be considered due to lesser number of flaws used in the demonstration.<br>Hit-miss analysis is not suitable for qualifying quantitative NDE data merit conditions. |
| Applicability | Applicable to specific 5 NDE methods | Applicable to specific written procedure for specific method for a specific flaw type. | Applicable within qualified flaw type domain for a given procedure. |
| Demonstration Requirements for Qualification of NDE Procedure | No demonstration needed.<br>Compliance to NASA-STD-5009 requirements is assessed.<br>Sometimes a delta qualification with demonstration on a few flaws is performed. | POD demonstration needed. | Demonstration on limited number of flaws. Each demonstration covers a flaw type domain. Limited validation is controlled by requirements similar to standard NDE |
| Use of Artificial Flaws: EDM notch, flat bottom hole (FBH), side drilled hole (SDH), simulated disbonds and delaminations, NDE reference standards | Measurements on the artificial flaws are not used beyond calibration. | Measurements on the artificial flaws are not used beyond calibration.<br>POD demonstration usually does not use artificial flaws such as EDM notches, FBH, and SDH. Accepted artificial flaws are disbonds and delaminations. Large number (34-60) of artificial disbonds and delaminations are needed for each flaw detectability type. | Measurements on the artificial flaws may be used beyond calibration such as for creating transfer function between real and artificial flaws. Smaller number of artificial flaws are used in the testing as artificial flaws provide low scatter in signal response. |
| Quality Aids: Image quality indicator, (IQI), quantitative quality indicator (QQI), representative quality indicator (RQI), resolution targets, contrast gages, | Used for verification of sensitivity of NDE procedure used per NASA-STD-5009 | Measurements on quality indicators are used for verification of quality as best assessed by the cognizant NDE professional. | Measurements on quality indicators are used for verification of quality as it relates to detection of qualified flaw size or as assessed by the cognizant NDE professional. |

## APPENDIX B: TABLE FOR VALIDATION CASES

| | Validation Cases | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | Designation | ST-1 | SP-1 | SP-2 | CU-1A | CU-1B | CU-1C | CU-1D | CU-1E | CU-2 | CU-3 |
| | NDE Description | Standard NDE | Special NDE | Special NDE | Custom NDE | Custom NDE | Custom NDE | Custom NDE | Custom NDE | Custom NDE | Custom NDE |
| | Validation Description | Delta Qualification | POD Qualification | POD Qualification | Hybrid of POD and Limited Validation | Limited Validation | Limited Validation | Limited Validation | Limited Validation | Limited Validation | Process Control Validation |
| | Flaw Detection Hits | Single Hit | Single Hit | Single Hit | Single Hit | Single Hit | Single Hit | Multiple Hit | Multiple Hit | Single Hit | Single Hit or Multiple Hit |
| | Data Type | Hit-miss or Signal Response | Hit-miss or Signal Response | Signal Response | Signal Response | Signal Response | Signal Response | Signal Response | Signal Response | Hit-miss | Hit-miss or Signal Response |
| | Flaw Size | $a_{90/95}$ | $a_{90/96}$ | $a_{90/97}$ | $a_{lv}$ | $a_{lv}$ | $a_{lv}$ | $a_{lv}$ | $a_{lv}$ | $a_{lv}$ | Not Quantified |
| **Flaw Set 1: Primary** | Geometry | Part Geometry | Part Geometry | Part Geometry | Simple Geometry | Part Geometry | Part Geometry | Part Geometry | Part Geometry | Part Geometry | Part Geometry or Simple Geometry |
| | Flaw Type | Real | Real | Artificial | Real | Real | Artificial | Real | Artificial | Real | Artificial (or Real) |
| | Quantity* | Limited Validation | POD | POD | POD | Limited Validation | Limited Validation | Limited Validation | Limited Validation | Limited Validation | ASTM or Industry Standards |
| **Flaw Set 2: Secondary** | Geometry | | | | Part Geometry | | | | | | |
| | Flaw Type | | | | Artificial | | | | | | |
| | Quantity* | | | | Limited Validation | | | | | | |
| **Flaw Set 3**: Combined Real and Artificial Flaws for Transfer Function Correlation | Geometry | | | | Simple Geometry | | Simple Geometry | | Simple Geometry | | |
| | Flaw Type | | | | Real and Artificial | | Real and Artificial | | Real and Artificial | | |
| | Quantity* | | | | Correlation / Transfer Function | | Correlation / Transfer Function | | Correlation / Transfer Function | | |
| **Flaw Set 4**: Calibration Reference Standard or Quality Aids | Geometry | Simple or Part | Simple or Part | Simple or Part | Simple or Part | Simple or Part | Simple or Part | Simple or Part | Simple or Part | Simple or Part | Simple or Part |
| | Flaw Type** | Artificial or Quality Aids | Artificial or Quality Aids | Artificial or Quality Aids | Artificial | Artificial | Artificial or Quality Aids | Artificial | Artificial | Quality Aids | Artificial or Quality Aids |
| | Quantity | Min. 1 | Min. 1 | Min. 1 | Min. 1 | Min. 1 | Min. 1 | Min. 1 | Min. 1 | Min. 1 | Min. 1 |

* - Quantity depends on validation method indicated

** - “Artificial” implies instrument standardization on artificial flaw and “Quality Aids” implies that any of IQI, RQI, QQI, line pair gages, image quality phantoms etc. are used to indicate quality.

# APPENDIX C: COMPARISON OF DATA ANALYSIS METHODS BY DATA TYPE

| | **Data Type** | | |
|---|---|---|---|
| Comparison Item | **Hit-miss (Binary Data) versus Flaw Size** | **Signal Response versus Flaw Size (*a-hat* versus "a") Analysis, Single Hit** | **Signal Response as 1D, 2D or 3D Map Generated using Controlled Scanning Scheme, Multiple Hit** |
| Acceptance Criteria Logic | Using flaw indication size and length-to-width aspect ratio to define acceptance criteria. May include criteria for clustering flaw indications in close proximity to each other as a single larger flaw. | Using calibration flaw response to set a decision threshold | Can use calibration flaw based decision threshold and/or size based decision threshold |
| Examples | Dye penetrant testing, magnetic particle testing, film radiography | Real time testing using ultrasonic A-scan and eddy current impedance display | Ultrasonic C-scan, eddy current C-scan, x-ray computed tomography |
| POD Analysis (Statistical) | 1. Binomial Point Estimate POD (NASA preferred method), Flaw detectability size > $a_{90/95}$<br>- Need minimum 34 flaws<br>2. General Linear Model POD (GLM) per MIL-HDBK-1823,<br>- Flaw detectability size = $a_{90/95}$<br>- Need minimum 60 flaws<br>- Need real flaws as transfer function between real and artificial flaws is not defined | 1. General Linear Model POD (GLM) per MIL-HDBK-1823<br>- Flaw detectability sixe = $a_{90/95}$<br>2. Physics Model and Curve Fit POD (Not Common)<br>3. Flaw detectability size = $a_{90/95}$<br>- Need real flaws as transfer function is not defined<br>- Need large number of flaws (minimum 40) for each flaw detectability type | 1. Convert the data to Hit-miss or signal response and use POD analysis<br>2. Need real flaws as transfer function is not defined<br>3. Need large number of flaws (minimum 34 for hit-miss analysis) |
| Custom NDE with Limited Validation[9,10,8,11] | Hit-Miss Analysis,<br>1. Flaw detectability size assumed > $a_{90/95}$<br>2. Need real flaws<br>3. Need smaller number (i.e. 16) flaws for each flaw detectability type<br>4. Qualified size may be much larger than corresponding hit-miss analysis per MIL-HDBK-1823 | Signal Response Analysis<br>1. Verifies that 90/95 POD/conf. conditions are met<br>2. Flaw detectability size > $a_{90/95}$<br>3. Combination of real and artificial cracks can be used with transfer function<br>4. Need smaller number (i.e. minimum 6) of flaws for each flaw detectability type | Contrast and Resolution Analysis<br>1. Verifies that 90/95 POD/conf. conditions are met for contrast and resolution of indications<br>2. Combination of real and artificial flaws can be used<br>3. Need smaller number (i.e. minimum 6) of flaws for each flaw detectability type |

## APPENDIX D: COMPARISONN OF NDE DATA ANALYSIS BY ANALYSIS METHODS

| **Comparison Item** | **MIL-HDBK-1823 Signal Response (*a-hat*) Method** | **MIL-HDBK-1823 Hit-miss Method** | **Binomial Point Estimate Method for Hit-miss Data** | **Engineering Method using Signal Response Curve Fit /Model Fit** | **Engineering Method using Limited Validation for Signal Response Data** | **Engineering Method using Limited Validation for Hit-miss Data** |
|---|---|---|---|---|---|---|
| | 1 | 2 | 3 | 4 | 5 | 6 |
| Flaw sizes | Provides smallest flaw size $a_{90/95\_a\_hat}$ | $a_{90/95\_hit\text{-}miss} >$ $a_{90/95\_a\_hat}$ | $a_{90/95\_point\_estimate} >$ $a_{90/95\_hit\_miss}$ | $a_{90/95\_curve\_fit}$ $\approx a_{90/95\_a\_hat}$ | $a_{a\_hat\_lv}$ $> a_{90/95\_a\_hat}$ | $a_{hit\_miss\_lv}$ $> a_{90/95\_a\_hat}$ |
| Comparison of Flaw sizes | $a_1$ smallest | $a_2 > a_1$ | $a_3 > a_2, a_1$ | $a_4 \approx a_1$ | $a_5 > a_1, a_4$ | $a_6 > a_1, a_4$ |
| Current Use at NASA | Rarely used | Rarely used except for research | Commonly used by NASA | Rarely used except for research | New method | New method |
| Number of Flaws | Minimum 40 real flaws of known size and distribution for each flaw detectability type | Minimum 60 real flaws of known size and distribution for each flaw detectability type,. 120 inspection opportunities will provide smaller $a_{90/95\_hit\text{-}miss}$ | Minimum 34 real flaws of same size for each flaw detectability type. Minimum 3 blank specimens needed. | Minimum 40 real flaws of known size and distribution for each flaw detectability type | Minimum 6 real flaws of known size and distribution for each flaw detectability type. Minimum 34 flaws for operator certification. | Minimum 16 real flaws of known size and distribution for each flaw detectability type. Minimum 34 flaws for operator certification. |
| Number of False Call Opportunities | Noise analysis. Based on scatter of minimum 40 real flaws or 40 measurements in unflawed region | For POF < 1%, at least 100 unflawed opportunities should be available. For POF < 0.1%, at least 1000 unflawed opportunities are needed | For POF < 1% at least 100 unflawed opportunities should be available. For POF < 0.1%, at least 1000 unflawed opportunities are needed | Noise analysis. Based on scatter of minimum 40 real flaws or 40 measurements on unflawed region | Noise analysis. Based on scatter of minimum 6 real flaws and minimum 40 measurements in unflawed region | For POF < 1%, at least 100 unflawed opportunities are needed. For POF < 0.1%, at least 1000 unflawed opportunities are needed |

# APPENDIX E: CUSTOM NDE WITH LIMITED VALIDATION FOR SIGNAL RESPONSE VERSUS FLAW SIZE DATA[9,10,11]

## 1.0 CU-1A: Custom NDE, Hybrid of POD and Limited Validation, Single Hit, Signal Response, Real and Artificial Flaws [9,10,11]

### 1.1 Assumptions

1.1.1 Linear correlation between signal response and flaw size.
1.1.2 Noise is unimodal.
1.1.3 Assumes conservative scatter factor or noise ratio of $R_\sigma \geq 2.0$
1.1.4 Provides flaw size estimate $\geq a_{90/95}$

### 1.2 Requirements

1.2.1 Demonstration shall have sufficient number of flaw data points.
There shall be minimum 3 flaws each for target and corresponding smaller or sub-target size, i.e. 6 flaws minimum for each extreme or worst case flaw type.
Sub-target size flaws shall have ~15-25% lower signal response or size than that of target size on the average.
There shall be minimum 34 flaws in the entire demonstration.
1.2.2 Noise shall be measured at minimum 40 unflawed locations.
1.2.3 Demonstration target and larger size flaws shall meet conditions 1, 2, and 3 for POF = 0.1%
1.2.4 Demonstration sub-target size flaws shall meet conditions 1, 2, and 3 for POF = 1%
1.2.5 NDE procedure shall be well documented (written procedure), optimized and controlled to provide repeatable results.
1.2.6. Demonstration shall use NAS410 certified and qualified operators.
1.2.6 Demonstration shall provide repeatable results.
Demonstration testing shall be repeated one more time preferably with different operator to verify repeatability of results.
1.2.7 Signal shall be measured as worst probable case of beam or probe orientation and scan line location with respect to the flaws used in the validation testing.
1.2.8 Demonstration and NDE procedure shall provide required coverage.
Worst case or extreme flaw detectability types (i.e. worst or extreme locations, orientations, surfaces, depths, part thicknesses) from detectability considerations and from design considerations are simulated in the demonstration samples.
All likely flaw type domains (i.e. locations, orientations, surfaces, depths, part thicknesses) are simulated covering all zones and areas that need to be inspected.
1.2.9 Validation shall address transfer function (or knockdown) due to differences in signal response of real and artificial flaw, if artificial flaws are used to augment real flaws.
Transfer function may be applied to the flaw size or signal response.
Transfer function may be determined by measuring signal response on separate sets of specimens, one set for artificial flaws and other for real flaws. Minimum 10 flaws shall be used for each real flaw detectability type (set 3A). Minimum 10 flaws shall be used for each artificial flaw detectability type (set 3B).
Transfer function study specimens may have simpler geometry to achieve manufacturability.
Transfer function data from similar applications may be used.
1.2.10 If artificial flaws are used for calibration minimum 3 flaws of different sizes shall be used (set 4). One of the three flaws shall have signal response close to the estimated target size real flaw response. Second flaw shall have a lower response by 15-25% compared to the first flaw. The third flaw may be as large as the target flaw size. Validation shall use procedure given in Ref. [10] and [11].
1.2.11 If artificial flaws are used, minimum 6 flaws per flaw detectability type shall be used.
1.2.12 NDE technique shall have monotonically increasing signal response about target flaw size.

Table 1: Merit ratios for reliable flaw detection

| **Condition** | **Description** | **Abbreviation** | **Condition**<br>**POF = 0.1%** | **Condition**<br>**POF = 1%** |
|---|---|---|---|---|
| **1** | Contrast-to-standard deviation of noise ratio | $CNR$ | ≥ 7.3 | ≥ 5.2 |
| **2** | Net decision threshold-to-standard deviation of noise | $TNR$ | ≥ 3.9 | ≥ 2.8 |
| **3** | Ratio of the contrast-to-net decision threshold ratio | $CTR$ | ~1.9 (min 1.5) | ~ 1.9 (min. 1.5) |

## 1.3 Procedure[9,10,11]

1.3.1 Step 1: Get the target flaw size, $a_{\mathrm{lv}}$ to be detected reliably from design/system engineering.

1.3.2 Step 2: Measure noise and determine standard deviation of noise on real part.

1.3.3 Step 3: By trial-and-error measure signal response for real flaws of target size made in simple geometry i.e. plate specimen.

1.3.4 Step 4: Determine 10 real flaw sizes such that their signal responses are within a range of $\hat{a}_{\mathrm{real\_target\_plate}} - 5\sigma$ to $\hat{a}_{\mathrm{real\_target\_plate}} + 3\sigma$. These response values shall be uniformly distributed as much as possible. Set 3A.

1.3.5 Step 5: Make the real flaw specimens in simple geometry (i.e. plate) specimens. Set 3A.

1.3.6 Step 6: Make identical artificial flaw specimens in simple geometry with same size flaws as the real flaws. This is called Set 3B. Set 3A and 3B together make set 3.

1.3.7 Step 7: Obtain the signal responses from both real and artificial flaws.

1.3.8 Step 8: Plot signal response versus flaw size. Determine correlation between the responses of real and artificial flaw and scatter within each dataset. Determine decision threshold $\hat{a}_{decision\_plate}$ needed from artificial flaw to detect real flaw of target size in the simple geometry specimen. See Ref. [11].

1.3.9 Step 9: Make a point estimate POD (min. 34 real flaws) or GLM POD (min. 40 real flaw). Set 1.

1.3.10 Step10: Determine artificial flaw size equivalent of size equivalent target real flaw. This will be called the target artificial flaw. Build three artificial target flaw specimens and three sub flaw specimens with signal response 15-25% lower. This is called set 2.

1.3.11 Step 11: Also build a calibration specimen with artificial flaws in real part geometry with flaw size same as the target artificial flaw. This is called set 4.

1.3.12 Step 12: Determine decision threshold for target flaw size using calibration flaws as given in Ref. [11] and inspect set 1 and 2.

1.3.13 Step 13: Verify that conditions 1, 2, and 3 for 0.1% POF are met for the target flaw size for set 2.

1.3.14 Step 14: Verify that conditions 1, 2, and 3 for 1% POF are met for the sub-target flaw size in set 2.

1.3.15 Step 15: If set 1 demonstration is successful, note merit ratios of flaws in set 1. If merit ratios from set1 and 2 flaws are comparable, and meet Table 1 requirements target flaw size is qualified as $a_{\mathrm{lv}}$. Else take a larger flaw size and repeat the procedure until Table 1 requirements are met.

## 2.0 Case CU-1B: Custom NDE, Limited Validation, Single Hit, Signal Response, Real Flaws[9,10,11]

### 2.1 Assumptions

2.1.1 Linear correlation between signal response and flaw size.
2.1.2 Noise is unimodal.
2.1.3 Assumes conservative scatter factor or noise ratio of $R_\sigma \geq 2.0$
2.1.4 Provides flaw size estimate $\geq a_{90/95}$.

### 2.2 Requirements

2.2.1 Demonstration shall have sufficient number of flaw data points.
There shall be minimum 3 flaws each for target and corresponding smaller or sub-target size, i.e. 6 flaws minimum for each extreme or worst case flaw type.
Sub-target size flaws shall have ~15-25% lower signal response or size than that of target size on the average.
There shall be minimum 34 flaws in the entire demonstration.
2.2.2 Noise shall be measured at minimum 40 unflawed locations.
2.2.3 Demonstration target and larger size flaws shall meet conditions 1, 2, and 3 for POF = 0.1%.
2.2.4 Demonstration sub-target size flaws shall meet conditions 1, 2, and 3 for POF = 1%.
2.2.5 NDE procedure shall be well documented (written procedure), optimized and controlled to provide repeatable results.
2.2.6. Demonstration shall use NAS410 certified and qualified operators.
2.2.7 Demonstration shall provide repeatable results.
Demonstration testing shall be repeated one more time preferably with different operator to verify repeatability of results.
2.2.8 Signal shall be measured as worst probable case of beam or probe orientation and scan line location with respect to the flaws used in the validation testing.
2.2.9 Demonstration and NDE procedure shall provide required coverage.
Worst case or extreme flaw detectability types (i.e. worst or extreme locations, orientations, surfaces, depths, part thicknesses) from detectability considerations and from design considerations are simulated in the demonstration samples
All likely flaw type domains (i.e. locations, orientations, surfaces, depths, part thicknesses) are simulated covering all zones and areas that need to be inspected.
2.2.10 Validation shall address transfer function (or knockdown) due to differences in signal response of real and artificial flaw, if artificial flaws are used to augment real flaws.
Transfer function may be applied to the flaw size or signal response.
Transfer function may be determined by measuring signal response on separate sets of specimens, one set for artificial flaws and other for real flaws. Minimum 10 flaws shall be used for each real flaw detectability type (set 3A).
Minimum 10 flaws shall be used for each artificial flaw detectability type (set 3B).
Transfer function study specimens may have simpler geometry to achieve manufacturability.
2.2.11 Artificial flaws are used for calibration (set 4). Minimum 3 flaws of different sizes shall be used. One of the three flaws shall have signal response close to the estimated target size real flaw response. Second flaw shall have a lower response by 15-25% compared to the first flaw. The third flaw may be as large as the target flaw size. Validation shall use procedure given in Ref. [10] and [11].
2.2.13 NDE technique shall have monotonically increasing signal response about target flaw size.

Table 1: Merit ratios for reliable flaw detection

| **Condition** | **Description** | **Abbreviation** | **Condition**<br>**POF = 0.1%** | **Condition**<br>**POF = 1%** |
|---|---|---|---|---|
| **1** | Contrast-to-standard deviation of noise ratio | $CNR$ | ≥ 7.3 | ≥ 5.2 |
| **2** | Net decision threshold-to-standard deviation of noise | $TNR$ | ≥ 3.9 | ≥ 2.8 |
| **3** | Ratio of the contrast-to-net decision threshold ratio | $CTR$ | ~1.9 (min 1.5) | ~ 1.9 (min. 1.5) |

## 2.3 Procedure

2.3.1 Step 1: Design the flaw size demonstration set for each flaw (worst or extremes) detectability type, minimum 6 flaws per flaw detectability type.
Measure noise and determine standard deviation of noise.
Based on trial-and-error testing, determine $a_{\mathrm{lv}}$ that can meet Conditions 1, 2, and 3 for 0.1% POF. This is the target flaw size.
Based on trial-and-error testing, determine $a_{\mathrm{lv}}$ that can meet Conditions 1, 2, and 3 for 1% POF. This is the sub-target flaw size. Signal response for this flaw size shall be 15-25% lower than that from target flaw size.

2.3.2 Step 2: Build the demonstration set.
There should be minimum 3 flaws for target size, 3 flaws for sub target size.

2.2.3 Step 3: Artificial flaws are used for calibration (set 4). Minimum 3 flaws of different sizes shall be used. One of the three flaws shall have signal response close to the estimated target size real flaw response. Second flaw shall have a lower response by 15-25% compared to the first flaw. The third flaw may be as large as the target flaw size.

2.2.4 Step 4: Determine decision threshold level with respect to calibration flaws. Validation shall use procedure given in Ref. [10] and [11].

2.3.5 Step 5: Conduct the NDE technique demonstration.
Calculate the ratios of conditions 2, 3, 4 for respective POF.
If ratios meet the conditions for each demonstration flaw, target flaw size is qualified as $a_{\mathrm{lv}}$.

# 3.0 Case CU-1C: Custom NDE, Limited Validation, Single Hit, Signal Response, Real and Artificial Flaws[9,10,11]

## 3.1 Assumptions

3.1.1 Linear correlation between signal response and flaw size.
3.1.2 Noise is unimodal.
3.1.3 Assumes conservative scatter factor or noise ratio of $R_\sigma \geq 2.0$.
3.1.4 Provides flaw size estimate $\geq a_{90/95}$.

## 3.2 Requirements

3.2.1 Demonstration shall have sufficient number of flaw data points.
There shall be minimum 3 flaws each for target and corresponding smaller or sub-target size, i.e. 6 flaws minimum for each extreme or worst case flaw type.
Sub-target size flaws shall have ~15-25% lower signal response or size than that of target size on the average.
There shall be minimum 34 flaws in the entire demonstration.
3.2.2 Noise shall be measured at minimum 40 unflawed locations.
3.2.3 Demonstration target and larger size flaws shall meet conditions 1, 2, and 3 for POF = 0.1%.
3.2.4 Demonstration sub-target size flaws shall meet conditions 1, 2, and 3 for POF = 1%.
3.2.5 NDE procedure shall be well documented (written procedure), optimized and controlled to provide repeatable results.
3.2.6. Demonstration shall use NAS410 certified and qualified operators.
3.2.6 Demonstration shall provide repeatable results.
Demonstration testing shall be repeated one more time preferably with different operator to verify repeatability of result.
3.2.7 Signal shall be measured as worst probable case of beam or probe orientation and scan line location with respect to the flaws used in the validation testing.
3.2.8 Demonstration and NDE procedure shall provide required coverage.
Worst case or extreme flaw detectability types (i.e. worst or extreme locations, orientations, surfaces, depths, part thicknesses) from detectability considerations and from design considerations are simulated in the demonstration samples.
All likely flaw type domains (i.e. locations, orientations, surfaces, depths, part thicknesses) are simulated covering all zones and areas that need to be inspected.
3.2.9 Validation shall address transfer function (or knockdown) due to differences in signal response of real and artificial flaw, if artificial flaws are used to augment real flaws.
Transfer function may be applied to the flaw size or signal response
Transfer function may be determined by measuring signal response on separate sets of specimens, one set for artificial flaws and other for real flaws. Minimum 10 flaws shall be used for each set for each grouping of a flaw detectability types
Transfer function study specimens may have simpler geometry to achieve manufacturability.
Transfer function data from similar applications may be used.
3.2.10 If artificial flaws are used for calibration minimum 3 flaws of different sizes shall be used. One of the three flaws shall have signal response close to the estimated target size real flaw response. Second flaw shall have a lower response by 15-25% compared to the first flaw. The third flaw may be as large as the target flaw size. Validation shall use procedure given in Ref. [9].
3.2.11 If artificial flaws are used, minimum 6 flaws per flaw detectability type shall be used. Some of these flaws can be same as calibration flaws.
3.2.12 NDE technique shall have monotonically increasing signal response about target flaw size.

Table 1: Merit ratios for reliable flaw detection

| **Condition** | **Description** | **Abbreviation** | **Condition POF = 0.1%** | **Condition POF = 1%** |
|---|---|---|---|---|
| **1** | Contrast-to-standard deviation of noise ratio | $CNR$ | ≥ 7.3 | ≥ 5.2 |
| **2** | Net decision threshold-to-standard deviation of noise | $TNR$ | ≥ 3.9 | ≥ 2.8 |
| **3** | Ratio of the contrast-to-net decision threshold ratio | $CTR$ | ~1.9 (min 1.5) | ~ 1.9 (min. 1.5) |

## 3.3 Procedure

3.3.1 Step 1: Get the target flaw size, $a_{\mathrm{lv}}$ to be detected reliably from design/system engineering.
3.3.2 Step 2: Measure noise and determine standard deviation of noise on real part.
3.3.3 Step 3: By trial-and-error measure signal response for real flaws of target size made in simple geometry i.e. plate specimen. Set 3A.
3.3.4 Step 4: Determine 10 flaw sizes such that their signal responses are within a range of $\hat{a}_{\mathrm{real\_target\_plate}} - 5\sigma$ to $\hat{a}_{\mathrm{real\_target\_plate}} + 3\sigma$. These response values shall be uniformly distributed as much as possible.
3.3.5 Step 5: Make the real flaw specimens in simple geometry (i.e. plate) specimens. This is called set 3A.
3.3.6 Step 6: Make identical artificial flaw specimens in simple geometry with same size flaws as the real flaws. This is called set 3B.
3.3.7 Step 7: Obtain the signal responses from both real and artificial flaws.
3.3.8 Step 8: Plot signal response versus flaw size. Determine correlation between the responses of real and artificial flaw and scatter within each dataset. Determine decision threshold $\hat{a}_{decision\_plate}$ needed from artificial flaw to detect real flaw of target size in the simple geometry specimen. See Ref. [11].
3.3.9 Step 9: Determine artificial flaw size equivalent of size equivalent target real flaw. This will be called the target artificial flaw. Build three artificial target flaw specimens and three sub flaw specimens with signal response 15-25% lower. This is called set 1.
3.3.10 Step 10: Also build a calibration specimen with artificial flaws in real part geometry with flaw size same as the target artificial flaw. This is called set 4.
3.3.11 Step 11: Determine decision threshold for target flaw size as given in Ref. [11].
3.3.12 Step 12: Verify that conditions 1, 2, and 3 for 0.1% POF are met for the target flaw size.
3.3.13 Step 13: Verify that conditions 1, 2, and 3 for 1% POF are met for the sub-target flaw size.
If ratios meet the conditions for each demonstration flaw, target flaw size is qualified as $a_{\mathrm{lv}}$.

## 4.0 Case CU-1D: Custom NDE, Limited Validation, Multiple Hit, Signal Response, Real Flaws[9,10,11]

### 4.1 Assumptions

4.1.1 Linear correlation between signal response and flaw size.
4.1.2 Noise is unimodal.
4.1.3 Assumes conservative scatter factor or noise ratio of ≥ 2.0 for correlation.
4.1.4 Provides flaw size estimate with $\geq a_{90/95}$.
4.1.5 Maps represent signal response.

### 4.2 Requirements

4.2.1 1D Flaw Image (cluster of pixels in one dimensional array.
  4.2 1.1 Flaw image shall be mapped either in a linear array of minimum 9 resolution size pixels.
  4.2.1.2 Contrast-to-noise ratio, CNR for the chosen pixel grid of flaw image shall be ≥ 2.5 for each flaw.
4.2.2 2D Flaw Image (cluster of pixels in 2 dimensional array)
  4.2.2.1 Flaw image shall be mapped either in an array of minimum 3 x 3 resolution size pixel or in an array of minimum 9 resolution size pixels.
  4.2.2.2 Contrast-to-noise ratio, CNR for the chosen pixel grid of flaw image shall be ≥ 2.5 for each flaw.
4.2.3 There shall be minimum 3 flaws each for target and corresponding smaller or sub-target size, i.e. 6 flaws minimum for each extreme or worst case flaw type.
4.2.4 Noise shall be measured at minimum 40 unflawed locations in the neighborhood of flaws.
4.2.5 POF shall be determined using MIL-HDBK-1823 hit-miss analysis and should be less than 0.1%.
4.2.6 Sub-target size flaws shall have ~15-25% lower signal response or size than that of target size on the average.
4.2.7 There shall be minimum 10 flaws in entire demonstration.
4.2.8 If above requirements are not met then requirements for flaw detection based on real time or single hit (Case CU-1B) apply with average signal amplitude taken as the flaw signal response.

### 4.3 Procedure

4.3.1 Step 1: Obtain the target flaw size, $a_{\text{lv}}$ from design/system engineering.
4.3.2 Step 2: Measure noise and determine standard deviation of noise on real part.
4.3.3 Step 3: By trial-and-error measure signal response for real flaws and assess, if necessary, conditions for validation can be met for given $a_{\text{lv}}$. Else, choose a flaw size where the conditions can be met. Get approval to set the validation for this flaw size and build target size and sub-target size flaw specimens with minimum three flaw each for each flaw type. Set 1.
4.3.4 Step 4: Also build a calibration specimen with artificial flaws in real part geometry with flaw size same as the target artificial flaw. This is called set 4.
4.3.5 Step 5: Determine decision threshold for target flaw size using calibration flaw as given in Ref. [11].
4.3.6 Step 6: Inspect the validation specimens and analyze the data per requirements above for multiple hits or 1D and 2D map of flaw. If the requirements are met for each demonstration flaw, target flaw size is qualified as $a_{\text{lv}}$.

## 5.0 Case CU-1E: Custom NDE, Limited Validation, Multiple Hit, Signal Response, Artificial Flaws[9,10,11]

### 5.1 Assumptions

5.1.1 Linear correlation between signal response and flaw size.
5.1.2 Noise is unimodal.
5.1.3 Assumes conservative scatter factor or noise ratio of ≥ 2.25 for correlation.
5.1.4 Provides flaw size estimate with ≥ $a_{90/95}$.
5.1.5 Maps represent signal response.

### 5.2 Requirements

5.2.1 1D Flaw Image (cluster of pixels in one dimensional array)
  5.2 1.1 Flaw image shall be mapped either in a linear array of minimum 9 resolution size pixels.
  5.2.1.2 Contrast-to-noise ratio, CNR for the chosen pixel grid of flaw image shall be ≥ 2.5 for each flaw.
5.2.2 2D Flaw Image (cluster of pixels in 2 dimensional array)
  5.2.2.1 Flaw image shall be mapped either in an array of minimum 3 x 3 resolution size pixel or in an array of minimum 9 resolution size pixels.
  5.2.2.2 Contrast-to-noise ratio, CNR for the chosen pixel grid of flaw image shall be ≥ 2.5 for each flaw.
5.2.3 There shall be minimum 3 flaws each for target and corresponding smaller or sub-target size, i.e. 6 flaws minimum for each extreme or worst case flaw type.
5.2.4 Noise shall be measured at minimum 40 unflawed locations.
5.2.5 POF shall be determined using MIL-HDBK-1823 hit-miss analysis and should be less than 0.1%.
5.2.6 Sub-target size flaws shall have ~15-25% lower signal response or size than that of target size on the average.
5.2.7 There shall be minimum 10 flaws in entire demonstration.
5.2.8 If above requirements are not met then requirements for flaw detection based on real time or single hit (Case CU-1C) apply with average signal amplitude taken as the flaw signal response.

### 5.3 Validation Procedure

5.3.1 Step 1: Obtain the flaw size, $a_{\text{lv}}$ from design/system engineering.
5.3.2 Step 2: Measure noise and determine standard deviation of noise on real part.
5.3.3 Step 3: By trial-and-error measure signal response for real flaws of target size made in simple geometry i.e. plate specimen.
5.3.4 Step 4: Determine 10 flaw sizes such that their signal responses are within a range of $\hat{a}_{\text{real\_target\_plate}} - 5\sigma$ to $\hat{a}_{\text{real\_target\_plate}} + 3\sigma$. These response values shall be uniformly distributed as much as possible.
5.3.5 Step 5: Make the real flaw specimens in simple geometry (i.e. plate) specimens. Set 3A.
5.3.6 Step 6: Make identical artificial flaw specimens in simple geometry with same size flaws as the real flaws. Set 3B.
5.3.7 Step 7: Obtain the signal responses from both real and artificial flaws.
5.3.8 Step 8: Plot signal response versus flaw size. Determine correlation between the responses of real and artificial flaw and scatter within each dataset. Determine decision threshold $\hat{a}_{decision_plate}$needed from artificial flaw to detect real flaw of target size in the simple geometry specimen. See Ref. [8].
5.3.9 Step 9: Determine artificial flaw size equivalent of size equivalent target real flaw. This will be called the target artificial flaw. Build three artificial target flaw specimens and three sub flaw specimens with signal response 15-25% lower. This is called set 1.
5.3.10 Step 10: Also build a calibration specimen with artificial flaws in real part geometry with flaw size same as the target artificial flaw. This is called set 4.
5.3.11 Step 11: Determine decision threshold for target flaw size using calibration flaw as given in Ref. [11].
5.3.14 Step 12: Inspect the validation specimens and analyze the data per Custom NDE with limited validation requirements for multiple hits or 1D and 2D map of flaw. If the requirements are met for each demonstration flaw, target flaw size is qualified as $a_{\text{lv}}$.

# APPENDIX F: CUSTOM NDE WITH LIMITED VALIDATION FOR HIT-MISS VERSUS FLAW SIZE DATA[8]

## 1.0 Case CU-2: Custom NDE, Limited Validation, Single Hit, Hit-miss, Real Flaws[8]

### 1.1 Assumptions

1.1.1 The limited validation method assumes that noise has constant standard deviation and POD function can be modeled by cumulative Standard distribution.

### 1.2 Requirements

1.2.1 Demonstration shall have a minimum of 16 flaws for each extreme or worst case flaw type. There shall be minimum 34 flaws in the entire demonstration.
1.2.2 Flaw sizes shall be uniformly distributed as much as possible as given below.
  1.2.2.1 Based on testing, determine largest missed flaw size and smallest detected flaw size. Approximately, 40-45% of data points shall span from smallest detected flaw to largest missed flaw.
  1.2.2.2 The remaining data points shall be equally distributed to the left of smallest detected flaw and to the right of largest missed flaw.
1.2.3 Validation shall use equations[8] provided to estimate the validated flaw size.
1.2.4 POF shall be determined using MIL-HDBK-1823 hit-miss analysis and should be less than 0.1%.
1.2.5 NDE procedure shall be well documented, optimized and controlled to provide repeatable results.
  1.2.5.1 Written procedure, qualified operators.
1.2.6 Demonstration shall provide repeatable results.
  1.2.6.1 Demonstration testing shall be repeated one more time preferably with different operator to verify repeatability of results.
1.2.7 Addresses transfer function between real and artificial flaws.
  1.2.8.1 Knockdown due to differences in test sample including test sample flaw and real part including real flaw
  1.2.8.2 Knockdown may be applied to the flaw size.
1.2.8 Scan shall be set up as worst probable case of beam or probe orientation and scan line location with respect to the flaws used in the validation testing.
1.2.9 Demonstration and NDE procedure shall provide required coverage.
  1.2.9.1 All likely flaw type domains (i.e. locations, orientations, surfaces, depths, part thicknesses) are simulated covering all zones and areas that need to be inspected.
1.2.10 NDE technique shall have monotonically increasing signal response or POD about target flaw size.

### 1.3 Procedure

1.3.1 Step 1: Design the flaw size demonstration set for each flaw detectability type, minimum 16 flaws per flaw detectability type
  1.3.1.1 Based on testing, attempt shall be made to determine largest missed flaw size and smallest detected flaw size.
  1.3.1.2 Approximately, 40-45% of data points shall span from smallest detected flaw to largest missed flaw.
  1.3.1.3 The remaining data points shall be equally distributed to the left of smallest detected flaw and to the right of largest missed flaw.
1.3.2 Step 2: Conduct the NDE technique demonstration.
1.3.3 Step 3: Analyze data using equations provided to determine $a_{lv}$.

# APPENDIX G: RECOMMENDED X-RAY RADIOGRAPHY NONDESTRUCTIVE EVALUATION REQUIREMENTS[14]

## 1.0 General

a. Requirements of NASA-STD-5009 supersede, in case of conflict with following requirements. NASA-STD-5009 Table 1 minimum detectable crack sizes for standard radiographic procedures are based solely on film radiography. With the conversion to digital radiographic techniques and methods, it is essential to assess the capability of such systems to achieve the table 1 detectable crack sizes. All cracklike indications shall be rejected.

## 2 0 Film Radiography

b. Film radiographic inspections shall be in accordance with ASTM-E-1742, Standard Practice for Radiographic Examination, or NASA fracture control and NASA NDE Engineering-approved contractor internal specifications with the following additional requirements:
c. The minimum radiographic inspection sensitivity level shall be 2-1T.
d. Film density shall be 2.5 to 4.0.
e. The center axis of the radiation beam shall be within +/-5 degrees of the assumed crack plane orientation.

## 3.0 Digital Radiography (DR) and Computed Radiology (CR)

### 3.1 General

a. Digital Radiographic (DR) inspections, shall be in accordance with ASTM-E-2698, Standard Practice for Radiological Examination Using Digital Detector Arrays, or NASA fracture control and NASA NDE Engineering-approved contractor internal specifications.
b. Computed Radiographic (CR) inspections, shall be in accordance with ASTM-E-2033, Standard Practice for Computed Radiology (Photostimulable Luminescence Method), or NASA fracture control and NASA NDE Engineering-approved contractor internal specifications.
c. DR and CR techniques shall be qualified in accordance with requirements of this section.
d. DR and CR system performance parameters such as those listed in table 1shall be measured or checked and the following requirements observed:
    (1) Acceptance limits shall be established per ASTM E2445 for CR systems and ASTM E2737 for DR systems.
    (2) Accompanying display monitors shall be checked per ASTM E2698.
    (3) Frequency of checks shall be approved by the responsible NDE engineering organization.
e. Minimum IQI (ASTM E1025) sensitivity shall be 2-1T.
f. Minimum contrast-to-noise ratio on the 1T hole shall be 2.5. Noise is measured as standard deviation.
g. Minimum 1T hole size-to-normalized unsharpness ratio shall be 3.
h. The center axis of the radiation beam shall be within +/-5 degrees of the assumed crack plane orientation.

*i. The limits on contrast-to-noise ratio, 1T hole size-to-normalized unsharpness ratio, and x-ray beam angle may need to be adjusted in order to achieve the table 1 detectable crack sizes. Actual limits shall be established based on qualification data obtained on a representative quality indicator (RQI) per ASTM E1817.*

## Table 1—System Performance Tests for CR/DR

The purpose of this table is to provide a table with CR and DR system performance monitoring methods.

Table 1—System Performance Tests for CR/DR

| | Test | Method/Reference |
|---|---|---|
| Computed Radiology | SRb (detector basic spatial resolution) | ASTM E2445, Duplex wire ASTM E2002 |
| | Plaque IQI sensitivity (visual) (contrast and spatial resolution) | ASTM E1742 |
| | Contrast sensitivity | ASTM E2445, ASTM E1647 |
| | Shading | ASTM E2445 |
| | Jitter | ASTM E2445 |
| | Banding | ASTM E2445 |
| | Erasure | ASTM E2445 |
| | Scanner slippage | ASTM E2445 |
| | Scan column dropout | ASTM E2445 |
| | Scan line integrity | ASTM E2445 |
| | Afterglow (Blooming/Flare) | ASTM E2445 |
| | Geometric distortion | ASTM E2445 |
| | EPS | ASTM E2445, ASTM E746 |
| | Geometric distortion | ASTM E2445 |
| | Photo multiplier tube non-linearity | ASTM E2445 |
| | Burn-In test | ASTM E2445 |
| | Spatial linearity | ASTM E2445 |
| | Central beam alignment | ASTM E2445 |
| | Imaging plate artifacts | ASTM E2445 |
| | Imaging plate response | ASTM E2445 |
| | Imaging plate fading | ASTM E2445 |
| Digital Detector Array | System spatial resolution | ASTM E2737 |
| | Contrast sensitivity (uses CNR) | ASTM E2737 |
| | Signal-to-Noise ratio (SNR) | ASTM E2737 |
| | Signal level | ASTM E2737 |
| | Bad pixel mapping | ASTM E2737 |
| | Gain calibration | ASTM E2737 |
| | Offset calibration | ASTM E2737 |
| | Plaque IQI sensitivity | ASTM E2737 |
| | Lag Calculation | ASTM E2737 |
| | Burn-In test | ASTM E2737 |
| | Motion system repeatability | User selected |
| Display Monitors | Full modulation check | ASTM E2698 |
| | Flicker check | ASTM E2698 |
| | Distortion check | ASTM E2698 |
| | 1% line contrast | ASTM E2698 |
| | 5% line contrast | ASTM E2698 |
| | Contrast ratio | ASTM E2698 |
| | Luminous intensity | ASTM E2698 |

## 3.2 Qualification of CR and DR Techniques for Reliable Detection of Cracks

### 3.2.1 Purpose and/or Scope

Techniques or set-ups for reliable detection of cracks shall be qualified through POD demonstration. The following requirements for limited validation testing would qualify a reliably detectable crack a/t ratio or flaw size if authorized by RFCB.

### 3.2.2 Limited Qualification Testing of CR and DR for Crack Detection.

a. Crack size is given as a/t ratio.
b. Shall use low and high thickness material samples with fatigue cracks with targeted size,
c. Shall use samples with all worst crack locations and orientations from point of inspectability and damage tolerance analysis.
d. Shall use representative quality indicators (RQI per ASTM E1817) that simulate part material and geometry as test samples except when flat plate samples can be substituted for RQI's, provided flat plate samples can adequately simulate part and x-ray set-up of the part to be inspected.
e. Shall use minimum three cracks for each flaw detectability type (i.e. low thickness, high thickness, worst locations and orientations)
f. Shall use limit angles for x-rays with cracks i.e. locate cracks at edge of inspectable region with sample orientation similar to that of the real part at the crack location.
g. Shall use optimized set-up.
h. Signal response is measured as contrast of crack indication and contrast-to-noise ratio is also measured. Noise is measured as standard deviation.
i. Set-ups shall be identical (double wall or single wall) to that of real part inspection.
j. Minimum contrast-to-noise ratio measured on RQI cracks shall be 2.5.
k. Minimum indication length of RQI crack-to-normalized image unsharpness ratio shall be 9 times geometric magnification.
l. Set-ups shall be repeatable and results shall be reproducible by demonstrating on minimum of three repeated set-ups for each shot configuration. This can be determined by monitoring crack indication signal, noise, signal-to-noise ratio and contrast-to-noise ratio on sample cracks.
m. Results shall show expected contrast and flaw detection variability between set-ups.

### 3.2.3 Qualified Set-ups

a. Set-up for thinnest and thickest sections shall be identical to or better than the qualified set-ups' total unsharpness and x-ray angle requirements.
b. Set-ups for in-between thicknesses shall have same or better total unsharpness.
c. X-ray angle for in-between thicknesses shall be enveloped by the qualified limit angles.

## 3.3 Limited Qualification Testing of X-ray CT for Detection of Volumetric Flaws and Cracks

d. X-ray CT shall use x-ray DR (ASTM E2737 and ASTM E2698) system performance monitoring methods given in Section 2.0.
e. X-ray CT using linear detector arrays shall use the resolution and density standard tests set forth in ASTM E1695 and ASTM E1935, or an alternate method approved by the cognizant engineering organization, to conduct system performance monitoring.
f. X-ray CT applications shall meet ASTM E1570 Standard Practice for Computed Tomographic (CT) Examination, ASTM E1695 Standard Test Method for Measurement of Computed Tomography (CT) System Performance, and ASTM E1935 Standard Test Method for Calibrating and Measuring CT Density.
g. CT qualification shall use demonstration of flaw detection using fatigue cracks and other simulated flaws in test samples (ASTM E1817 RQI) that simulate part for the purpose of x-ray CT.
h. The RQI's shall be tailored to the applications demonstrating necessary dimensional capability, including wall thicknesses measurement, accuracy.

i. The RQI’s shall be tailored to the applications demonstrating necessary flaw detection capability.
j. Validation of detection of volumetric flaws does not require POD level demonstration but requires flaw detectability demonstration on minimum 3 flaws of target level (i.e. cracklike flaws) and reject level (volumetric flaws) given in acceptance criteria.
k. Shall use samples with all worst flaw locations and orientations from point of x-ray CT inspectability and damage tolerance analysis.
l. Flaws shall be distributed throughout the relevant RQI to cover all inspection zones.
m. Minimum contrast-to-noise ratio on the demonstrated reject level volumetric flaws shall be 2.5.
n. Minimum reject level volumetric flaw size-to-demonstrated resolution at the flaw location ratio shall be 3.
o. Flaw detectability shall to be demonstrated on minimum 3 target level cracks in the RQI for each inspection zone where cracks are likely to occur.
p. Minimum contrast-to-noise ratio measured on target size RQI crack indications shall be 2.5.
q. Minimum ratio of length of target level crack indication from RQI to demonstrated resolution in RQI at the crack location shall be 9.
r. Qualification plan shall be prepared.
s. CT procedure shall document CT set-up and calibration set-up.
t. Qualification report shall be prepared to document, qualification plan, CT procedures, acceptance criteria, RQI description, demonstration test results and qualification process conclusion.