Directed Design of Experiments for Validating Probability of Detection Capability of NDE Systems (DOEPOD)

Edward R. Generazio
Langley Research Center, Hampton, Virginia

NASA STI Program . . . in Profile

Since its founding, NASA has been dedicated to the advancement of aeronautics and space science. The NASA scientific and technical information (STI) program plays a key part in helping NASA maintain this important role.

The NASA STI program operates under the auspices of the Agency Chief Information Officer. It collects, organizes, provides for archiving, and disseminates NASA's STI. The NASA STI program provides access to the NTRS Registered and its public interface, the NASA Technical Reports Server, thus providing one of the largest collections of aeronautical and space science STI in the world. Results are published in both non-NASA channels and by NASA in the NASA STI Report Series, which includes the following report types:

- TECHNICAL PUBLICATION. Reports of completed research or a major significant phase of research that present the results of NASA Programs and include extensive data or theoretical analysis. Includes compilations of significant scientific and technical data and information deemed to be of continuing reference value. NASA counter-part of peer-reviewed formal professional papers but has less stringent limitations on manuscript length and extent of graphic presentations.

- TECHNICAL MEMORANDUM. Scientific and technical findings that are preliminary or of specialized interest, e.g., quick release reports, working papers, and bibliographies that contain minimal annotation. Does not contain extensive analysis.

- CONTRACTOR REPORT. Scientific and technical findings by NASA-sponsored contractors and grantees.

- CONFERENCE PUBLICATION. Collected papers from scientific and technical conferences, symposia, seminars, or other meetings sponsored or co-sponsored by NASA.

- SPECIAL PUBLICATION. Scientific, technical, or historical information from NASA programs, projects, and missions, often concerned with subjects having substantial public interest.

- TECHNICAL TRANSLATION. English-language translations of foreign scientific and technical material pertinent to NASA's mission.

Specialized services also include organizing and publishing research results, distributing specialized research announcements and feeds, providing information desk and personal search support, and enabling data exchange services.

For more information about the NASA STI program, see the following:

- Access the NASA STI program home page at http://www.sti.nasa.gov

- E-mail your question to help@sti.nasa.gov

- Phone the NASA STI Information Desk at 757-864-9658

- Write to:
  NASA STI Information Desk
  Mail Stop 148
  NASA Langley Research Center
  Hampton, VA 23681-2199

Directed Design of Experiments for Validating Probability of Detection Capability of NDE Systems (DOEPOD)

Edward R. Generazio
Langley Research Center, Hampton, Virginia

March 2015

DOEPOD[1] Manual

Directed Design of Experiments for Validating Probability of Detection Capability of NDE Systems (DOEPOD)
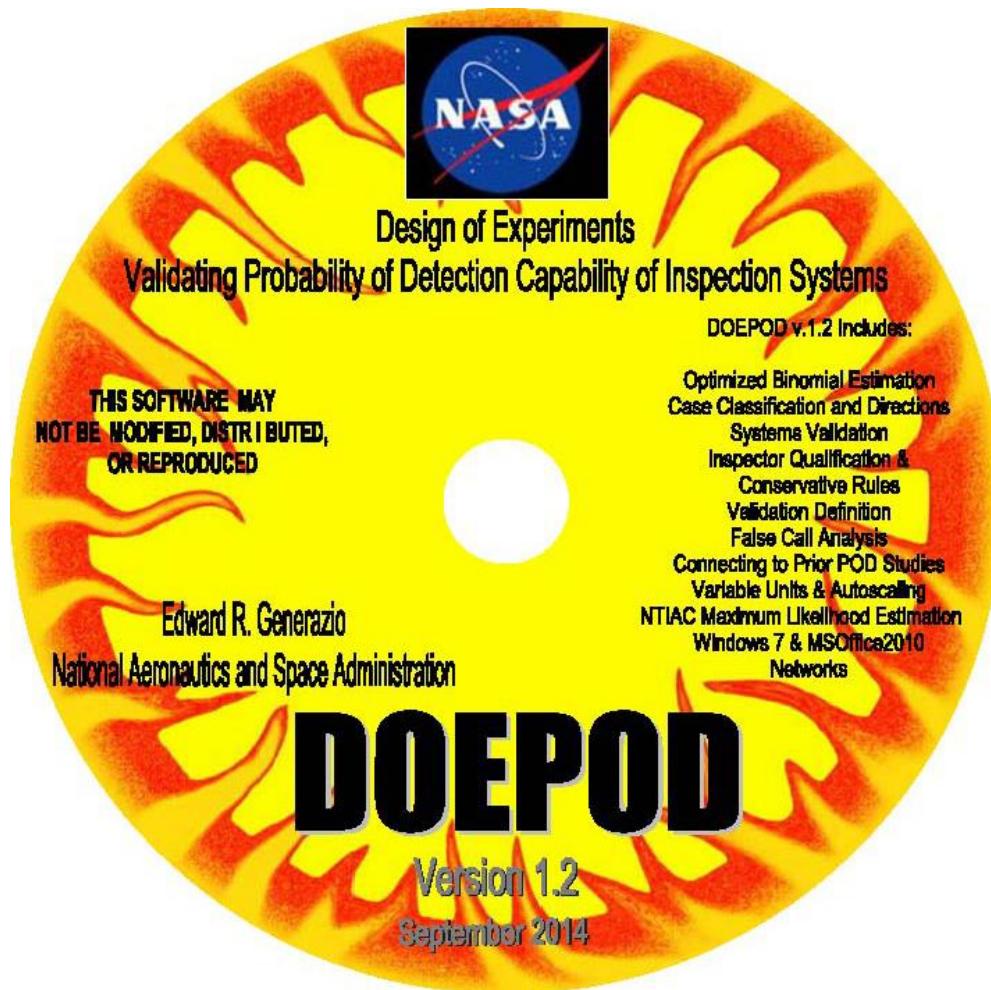
Sponsor
Office of Safety and Mission Assurance
National Aeronautics and Space Administration
Washington, DC

---

[1] Generazio, E. R. "Directed Design of Experiments for Validating Probability of Detection Capability of a Testing System" US Patent Serial Number: US 8,108,178

INDEX

Directed Design of Experiments for Validating Probability of Detection Capability of NDE Systems (DOEPOD)

DOEPOD OVERVIEW

The capability of an inspection system is established by applications of various methodologies to determine the probability of detection (POD). One accepted metric of an adequate inspection system is that there is 95% confidence that the POD is greater than 90% (90/95 POD). Design of experiments for validating probability of detection capability of nondestructive evaluation (NDE) systems (DOEPOD) is a methodology that is implemented via software to serve as a diagnostic tool providing detailed analysis of POD test data, guidance on establishing data distribution requirements, and resolving test issues. DOEPOD demands utilization of observance of occurrences. The DOEPOD capability has been developed to provide an efficient and accurate methodology that yields observed POD and confidence bounds for both Hit-Miss or signal amplitude testing. DOEPOD does not assume prescribed POD logarithmic or similar functions with assumed adequacy over a wide range of flaw sizes and inspection system technologies, so that multi-parameter curve fitting or model optimization approaches to generate a POD curve are not required. DOEPOD applications for supporting inspector qualifications is included.

DOEPOD utilizes the concept of "point estimate Probability of a Hit" (POH) at any flaw size (Generazio, 2008, 2009)[2]. That is, the number of Hits observed per set of specimens exhibiting flaws of similar characteristics (e.g., flaw lengths). The determination of estimated POH at any selected flaw size is a measured or observed quantitative value between zero and one, and knowledge of the estimated POH also yields a quantitative measure of the lower confidence bound. This process is statistically referred to as "observation of occurrences" and is distinct from use of functional forms that predict probability of detection (POD). The driving parameters of DOEPOD are the observed estimated POH and the lower confidence bounds of the observed estimated POH. Flaw size is referred to throughout the subsequent text as a "class length" for length, depth, area, etc.

The binomial distribution has been used previously for determining POD by observation of occurrences. Prior work (Yee, 1976, Rummel, 1982) used a selection of arrangements for grouping flaws of similar characteristics. Yee (1976) used smoothing optimized probability and overlapping sixty point methods, grouped by number of flaws into a class and by cumulative sums of fixed flaw size class intervals, while Rummel (1982) used fixed class widths. These binomial approaches have lead to the acceptance of using the 29 out of 29 (29/29) point estimate (Yee, 1976, Rummel, 1982, MSFC-STD-1249 method, in combination with validation that the POD is increasing with flaw size, to meet the requirements of MSFC-STD-1249 and NASA-STD-5009. DOEPOD extends work in binomial applications for POD by adding the concept of lower confidence bound maximization as the driver for establishing that there is 95% confidence that the POD is greater than 90% (90/95 POD). DOEPOD satisfies the requirement for critical applications where validation of inspection systems, individual procedures, and operators are required even when a predicted POD curve (NTIAC, 1997) is estimated. Inspection processes and procedures are to fixed and under control before applying DOEPOD analysis.

DOEPOD follows a series of defined processes to evaluate inspection data that is placed in the user friendly data template files. Details of the processes used are identified in the references at the end of the manual. During operation DOEPOD statistically evaluates the inspection data and identifies the data sets as being a specific case from a particular class of data set classes. The classes range from CASE 1 to CASE 7, referring to fully validated at a 90/95 POD level to extremely far from validation, respectively. Once this class or CASE is

---

[2] References are on page 50 of this manual.

known, DOEPOD identifies a series of ordered steps, that if pursued successfully, will lead to full validation.

In addition to validating inspection systems, DOEPOD provides support for the qualification of inspectors. DOEPOD includes the capability to evaluate false call rates for both linear and area inspection windows, and to validate the connection of DOEPOD POD results with other POD results obtained from other previous testing.

DOEPOD Update History

DOEPOD has been updated since the original beta release. The current release is Prerelease v.1.0.3 and additions include: (1) Test to validate that 90/95 POD (or greater exists) for flaw sizes greater than the 90/95 POD flaw size, X_{pod}, (2) False call warnings, (3) Inspector qualification, and (4) Use of variable units.

Beta   Is the original Beta software

Beta.2   Permits utilization of all exactly identically sized flaws for simulations. A flaw at 0.00002" is added to satisfy the requirement for at least two different flaw sizes.

Beta.3   When utilizing the F-table for determining the confidence bounds, DOEPOD uses the conservative table value for all determinations of the lower confidence bounds. No interpolating of the table values are used. This may create an inconsistency with prior estimated accepted 90/95 confidence levels for 29 out of 29, 45 out of 46, 59 out of 61, etc, sample sets. DOEPOD evaluates the conservative lower confidence bounds and compatibility with prior accepted lower confidence bounds. If these conflicts are present then the conservative lower confidence bounds obtained from the F-table are rounded up from 0.9872 to 0 .9001 to assure compatibility with prior work that may have used the less conservative F-table values. This represents an error less than 0.3% in the confidence bound and is typical when comparing table derived values.

Beta.4   Number of false calls and number of false call opportunities may be entered by integer numbers.

False call opportunities may be determined by the inspection length or area windows.

Excessive false call rate is announced as a warning.

If 90/95 X_{POD} is achieved, class lengths are also grouped by number, from large to small class lengths, in order to determine if 90/95 POD is reached at class sizes from X_{POD} to X_{XL}. This knowledge is used to support validation at larger flaw sizes when 90/95 X_{POD} has been reached, or to possibly yield a 90/95 POD at large class lengths when no further specimens are or will be available.

Maximum Likelihood Estimation (MLE) is added and executed. The MLE results are for comparison to work of others and are not used by DOEPOD to support validation of the inspection system or the qualification of inspectors.

Beta4 takes noticeably longer than Beta3 to execute due to the execution of the three additional analysis providing for MLE, large flaw validation, and false call evaluation. User options are added to inhibit large flaw validation and MLE analyses.

There is a new template TEMPLATE Beta4.xls that is used for DOEPOD Beta4. DOEPOD Beta4 is compatible with prior templates.

Prerelease
v. 1.0

The following was not being executed in prior versions. "If 90/95 X_{POD} is reached and the inspection widow is not provided, then DOEPOD will use the X_{POD} class length to determine the inspection widow." This has been corrected to execute when: "If number of false call opportunities in not known, and Xpod is reached, then the length or circular area (inspection windows) swept out by the Xpod length is used to estimate the number of false call opportunities when either the total inspection length or total inspection areas is given."

Highlighted Misses that may be present within the X_{pod} class width group are now highlighted in RED in the "Analysis Data" sheet.

Validation definition expanded to clarify: …" flaw types in the test specimen set", etc.

Flaw areas (as well as other parameters) may be used as a class length. Typical flaw areas, etc., may occur below the reserved class length number 0.00002 therefore, flaw areas need to be scaled by the user to exceed this number in order for DOEPOD to recognize these flaw areas at test samples rather than false call opportunities. Added note in manual on use of DOEPOD for analysis by other than length or depth flaw sizes, e.g., flaw area.

Hit/Miss values less than zero are considered Misses.

The graphic visualization of the estimated POH by use of a curve fitted log-odds functions provided very limited understanding. Unfortunately, the risk of utilization of un-validated math models is high, therefore this analysis and visualization is no longer supported.

False call "Warnings" explicitly stated on analysis chart. Added an "acceptable" statement when the false call upper bound indicates that the X_{POD} is only negligibly [see text] affected by the presence of false calls observed.

Added new priority requirements for sample selection and execution priorities to the operational instructions. Recommend initial survey set in order to minimize total number of samples required.

Faster execution speed is available by turning off screen updates off during processing. This is added as an option in the DOEPOD v.1.0 template.

Units labels (e.g., inch, in^2) may now be changed. This is added as an option in the DOEPOD Prerelease.v.1.0 template.

Auto scaling has been added to allow class lengths greater than 10" or any units with values greater than 10.

Test to validate that 90/95 POD (or greater exists) for flaw sizes greater than the 90/95 POD flaw size, X_{pod}, is added.

Large flaw validation is now a requirement for validating that 90/95 POD or better exist for flaws greater than the 90/95 POD class length. Large flaw validation analysis can no long be disabled in DOEPOD Prerelease.v.1.0.3

User may now indicate the maximum allowed flaw size for the large flaw validation. This prevents DOEPOD from requesting flaws sizes that exceed sample dimensions, etc.

User may now add a validated 90/95 POD flaw size obtained by other POD analyses to support large flaw validation and to connect the current DOEPOD analysis with prior 90/95 POD flaw size results.

Inspector qualification is now included. DOEPOD v.1.0.3.6 allows a broader range of large flaws to be used during inspector qualification.

There is a new template TEMPLATE Prerelease.xls that is used for DOEPOD Prerelease.v.1.0. DOEPOD Prerelease.v.1.0 is compatible with all prior templates.

Prerelease v. 1.0.3.6	Will not re-analyze files in the Analysis Folder that have already been run. Re-analyzing the same file requires that the input file name be changed. This has been done in order to maintain data integrity.
	Will not restart an analysis if the system or user aborts the DOEPOD analysis and tries to retstart DOEPOD in a mid-analysis condition. DOEPOD will restart from the original DOEPOD.xls file.
Prerelease v.1.0.3.19	Reorders identical flaw sizes so that any Misses are adjacent to the next largest flaw size in the data set. This reordering is not performed when using the original data files from the NTIAC NDE Capabilities Data Book, 1997.
	Includes additional conservative rules for inspector qualifications yielding a CONDITIONAL PASS from the presence of exceptions in false call and large flaw validation analyses, and Misses.
DOEPOD v.1.0	Released July 2009. DOEPOD v.1.0 is compatible with prior templates. TEMPLATE v.1.0.xls and TEMPLATE Prerelease.xls are equivalent.
DOEPOD v.1.2	Released September 2014. DOEPOD folder may be placed anywhere on PC including at a network location. Apple® computers without PC simulation are not supported, e.g., Apple® iMac®.

DEFINITIONS

C_L Class length, e.g., inspection parameter (length, depth, area, etc.)

C_W Class width (width of the moving class; all flaws within the range C_L to C_L - C_W, inclusively, are group together)

Hit   Flaw is detected

Miss   Flaw is not detected

MLE   Maximum Likelihood Estimate of POD using a two parameter statistical model. The MLE is included in DOEPOD as a user request for comparison. *The included method is that of the* NDE Capabilities Data Book, 3rd ed., Nov. 1997, NTIAC DB-97-02, DoD. *The use of MLE estimated POD is not recommend unless a full validation of the estimated POD is performed (see Generazio, E. R., Interrelationships Between Receiver/Relative Operating Characteristics Display, Binomial, Logit, and Bayes' Rule Probability of Detection Methodologies, NASA-TM-2014-21818, April 2014.*

Need   Add new samples to the existing specimen set in order to reach the number of samples required at the class length. Note that a single specimen may contain more than one flaw, so that "add samples" refers to "add flaws".

LCL   Lower confidence bound (value) of POH @ 95% confidence

Opt. X_{POH} Optimum X_{POH} is identified for non-survey data sets. Optimum X_{POH} is the smallest class length and largest class width at which the minimum $X_{POH} = 1$ occurs. Optimum X_{POH} may be more aggressive than optional, X_{PODopt}, or $X_{Best\ LCL}$, when the class width is constrained to the companion Optimum X_{POH} class width listed. DOEPOD does not force use of Optimum X_{POH} over X_{PODopt}. or $X_{Best\ LCL}$ Stability has not been demonstrated at Optimum X_{POH}, therefore there is an additional risk that Optimum X_{POH} can not be satisfied to reach X_{POD}

POH   Estimate of Probability of Hit (Number of Hits in Class Length/Total Number of Trials in Class Length)

POD   Probability of Detection (the true POD obtained if an infinite number of samples are used)

Signal Amplitude   Scalar amplitude output of NDE inspection system

Survey Data Sets   Survey Data Sets are data sets that have a sparce or disperse collection of samples. The moving class width optimization has identified this data set as having limited applications where the classwidth has exceeded $X_L/3$ <u>and</u> X_{POD} has not been reached. An alternate optimization of X_{POH} is used to provide guidance. The Survey Set is the recommended initial set for DOEPOD.

Survey X_{POH}	Survey X_{POH} is only identified for data sets determined to be Survey Data Sets. Survey X_{POH} is the smallest class length and largest class width at which the minimum $X_{POH} = 1$ class length occurs. Survey X_{POH} is the minimum class length at which X_{POD} may be achieved when the class width is constrained to the companion survey class width listed. Survey X_{POH} is utilized in all cases in which a Survey Set is identified by DOEPOD.
$X_{Best\ LCL}$	Class length exhibiting the maximum or "best" LCL. The best class length is determined by increasing the moving class width until a maximum LCL is obtained
X_i	Class length X at point "i"
X_L	Largest class length in entire data set
X_m	Class length near the mid-point between the largest and the smallest class lengths having no Misses
X_P	90/95 POD or greater is achieve, by grouping numbers of specimens, for the range X_P to X_L. X_P is only provided when X_{POD} has been identified. For inspector qualification, X_P can not be less than the largest flaw Missed. The class width of flaw set used for inspector qualification is listed as Inspector Classwidth @ Xp in the charts. The flaw sizes used for inspector qualification range from Xp to (Xp - Classwidth @ Xp).
X_{POD}	Class length at which the lower confidence bound (value) is 0.90 (90/95 POD) @ 95% confidence.
$X_{POH=1}$, X_{POH}	Class length where there are no Misses above this class length, and POH = 1 above this class length.
X_{PODopt}	Optional existing smaller class length where X_{POD} may also be achieved if additional samples are added and Hits are identified.
X_S	Smallest class length in the data set
UCL	Upper confidence bound (value) of the false call rate @ 95% confidence
**Validated	90/95 POD has been reached at a classlength, X_{POD}. In order to achieve 90/95 POD for the class length range between X_{POD} and the largest class length in the data set, X_L, inclusively, validation at a classlength near the mid-point and largest classlength is required[ξ]. If, in addition, there exists a class length, X_P, where 90/95 POD or greater exits for all class lengths in the range X_P to X_L, and $X_P = X_{POD}$, and there is a sufficient number and adequate range and distribution of classlengths greater than X_{POD}, then the validation extends from X_{POD} to X_L. When this occurs, validation at a classlength near the mid-point and largest classlength is satisfied. [ξ]WARNING: There are inspection

systems that exhibit an oscillating or non-uniform POD. For example when the flaws are greater than the eddy current footprint, when large flaws are loaded to closure, or when the physics of the inspection processes changes modes over the flaw size range of interest. If flaws in these ranges or conditions are to be detected with a 90/95 POD, then samples in these ranges need to be included. When multiple base parameters are combined, e.g., (length)x(width) = area, and the combine parameter (e.g., area) is used as the class length, then 90/95 POD is only valid if the inspection technology has been validated to <u>quantitatively measure</u> each of the base parameters, or if the inspection technology is validated to <u>quantitatively measure the new combine parameter</u>. When all CASE 1 or CASE 1+ requirements are met, and the above warnings have been evaluated and the upper confidence bound of the false call rate is not excessive, then the inspection system is validated between X_{POD} and the largest class length X_L for the flaw types, materials, and structure of the test specimen set. Validated is defined here to be: "This confidence bound procedure has a probability of at least 0.95 to give a lower bound for the 90% POD point that exceeds true (unknown) 90% POD point. This is referred to as 90/95 POD, and for larger flaws in the evaluation range 90/95 POD is met or exceeded. THIS SOFTWARE AND ANY ACCOMPANYING DOCUMENTATION IS RELEASED "AS IS". THE U.S. GOVERNMENT MAKES NO WARRANTY OF ANY KIND, EXPRESSED OR IMPLIED, INCLUDING, WITHOUT LIMITATION, ANY WARRANTIES OF MERCHANTABILITY AND FITNESS FOR A PARTICULAR PURPOSE. IN NO EVENT WILL THE U.S. GOVERNMENT BE LIABLE FOR ANY DAMAGES, INCLUDING ANY LOST PROFITS, LOST SAVINGS, OR OTHER INCIDENTAL OR CONSEQUENTIAL DAMAGES ARISING OUT OF THE USE, OR INABILITY TO USE THIS SOFTWARE OR ANY ACCOMPANYING DOCUMENTATION, EVEN IF INFORMED IN ADVANCE OF THE POSSIBILITY OF SUCH DAMAGES. THIS SOFTWARE MAY NOT BE MODIFIED, DISTRIBUTED, OR REPRODUCED.

DOEPOD CASE EXAMPLES

DOEPOD VALIDATION DYNAMICS

When X_{pod} is identified, DOEPOD further attempts to validate that 90/95 POD (or greater) exists for class lengths greater than X_{pod}. (See section on VALIDATION AT LARGE CLASS LENGTHS). DOEPOD establishes this large flaw validation in two steps. (1) If validation also exists for class lengths greater than X_{pod}, then the validation range is extended to range from $X_{pod} = X_p$ to X_L. The validation range is indicated by a shaded horizontal line (purple) extending from $X_{pod} = X_P$ to X_L. If $X_{pod} < X_p$ then, there is a validation gap. The validation range is indicated by a shaded horizontal line (purple) in subsequent figures extending from X_P to X_L. Groupings of class lengths that exhibited 90/95 POD or greater are shown as individual points on the shaded line. (2) When X_{pod} is identified, DOEPOD has an additional requirement that there be 25 unique class lengths uniformly spaced in size from X_{pod} to X_L.

In order to allow comparison with other POD methodologies, a predicted POD and it's 95% lower confidence bound are available by using the maximum likelihood estimation method (MLE) (NTIAC, 1997). MLE of parameters α (alpha) and β (beta) for the math model (NTIAC, 1997) are shown on the MLE worksheet as "new alpha" and "new beta". MLE results may not be available or valid due to a variety of issues including non-convergence and inadequacy of the MLE mathematical model for NDE systems. The MLE of the predicted POD is provided solely for comparison and is not used in the DOEPOD analysis. Please see warnings[3] concerning adequacy of math models used in MLE methodologies. **Use of MLE POD methods for fracture critical POD inspection demonstrations is not recommended due to the lack of validated NDE math models used in MLE.**

DOEPOD CASE EXAMPLES FOR SYSTEMS VALIDATION

DOEPOD classifies the POD data as being one of seven different cases. The cases are identified as CASE 1, 2, 4, 5, 6, 7, and Survey Data sets. During the development of DOEPOD, the number of unique cases was not known, and CASE 0 (all Hits) and CASE 3 (multiple flaws sizes where 90/95 POD is observed for a fixed class width) are now included in CASE 1 and 2, respectively.

CASE 1 is the only case exhibiting full validation when false calls analysis results are acceptable. CASE 1 has three sub-cases (not shown), CASE 1+, CASE 1#, and CASE 1* that indicate specific reasons why the full validation CASE 1 has not occurred. The differences in the cases are highlighted in Table C.

CASE 1 is the best case and is shown in Figure 6a. There is an adequate distribution of flaws at X_{pod} and there is a sufficient number of well distributed large flaws above the X_{pod} flaw size. 90/95 POD is reached at a class length, X_{pod}, and there are Misses only below X_{pod}. 90/95 POD validation from X_{pod} to largest flaw, X_L, is demonstrated when any false call warnings are addressed.

---

[3] See the *Comparison Between the Observed POD from DOEPOD and the Predicted POD from the Maximum Likelihood Estimation (MLE) Method* section in the Design of Experiments for Validating Probability of Detection Capability of NDE Systems (DOEPOD) and for Qualification of Inspectors and Validating Design of Experiments for Determining Probability of Detection Capability (DOEPOD) at the end of this manual.
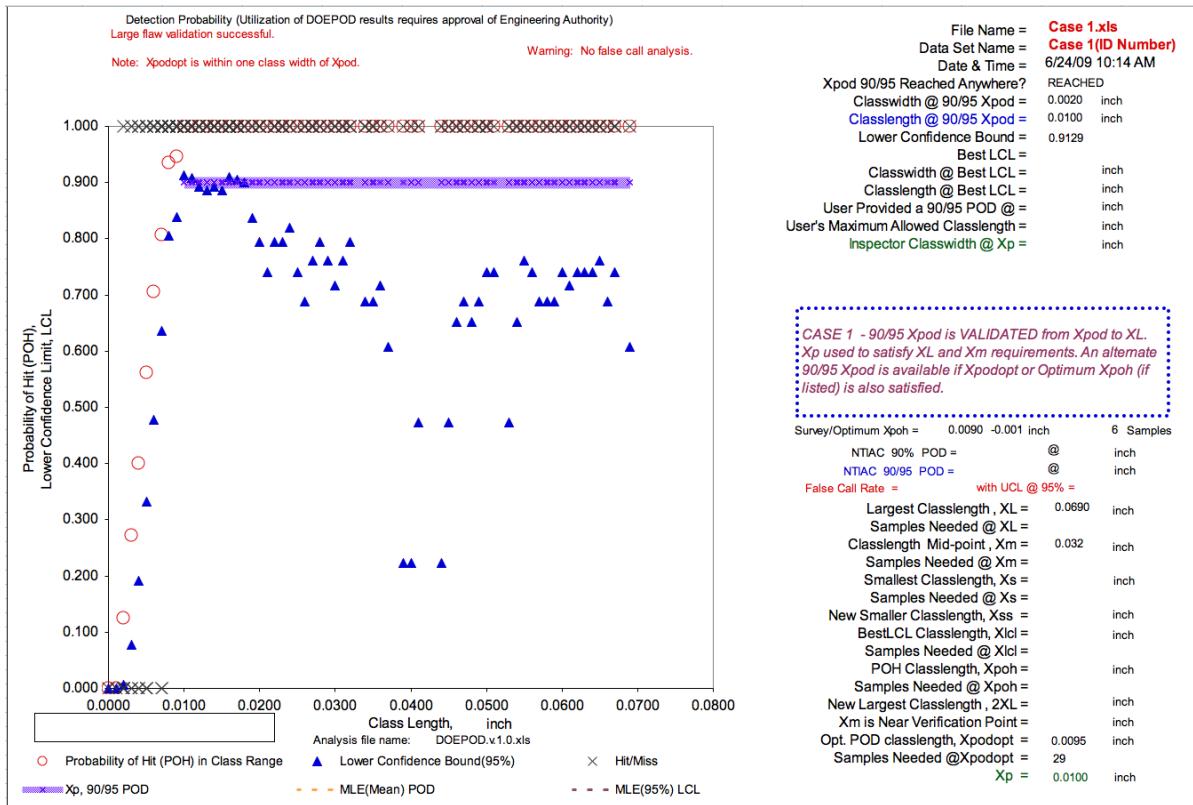
FIGURE 6a. CASE 1 example of DOEPOD analysis

CASE 1+ is the next best case and is shown in Figure 6b. There is an adequate distribution of flaws at X_{pod} and there is a sufficient number of well distributed large flaws above the X_{pod} flaw size. 90/95 POD at X_{pod} is reached at a class length, and there are Misses above X_{pod} that need to be explained and resolved. Any false call warnings need to be addressed before POD validation from either X_{pod} or X_p to largest flaw, X_L is demonstrated. If X_p is greater than X_{pod}, then there is a validation gap.
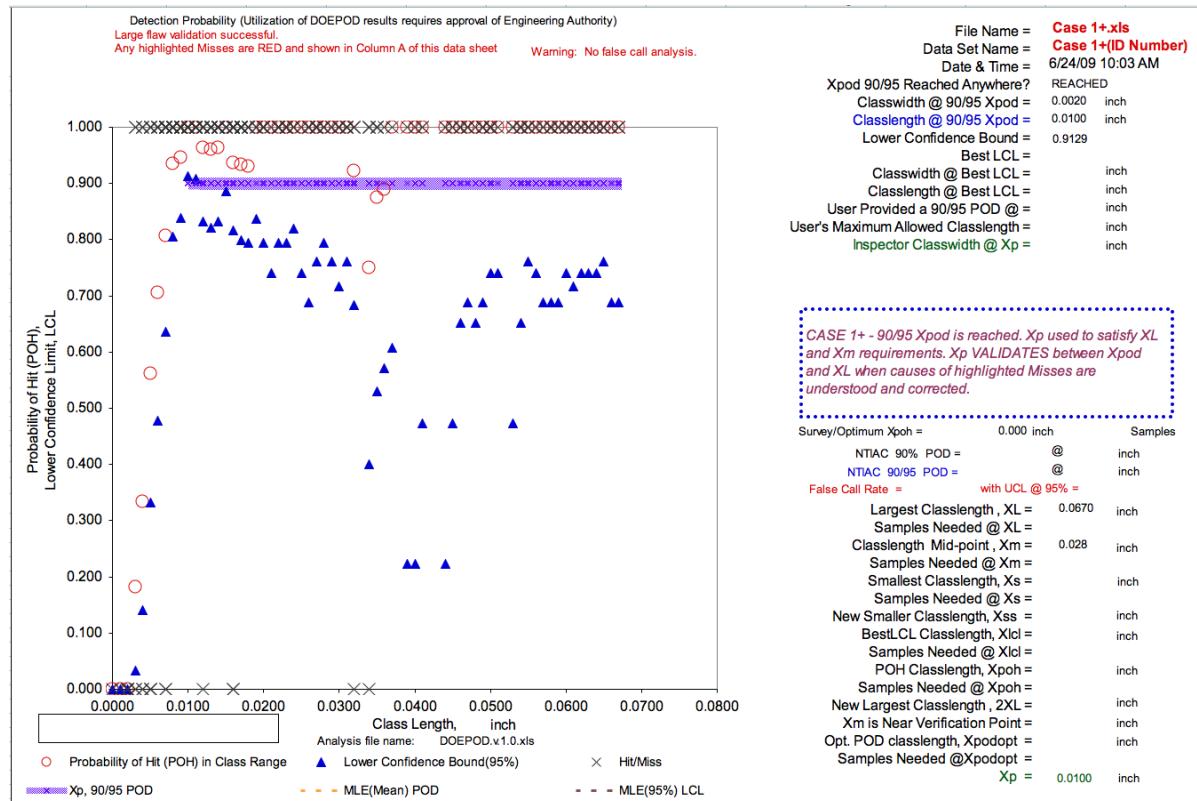
FIGURE 6b. CASE 1+ example of DOEPOD analysis

An explanation and resolution of all Misses above X_P is required. Class lengths exhibiting Misses that require explanation are highlighted in red in column A of the "Analysis Data" worksheet. The specific companion sample identification numbers for these Misses are listed as "Explain Miss : Sample ID = "of the "Analysis Data" worksheet in column I starting in row 64. An example output is shown below:

	A	B
	Crack size, length, etc	Hit = 100; Miss =0
	0.0120	0
	0.0160	0
	0.0160	0
	0.0320	0
	0.0340	0

◇	H	I
64	Explain Miss: Sample ID =	89
65	Explain Miss: Sample ID =	95
66	Explain Miss: Sample ID =	187
67	Explain Miss: Sample ID =	148
68	Explain Miss: Sample ID =	208

CASE 1# is shown in Figure 6c. There is an adequate distribution of flaws at X_{pod}, however, there is an insufficient number of well distributed large flaws above X_{pod} flaw size. 90/95 POD at X_{pod} is reached at a class length, and there are Misses only below X_{pod}. Further validation is still required in order to verify that the POD is actually increasing with increasing class length. The large flaw validation has failed due to a lack of sufficient number of flaws, or an inadequate spacing of the large flaw size, or inadequate range of large flaw sizes. The DOEPOD recommendations are to add the specified large flaws identified in the large flaw validation table of the "Analysis Data" worksheet in columns CE-CG, rows 1-29 (Figure 6d) that are greater than the X_{pod} flaw size.
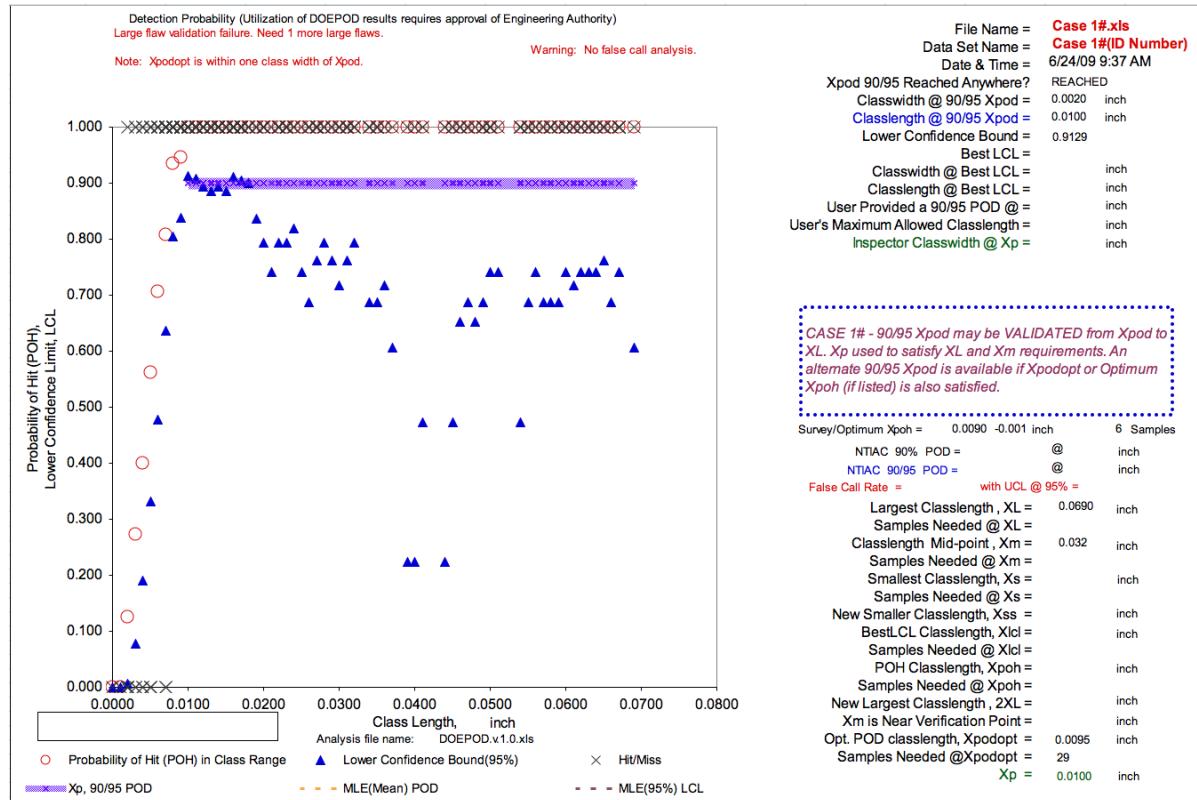
FIGURE 6c. CASE 1# example of DOEPOD analysis

CE	CF	CG
LARGE FLAW VALIDATION*		
NEEDED LARGE FLAWS =		**1**
Flaw Size Must be Greater Than:	**Flaw Size Must be Less Than or Equal to:**	**Available Flaw Sizes**
inch	inch	inch
0.0112	0.0135	0.0120
0.0135	0.0159	0.0140
0.0159	0.0183	0.0160
0.0183	0.0206	0.0190
0.0206	0.0230	0.0210
0.0230	0.0253	0.0230
0.0253	0.0277	0.0260
0.0277	0.0301	0.0280
0.0301	0.0324	0.0310
0.0324	0.0348	0.0340
0.0348	0.0371	0.0350
0.0371	0.0395	0.0390
0.0395	0.0419	0.0400
0.0419	0.0442	0.0440
0.0442	0.0466	0.0450
0.0466	0.0489	0.0470
0.0489	0.0513	0.0490
0.0513	0.0537	**NEEDED**
0.0537	0.0560	0.0540
0.0560	0.0584	0.0570
0.0584	0.0607	0.0590
0.0607	0.0631	0.0610
0.0631	0.0655	0.0640
0.0655	0.0678	0.0660
0.0678	0.0702	0.0690

FIGURE 6d. CASE 1# example of DOEPOD analysis large flaw requirement where a large flaw between 0.0513" and 0.0537" is needed to complete the validation.

CASE 2 is the most interesting case and is shown in the Figure 7 and 8. There is an adequate distribution of flaws at X_{pod} however, there are too many Misses above X_{pod}. In this case, 90/95 POD is reached at a class length, X_{pod}. There are Misses below X_{pod} and excessive Misses above X_{pod}. The number of flaws with sizes greater than X_{pod} needs to be increased. Therefore, the 90/95 POD at X_{pod} can not be accepted as a validation flaw size. The term excessive is used here since the binomial analysis number of flaws yields a Best LCL less than 0.90. Since excessive Misses exist at class lengths, X_i, above X_{pod}, then these greater lengths need to be validated by adding more test data.

The DOEPOD recommendations are listed as two options that may be executed to establish an acceptable and generally larger 90/95 POD flaw size. Successful execution of the recommendations will transition this CASE 2 to CASE 1. Option 1 is to add flaws of class length X_i where POH<1 (Figure 8, TABLE A). Starting from largest class length, X_i, and work toward small class lengths until reaching an new acceptable larger X_{pod} or reaching X_{pod}. Option 2 is to add flaws of class length X_i where POH=1 (Figure 8, TABLE B, below), and accept a larger X_{pod} class length at the X_i selected. This acceptance is valid as long as any class lengths larger than the new X_{pod} class length where POH<1 are shown [via Option 1 above] to be at 90/95 POD or greater. Acceptance of a larger X_{pod} is not necessarily the ultimate X_{pod} capability of the inspection system, but rather the current demonstrated capability of the inspection system. It is also important to recognize that by introducing additional data an acceptable or larger X_{pod} may never be obtained. In summary, the initial DOEPOD recommendations for CASE 2 are to satisfy the smallest X_{pod} in Table B that is greater than the largest X_{pod} in Table A, and/or the largest X_{pod} in Table A.

FIGURE 7. CASE 2 example of DOEPOD analysis
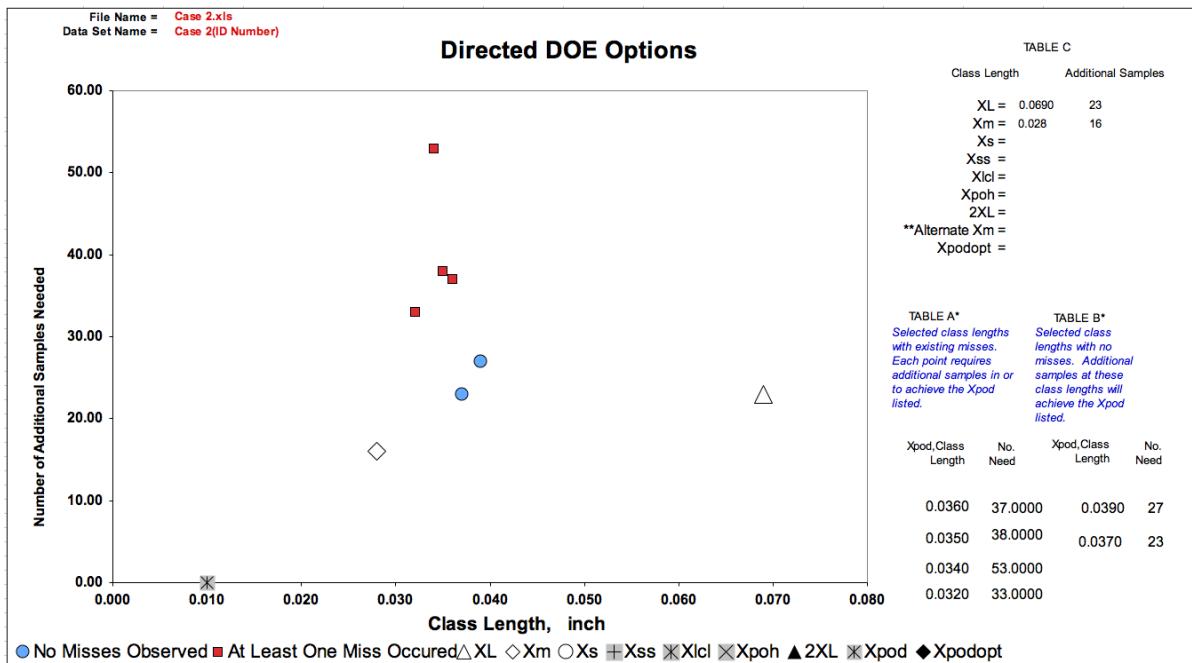
FIGURE 8. CASE 2 example of DOEPOD analysis recommendations

X_L and X_M sample requirements are shown for historical record. This is for information only in CASE 2, and these values may change when the above recommendations are executed.

CASE 2 will be automatically upgraded to CASE 1* if X_P exists then the large flaw validation *by number of flaw sizes* has occurred. However, large flaw validation *by distribution of large flaw sizes* has not occurred. Validation by both **number** and **distribution** of large flaws is required to complete the validation. An example of CASE 1* is shown in Figure 9. When CASE 2 is upgraded to CASE 1*, the table requirements are no longer necessary and are deleted. Further validation is still required in order to verify that the POD is actually increasing with increasing class length. The large flaw validation has failed due to a lack of sufficient number of flaws, or an inadequate spacing of the large flaw size, or the large flaw size range in inadequate. For CASE 1* there is an adequate distribution of flaws at X_{pod}, however, there is an insufficient number of well distributed large flaws above X_{pod} flaw size. If X_p is greater than X_{pod}, then there is currently a validation gap.

An explanation and resolution of all Misses above X_P is required. Class lengths exhibiting Misses that require explanation are highlighted in red in column A of the "Analysis Data" worksheet, and the specific companion sample identification numbers for these Misses are listed as "Explain Miss : Sample ID = "of the "Analysis Data" worksheet in column I starting in row 64. An example output for the analysis shown in Figure 9.

A	B
Crack size, length, etc	**Hit = 100; Miss =0**
0.0120	0
0.0160	0
0.0160	0
0.0320	0
0.0340	0

◇	H	I
64	Explain Miss: Sample ID =	89
65	Explain Miss: Sample ID =	95
66	Explain Miss: Sample ID =	187
67	Explain Miss: Sample ID =	148
68	Explain Miss: Sample ID =	208

FIGURE 9.   CASE 1* example of DOEPOD analysis

CASE 4 is similar to CASE 1 except that 90/95 POD at X_{pod} is not reached anywhere as shown in Figure 10. There is an inadequate number of flaws with similar sizes, therefore, the number of flaws needs to be increased. This is a well behaved data set as defined by the absence of Misses above $X_{Best\ LCL.}$ The best lower confidence bound, Best LCL, is below 0.9 for the best class width group. There are no Misses at or greater than the $X_{Best\ LCL}$ class length, or within the class width group exhibiting the best LCL, $X_{Best\ LCL}$.

The DOEPOD recommendations are to add flaws of $X_{Best\ LCL}$ or $X_{POH=1}$ in class length in order to achieve 90/95 X_{POD} at $X_{Best\ LCL.}$ or X_{POH}, respectively. The class width for added samples at is listed as Classlength@Best LCL in Figure 10. $X_{Best\ LCL}$ may equal X_L or $X_{POH=1}$ so that the number of samples listed at this class length are redundantly the same and only one set of samples is needed. There is also a more aggressive option that may be executed. If Optimum X_{POH} $<X_{POH=1}$ then the user may add samples at Optimum X_{POH} rather than at $X_{POH=1.}$ This option identifies the unique class width and class length, Optimum $X_{POH,}$ for which there are no Misses above Optimum X_{POH} for the class width identified. This example shows Optimum X_{POH} to have a class length of 0.0976" with a class width of 0.004". This is listed as more aggressive since the lower confidence bound at this class length is very low due to the limited number of samples in the class width, and 28 additional samples will be needed at Optimum $X_{POH.}$
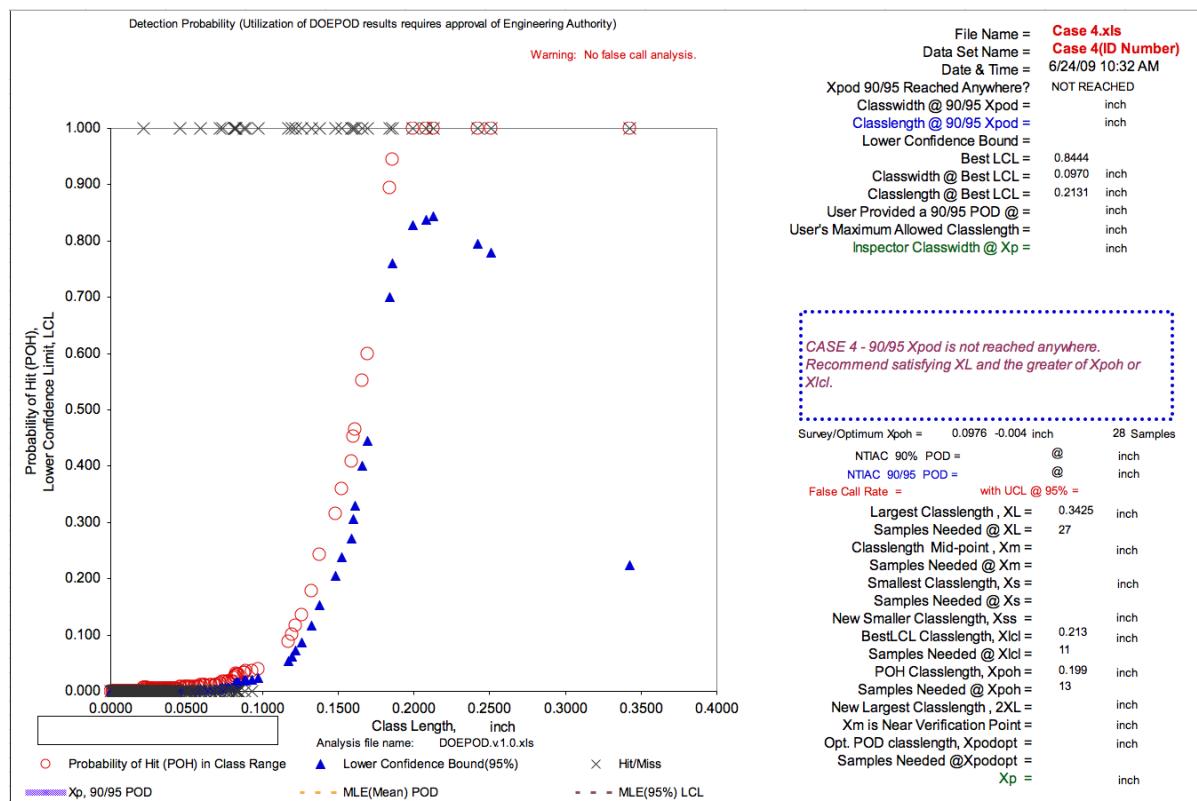
FIGURE 10. CASE 4 example of DOEPOD analysis

CASE 5 is similar to CASE 2 except 90/95 POD at X_{pod} is not reached anywhere as shown in Figure 11. The POH is well behaved for flaw sizes at and above $X_{POH=1}$, therefore, the number of flaws with sizes at X_{POH} needs to be increased. There is an inadequate number of flaws at $X_{Best\ LCL}$ and there are misses above $X_{Best\ LCL}$. There are Misses at or greater than the class length $X_{Best\ LCL}$ or within the $X_{Best\ LCL}$ class width group. There exists a class length, $X_{POH=1}$, above which there are no Misses. There are no Misses for class lengths equal to greater than $X_L/3$ (i.e., $X_{POH=1} \leq X_L/3$). $X_{POH=1} \leq X_L/3$ so that POH is not fluctuating at larger class lengths.

DOEPOD recommendations are to use $X_{POH=1}$ as the trial X_{pod} by adding flaws at $X_{POH=1}$.
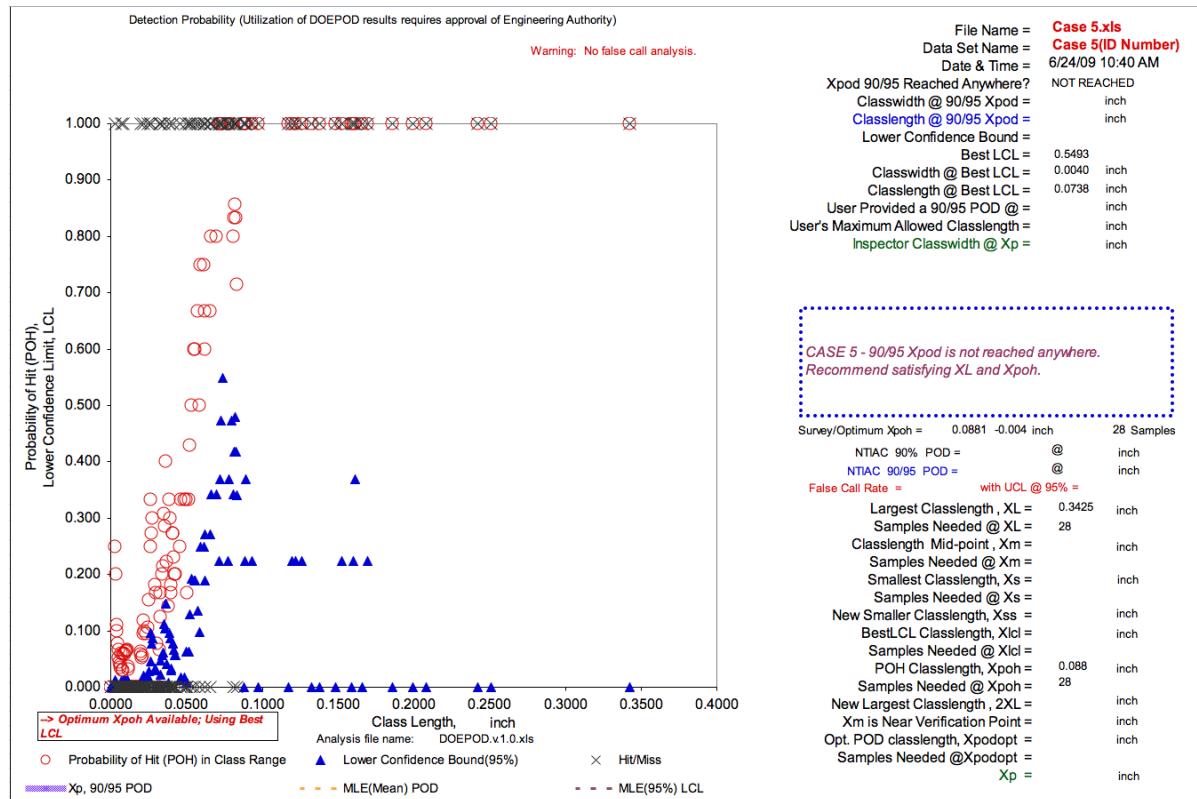
FIGURE 11. CASE 5 example of DOEPOD analysis

CASE 6 is similar to CASE 5, 90/95 POD at X_{POD} is not reached anywhere as shown in Figure 12. The POH is fluctuating throughout a considerable range of flaw sizes used, therefore, the range of flaw sizes needs to be increased. The Best LCL is below 0.9 for the best class width group. There are Misses at $X_{Best\ LCL}$ or within the $X_{Best\ LCL}$ class width group or at class lengths greater than class length $X_{Best\ LCL}$. There exists a class length, $X_{POH=1}$, above which there are no Misses. There are are Misses for class lengths greater than $X_L /3$ (i.e., $X_{POH=1} > X_L /3$). $X_{POH=1} > X_L /3$ so that POH is fluctuating at larger flaw sizes. Since POH is fluctuating at large class lengths, there is a need to expand current range of flaw sizes.

The DOEPOD recommendations are to add flaws with class lengths of $2X_L$ or greater, and add flaws at $X_{POH=1}$.
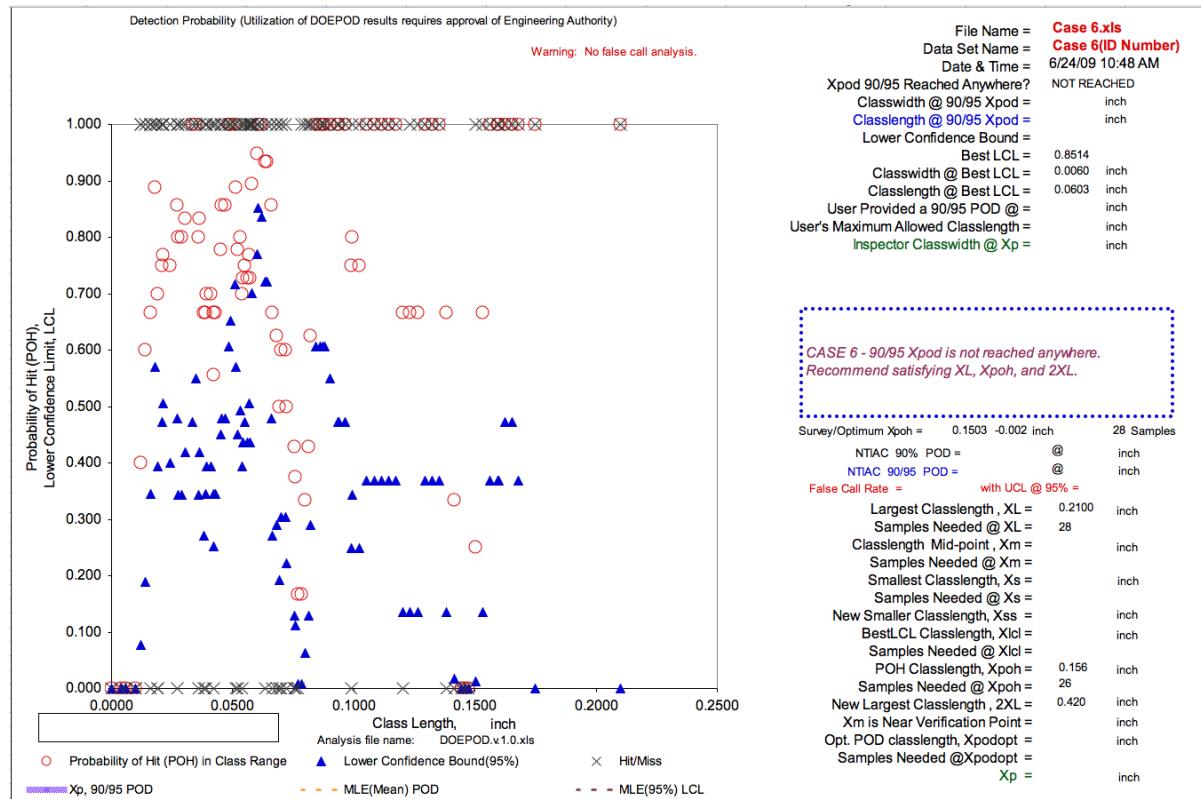
FIGURE 12. CASE 6 example of DOEPOD analysis

CASE 7 is similar to CASE 6, 90/95 POD at X_{POD} is not reached anywhere as shown in Figure 13. The POH is fluctuating throughout the entire range of flaw sizes used, therefore, the range of flaw sizes needs to be increased. The Best LCL is below 0.9 for the best class width group. There are Misses at $X_{Best\ LCL}$ or within the $X_{Best\ LCL}$ class width group or at class lengths greater than class length $X_{Best\ LCL}$. There does <u>not</u> exist a class length, $X_{POH=1}$, above which there are no Misses. POH is fluctuating or there may be no Hits anywhere.

DOEPOD recommendations are that inspection system may not be appropriate for meeting inspection criteria, or there is a need to expand current range of X_L by adding 29 new samples with class lengths of $2X_L$ or greater.
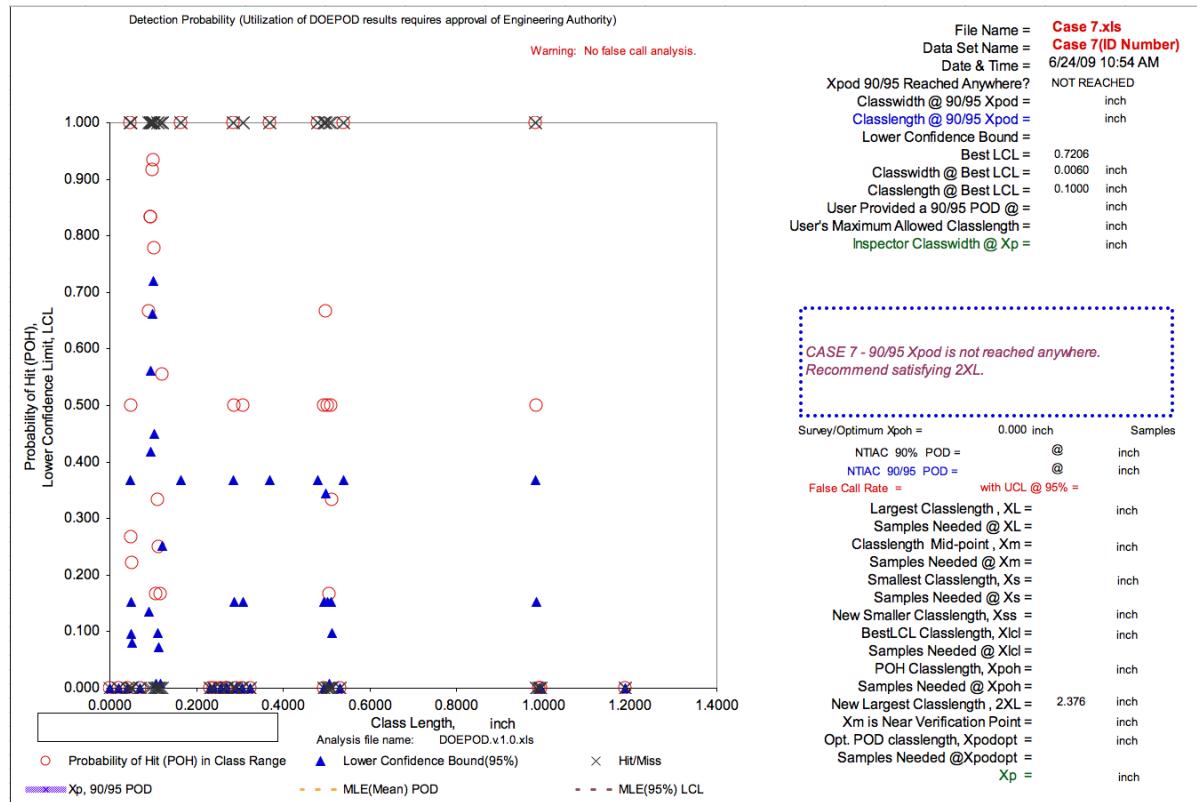
FIGURE 13. CASE 7 example of DOEPOD analysis

DOEPOD serves as a tool for optimizing the flaw size distribution requirements when analyzing Survey Data Sets. DOEPOD identifies Survey Data Sets when there is an insufficient number of flaws for unconstrained class width optimization as shown in Figure 14. This occurs when the optimized class width exceeds 1/3 X_L <u>and</u> 90/95 POD at X_{pod} has not been reached. The class width optimization has determined that there is a survey class width for which the smallest $X_{POH=1}$ class length is identified. In survey data sets the optimization procedure that maximizes LCL by increasing class width is automatically superceded. Here, $X_{Best\ LCL}$ is identified for survey data sets by determining the maximum C_W at X_{POH} for which there are no Misses within the grouping.

DOEPOD recommendations are to add flaws in the range X_{POH} to $X_{POH} - C_W$, inclusively. The Survey X_{POH} class length and class width are identified on the charts as Survey/Optimum X_{POH}. For example, the listing:

Survey/Optimum X_{POH} = 0.4600 – 0.039 inch (need 28 samples)

indicates that a class width of 0.039" is used, and the Survey or (Optimum X_{POH}) occurs at 0.4600", and that 28 additional flaws may be added in order to attempt to achieve X_{POD} at that class length. The added flaws should have flaw sizes that range anywhere between 0.4600" and 0.4210", inclusively.
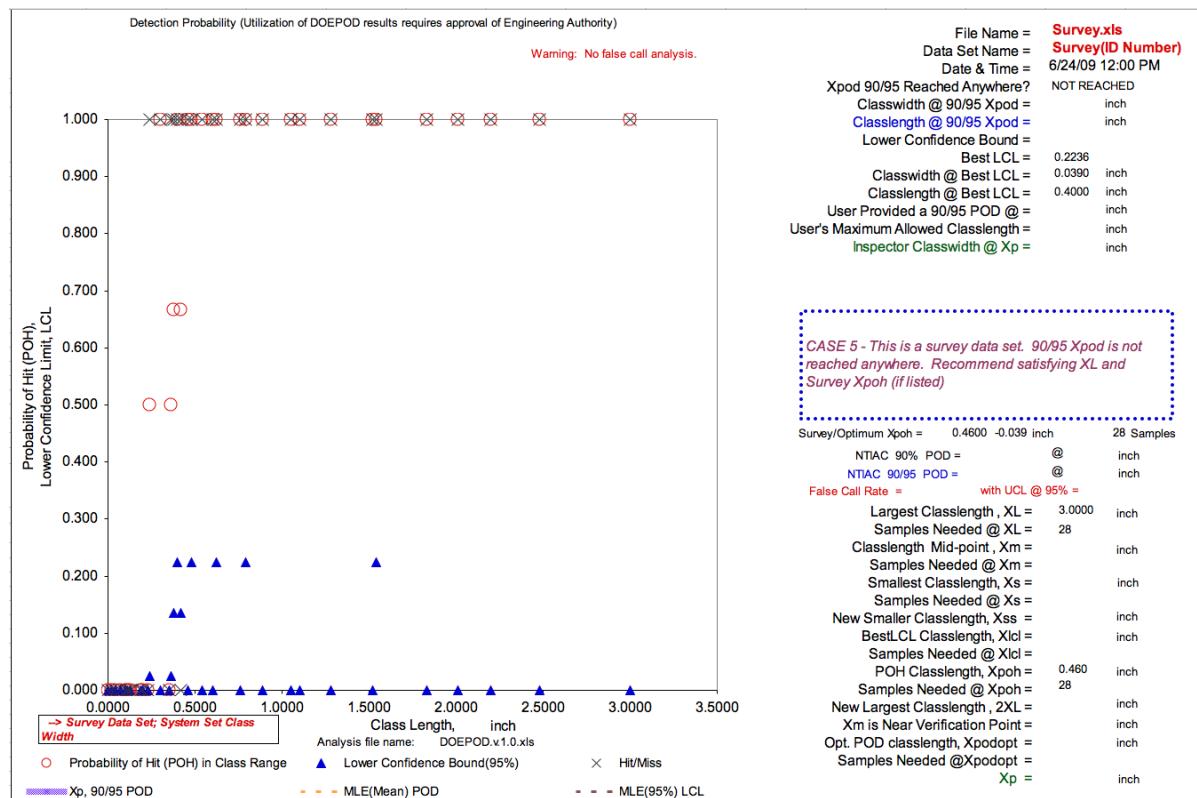
FIGURE 14. Survey Case example of DOEPOD analysis

DOEPOD Analysis Summary and Recommendations for all cases are shown in Table C.

|	Is 90/95 POD at X_{pod} reached? (i.e., lower confidence bound, X_{Best_LCL}, is equal to or greater than 0.9)	Does $X_{POH=1}$ exist?	Is POH = 1 everywhere greater than X_{Best_LCL}?	Is X_{poh} less than or equal to $X/3$?	Large flaw validation complete?	DOEPOD Analysis Summary and Recommendations
CASE 1	YES	YES	YES	NA	YES	90/95 POD at X_{pod} has been reached. Actions: Address any false call warnings.
CASE 1+	YES	YES	NO	NA	YES	90/95 POD at X_{pod} has been reached. Actions: Misses above Xpod need to be explained and resolved. Address any false call warnings.
CASE 1#	YES	YES	YES	NA	NO	90/95 POD at X_{pod} has been reached. Actions: Further validation at flaw sizes greater than Xpod is required. Add large flaws. Address any false call warnings.
CASE 1*	YES	YES	NO	NA	NO	90/95 POD at X_{pod} has been reached. Actions: Further validation at flaw sizes greater than Xpod is required. Add large flaws. Misses above Xpod need to be explained and resolved. Address any false call warnings.
CASE 2	YES	YES	NO	NA	NA	90/95 POD at X_{pod} has been reached, however, there are an excessive number Misses above X_{pod}. Actions: Additional validation at identified flaw sizes is required. Add flaws per instructions.
CASE 4	NO	YES	YES	NA	NA	90/95 POD at X_{pod} has not been reached. Actions: Increase number of flaws at $X_{POH=1}$ or X_{Best_LCL}.
CASE 5	NO	YES	NO	YES	NA	90/95 POD at X_{pod} has not been reached and there are Misses above X_{Best_LCL}. Actions: Increase the number of flaws at $X_{POH=1}$.
CASE 6	NO	YES	NO	NO	NA	90/95 POD at X_{pod} has not been reached. The POH is fluctuating above X_{Best_LCL} and X_{poh} is greater than $X/3$. The inspection system is unstable for the flaw size range analyzed. Actions: Increase the flaw size range by a factor of two.
CASE 7	NO	NO	NA	NA	NA	90/95 POD at X_{pod} has not been reached. The inspection system is unstable for the entire flaw size range analyzed. Actions: The inspection system may not be appropriate or increase the flaw size range by a factor of two.
SURVEY CASES	NO	YES	NA	NA	NA	The optimized class width exceeds 1/3 XL and X_{pod} has not been reached. The class width optimization has determined that there is a class width for which the smallest $X_{POH=1}$ class length is identified. Actions: Add flaws at Survey/Optimum X_{POH}

= YES   = NO   NA = NOT APPLICABLE

Table C. Summary of all CASES and actions.

DOEPOD FOR INSPECTOR QUALIFICATION

DOEPOD analysis may be applied to evaluate the capability of inspectors. This is similar to validating that the inspection system meets the inspection requirements, except that the requirement for validation at large flaws is not strictly required as it is already included in the systems validation. The 90/95 POD capability of the inspection system must be demonstrated first, by obtaining CASE 1 with inspection processes and procedures fixed and under control, before asking inspectors to demonstrate their inspection capability using the inspection system. There are situations where critically large flaws have been missed by inspectors even though the inspection system had a demonstrate capability to finding large flaws. Since human factors plays an important and possibly large role here, it is good engineering practice to include large flaws in the sample set when performing inspector qualification. It is recommended, as a minimum, that 29 unique flaws at the target flaw size, X_{pod}, and 5 equally spaced unique larger flaws, along with a minimum of 84 false call opportunity sites, be included in all inspector qualification tests. The largest flaw size is to be the smaller of the largest flaw expected in the component or 3 times the target flaw size, X_{pod}. Ideally, the number of large flaws is to be 25 in order to strictly assure that the inspector is capable of demonstrating 90/95 POD over the entire expected flaw size range. A minimum of five flaws is reasonably set by experience of current industry qualification test practices, and is solely established by good engineering judgment. POD testing for qualifying inspectors is only one element of inspector qualification. Other elements included in inspector qualification are calibration, adherence to procedures, visual acuity, etc.

There are often specimen constraints that are imposed when there are insufficient number or range of large flaws, or when the large flaws are poorly distributed in size. DOEPOD addresses these constraints by identifying a CONDITIONAL PASS and lists specific conditions when there are insufficient number or range of large flaws, or the large flaws are poorly distributed in size. In this manner, the examiner is required to explain and justify the large flaw results.

There are concerns when implementing false call analysis for inspector qualifications. An issue arises when there is a large amount of false call opportunities as would be available for a area inspections such as penetrant inspections. In this scenario, the inspector could have many false calls while yielding an upper confidence bound of the false call rate that is acceptable. Even though, statistically, the inspector's false call rate is acceptable and does not affect the POD results, the presence of many false calls is a cause of concern. Specifically, the test specimens are generally free of non-flaw blemishes, such as scratches, so that false calls are expected to be small.

During an inspection of failure critical components, the focus on the worse case of missing a critical flaw so that all indications regardless of size are to be noted. For example, if the inspection drawing note calls out a 90/95 POD flaw size of 0.150", then the inspector does not ignore the 0.010" flaw found. The accept or reject decision is not in the hands of the inspector, and therefore a false call is preferred over missing a critical flaw in a failure critical component. **It is better to disposition a false call than to Miss a critical flaw. The presence of a critical flaw can not be tolerated.**

The discussion above support the acceptance of false calls during inspector qualification to some extent, but the presence of many false calls is a warning.

DOEPOD resolves this issue by identifying a CONDITIONAL PASS and lists specific conditions when one or more false calls is observed, or if the upper confidence bound of the false call rate is to high, or if there are no false call opportunities. In this manner, the examiner is required to explain and justify the false call results.

For system validations, 90/95 POD at X_{pod} is often less than the largest flaw Missed, as this is statistically acceptable. However, for inspector qualifications the conservative rule, "*Qualification flaw size can not exceed the largest Miss observed.*" has historically been applied when determining the inspector's qualification 90/95 POD flaw size. The overall implementation of this rule requires that DOEPOD update X_p to a more conservative value that satisfies this conservative rule for inspector qualifications. X_p (and not X_{pod}) is the qualification 90/95 POD flaw size for the inspector.

An example of PASS and CONDITIONAL PASS inspector tests are shown in Figures 15 and 16, respectively. The CONDITIONAL PASS may only be accepted when the miss at $X_p = 0.085$" is explained and resolved. Inspector Classwidth @ X_p identifies the range of flaw sizes used to identify X_p. That is, the range includes all the flaws from (X_p - Inspector Classwidth) to X_p, inclusively.

FIGURE 15. Example of inspector PASS

FIGURE 16. Example of inspector CONDITIONAL PASS requiring further explanation before pass is accepted.   The miss to be explained is at 0.085".

[4]REQUIREMENTS

- Inspection processes are to be under control, fixed, reproducible, and repeatable.

- The minimum number of flawed sites for systems validation is 29 flaws of the target flaw size and 25 flaws larger than the target flaw size.

- The minimum number of flawed sites for inspector qualification is 29 flaws of the target flaw size and 5 flaws larger than the target flaw size.

- Test samples or inspection sites with no flaws present are to be included for determination of false call rate and the upper confidence bound of the false call rate at 95% confidence. There are two methods for including false calls (see the FALSE CALL ANALYSIS section). There are to be a minimum of 84 of unflawed specimens or unflawed inspection sites during any test. This is a minimum requirement that is coupled to the false call rate and it's upper confidence bound. If there are no false call opportunities listed, then a false call analysis is not performed, and the DOEPOD results are subject to this uncertainty, and validation and qualification is not assured.

- Multiple inspection processes may be used on the same set of test samples with the requirement that DOEPOD is to be executed for each process separately. When multiple inspection processes or systems are used, the resulting directed sample requirements may be overlapping. In this situation, the user is to keep the non-overlapping directed sample requirements applied to the appropriate inspection process, while utilizing overlapping directed sample requirements for the multiple processes in order to minimize the number of generated test samples.

- Class lengths (e.g., flaw lengths) must be greater than 0.00002 (any units). DOEPOD varies the class width in 0.001 increments from 0.001 to 0.100 and then varies class width in 0.1 increments for larger class widths. Any flaws with class lengths less than 0.001 are grouped and assigned the class length of 0.001.   The uncertainty in the optimized class width is $\pm$ 0.0005 for class widths of 0.1 and below, and the uncertainty for optimized class widths above 0.1" is $\pm$ 0.05. If the data set primarily contains flaws greater than 0.5, then the user may want to rescale (to reduce) all flaw

---

sizes by a factor of 10 in order to obtain a better resolution. DOEPOD checks the maximum class length entered by the user. If the maximum class length exceeds 10, independent of the units used, DOEPOD attempts to rescale the data downward so that the class lengths are in the range 0.001 – 10. Any flaws with class lengths less than 0.001 are grouped and assigned the class length of 0.001.   Upward rescaling is not attempted by DOEPOD, and if upward rescaling is required, the user must pre-scale the data before entry. Any pre-scaling, upward or downward, done by the user may be recorded by the user using the units section of the template "Data.xls" file (see **ADVANCED DOEPOD INSTRUCTIONS)**. If the user performs pre-scaling and a false call analysis is to be performed, the same pre-scaling must also be applied to the false call lengths (and areas) in the template "Data.xls".

- DOEPOD has the capability to label different units, such as cm, in^2, pixels, etc. The units of measure are listed by the user in the template "Data.xls" file (see **ADVANCED DOEPOD INSTRUCTIONS)**. However, DOEPOD still varies the class width in 0.001 increments (of the units listed by the user) from 0.001 to 0.100 and then varies class width in 0.1 increments (of the units listed by the user) for larger class widths.   The flaw size (class length) data must be greater than 0.00002 for any units listed by the user, and the preferred range is 0.00002 – 10, exclusively.

- A moving class that groups flaws of similar size is used to optimize the lower confidence value. This moving class and best lower confidence bound (value) optimization will be invoked if there are more than four (4) samples at different class lengths.

- The total number of unflawed and flawed sites can not exceed 1999. Also see options in the FALSE CALL ANALYSIS section for relaxing this requirement.

- Be prepared to generate, inspect, and evaluate test samples during the NDE technology capability determination.

- Flawed sites are to be added in the order of priority listed in the OPERATIONAL INSTRUCTIONS.

- Validated** 90/95 POD at X_{POD} is obtained when the user has reached and satisfied the sample requirements of either CASE 1 or CASE 1+ with highlighted Misses explained and resolved, and there are no false call warnings or large flaw validation failures.   See full definition of VALIDATED in definitions section.

SOFTWARE INSTALLATION

PC Installation Instructions
Microsoft Office Excel 2007 runs 20 times slower than prior versions, this is being remedied in Microsoft Office Excel 2010. Recommend using a prior version.

- Screen savers must be turned off during extended DOEPOD operations.
- Copy "DOEPOD" folder to your computer system.

MAC Installation Instructions
Microsoft Office Excel 2008 for Mac does not include the required Visual Basic for Applications, this is being remedied in Microsoft Office Excel 2010 for Mac. Recommend using a prior version.

- MAC is no longer supported. Use a prior version of DOEPOD with older Macs.

DIRECTIONS

  The goal is to reach and satisfy the sample requirements of CASE 1 or CASE 1+ described below. That is, 90/95 POD at X_{POD} is quantitatively determined, and all class lengths larger than X_{POD} are validated to be equal to or greater than 90/95 POD within the range of sample class lengths used, and there are no false call warnings.

  Efficient operation of this program may be obtained by manufacturing the directed samples, inspecting the samples; break down the samples to determine class lengths or use alternate method to establish class lengths, and add the obtained data to the existing data set. Break down of samples is not required if the flaws sizes are known by process control, etc. Note that movement between cases, as obtained by meeting the directed DOEPOD requirements, is not necessarily sequential.

  Follow the numbered instructions below in the order that they are presented.

DOEPOD OPERATIONAL INSTRUCTIONS

Step 1)   Generally, flaws can not be manufactured on demand. Therefore, it is recommended that a set of flaws that spans the range of flaw sizes of interest is produced for validating the inspection system capability. This is a Survey Set. An example Survey Set for an aerospace system may have about 16 flaws. The flaws sizes in an example Survey set may include one flaw of each of these sizes: 0.010", 0.020", 0.030", 0.040", 0.050", 0.060", 0.070", 0.080", 0.090", 0.100", 0.200", 0.300", 0.400", 0.500", 0.600", and 0.700". When qualifying inspectors, there is only one set comprised of 29 flaws of the qualification target flaw size, and 5 large flaws distributed equally in size between the target flaw size and the largest flaw expected to be found.

For both system validation and inspector qualifications there are to be a minimum of 84 opportunities for false calls.

Notes: When validating the inspection system capability, flaws sets with 100's of flaws exhibiting any combination of Hits or Misses may also be used as the initial flaw set. Alternatively, the minimum number of initial flaws is five (5) with one (1) flaw with a class length for which there will be a Miss, and four (4) or more other flaws of different class lengths. These four (4) may exhibit any combination of Hits or Misses. This alternative minimum number of initial flaws should only be used when a Survey Set is unobtainable.

An additional set of 25 flaws uniformly distributed in size between the 90/95 X_{pod} flaw size and three times the 90/95 X_{pod} flaw size are required to complete a systems validation. The 90/95 POD flaws size is not know a priori, therefore, these larger flaws sizes and their range will be identified after 90/95 X_{pod} is reached at a flaw size.

The minimum number of flawed sites for systems validation is 29 flaws of the target flaw size and 25 flaws larger than the target flaw size.

The minimum number of flawed sites for inspector qualification is 29 flaws of the target flaw size and 5 flaws larger than the target flaw size.

Step 2)   Inspect samples and identify a Hit (or a Miss) or Signal Amplitude for each inspection site.

Step 3)   Breakdown samples or use an alternate method to establish actual class lengths (e.g., flaw length).

Step 4)   Enter class lengths (flaw size) and Hit/Miss (or Signal Amplitude) data in columns labeled "Crack Size" and "Hit/Miss" or "Signal Amplitude" of the "Data.xls" spreadsheet. The data is entered starting in row two (2). If signal amplitude data is used then a "Signal Threshold" value is required in row two (2) of the column labeled "Signal Threshold" in Data.xls. A template Data.xls is provided. The label "Data" in the template "Data.xls" file name may be replaced by any file name of interest. The Data.xls spreadsheet must be in the DATA folder.   Example data entries for both Hit/Miss or Signal

Amplitude data are shown below and are the minimum DOEPOD data entry requirements. Here a Hit = 100 and a Miss = 0. Flaw identification labels may and should be listed in the column labeled "ID Number" starting in row 2.

Hit/Miss Data

	A	B	C	D	Q	R
	ID Number	CRACK SIZE	DEPTH	HIT/MISS (0 or 100)	Signal Amplitude Measured (Arbitrary Units)	SIGNAL TREASHOLD
1						
2	GBDY 1	0.342		100		
3	GBDY 2	0.251		100		
4	GBDY 3	0.242		100		
5	GBDY 4	0.213		100		
6	GBDY 5	0.208		100		
7	GBDY 6	0.199		100		
8	GBDY 7	0.186		100		
9	GBDY 8	0.184		0		
10	GBDY 9	0.169		100		
11	GBDY 10	0.166		0		

Signal Amplitude Data

	A	B	C	D	Q	R
	ID Number	CRACK SIZE	DEPTH	HIT/MISS (0 or 100)	Signal Amplitude Measured (Arbitrary Units)	SIGNAL TREASHOLD
1						
2	GBDY 1	0.342			0.5	0.25
3	GBDY 2	0.251			0.4	
4	GBDY 3	0.242			0.3	
5	GBDY 4	0.213			0.2	
6	GBDY 5	0.208			0.1	
7	GBDY 6	0.199			0.05	
8	GBDY 7	0.186			0.025	
9	GBDY 8	0.184			0.01	
10	GBDY 9	0.169			0.025	
11	GBDY 10	0.166			0.01	

a. Hit/Miss data is entered as a "100" and "0" for a Hit and Miss, respectively. Crack sizes (class lengths) are defaulted to be inches. Note: When using of DOEPOD for analysis by other than length or depth flaw sizes, e.g., flaw area, flaw areas may occur below the reserved number 0.00002. Flaw areas need to be scaled, by the user to exceed this number in order for DOEPOD to recognize these flawed areas as test flaws rather than false call opportunities.

b. If Signal Amplitude is used then the threshold value of a Hit is required. All amplitudes at and above the threshold value are considered Hits. All amplitudes below the threshold value are considered Misses.

c. The crack sizes (class lengths) need not be in a particular order, but the Hit/Miss or signal amplitude data must be in the same row as it's companion crack size (class length). The data is to be contiguous, and the absence of an entry in the "Crack Size" (class length) row indicates the end of the data set. No other data is to be included in the "Crack Size" and "Hit/Miss" columns.

d. Depth data column is for record. Analysis by depth is done by moving any depth data to the "Crack Size" column.
e. Optional: Enter false call information (see FALSE CALL ANALYSIS).
f. Optional: Enter enable MLE analyses or to disable screen updating for faster processing per ADVANCED DOEPOD INSTRUCTIONS.
g. Optional: Enter to indicate Inspector Qualification analyses per ADVANCED DOEPOD INSTRUCTIONS.

Step 5) Run DOEPOD program.

a. Open the "DOEPOD" folder
b. Open the Excel "DOEPOD.xls" program (enabling Macros). There is also a version number listed in the DOEPOD file name listed above.
c. Select "Enable Macros"
d. Select "DOEPOD" button

Mac | PC
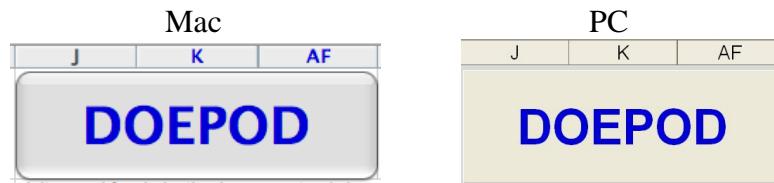
e. All data files in the DATA folder will then be analyzed, and the analysis results will be placed in the ANALYSIS folder.

Step 6) Read DOEPOD CASE identification and the brief description of recommendations in the text box (outlined with dotted lines) on the chart in output file: Analysis.Data.xls which is in the ANALYSIS folder. Pay particular attention to instructions in the charts, before generating more samples. Follow the instructions below. Print file "Analysis.Data.xls", for hard copy of charts.

When opening Analysis files the Macros may be disable by selecting "Disable Macros". The following warning is normal and "yes" should be selected as DOEPOD is protected.

Select "yes"

Step 7) Instructions for Systems Validation when CASE 1, CASE 1+, CASE 1#, or CASE 1* is reached:

a. If **CASE 1** is reached and there are no false call warnings , then validation is complete.
   i. If X_p is absent, validation** is from X_{pod} to X_L.
   ii. If $X_p = X_{pod}$, validation** is from X_{pod} to X_L.

iii. If $X_p > X_{pod}$, validation** is from X_p to X_L.

There is no further action. The user may execute the **X_{podopt} Option** below if desired.

b. If **CASE 1** is reached and there are false call warnings, then validation is not complete. Increase false call opportunities to a minimum of 84 or greater, and resolve any false calls. Return to Step two (2).

c. If **CASE 1+** is reached and there are no false call warnings , then validation is complete when causes of highlighted Misses are understood and resolved.
    i. If X_p is absent, validation** is from X_{pod} to X_L.
    ii. If $X_p = X_{pod}$, validation** is from X_{pod} to X_L.
    iii. If $X_p > X_{pod}$, validation** is from X_p to X_L.

There is no further action. The user may execute the **X_{podopt} Option** below if desired.

d. If **CASE 1+** is reached and there are false call warnings, then validation is not complete. Increase false call opportunities to a minimum of 84 or greater, and resolve any false calls. Return to Step two (2).

e. If **CASE 1# or CASE 1*** is reached and the DOEPOD analysis is for validating that the inspection system meets the inspection requirements, then there are large flaw sample requirements as indicated in the large flaw validation failure note in the output chart. Follow the steps listed below in order to complete the validation:

    i. CASE 1# or CASE 1*: Address all false call warnings. If the false call analysis is successfully executed, then a false call data summary is listed (see False Call Analysis). An estimate of the false call rate and the upper confidence bound of the false call rate is listed in the output charts. Increase false call opportunities to a minimum of 84 or greater, and resolve any false calls.

    ii. CASE 1#: Validation is not completed. The user has two options. (1) Extend the large flaw validation range or add samples as indicated in the large flaw validation failure note. When extending the large flaw range or adding large flaws it is required to assure that 25 flaws (or the number of large flaws indicated in the large flaw validation failure note) flaws uniformly spaced in size between Xpod and the extended large flaw size are included. Any X_L and X_m sample requirements listed are no longer required when meeting large flaw validation failure requirements. Return to Step two (2). Or (2) Execute **Optimum X_{poh} Option** below. Executing the **Optimum X_{poh} Option** is at risk, since this option represents an attempt to move the X_{pod} flaw size to a smaller value.

        iii. CASE 1*: Validation is not completed. Cause of highlighted Misses need to be understood and resolved. The user has two options. (1) Extend the large flaw validation range or add samples as indicated in the large flaw validation failure note. When extending the large flaw range or adding large flaws it is required to assure that 25 flaws (or the number of large flaws indicated in the large flaw validation failure note) flaws uniformly spaced in size between Xpod and the extended large flaw size are included. Any X_L and X_m sample requirements listed are no longer required when meeting large flaw validation failure requirements. Return to Step two (2). Or (2) Execute **Optimum X_{poh} Option** below. Executing the **Optimum X_{poh} Option** is at risk, since this option represents an attempt to move the X_{pod} flaw size to a smaller value.

Notes: **Optimum X_{poh} Option:** The user may optionally add samples at Optimum X_{poh} or X_{podopt} (see notes below) in an effort to demonstrate the existence of X_{pod} at a lower value. Optimum X_{poh} samples added are contained within the size and tolerance range range listed in the analysis chart (See the Survey Set in the **CASE EXAMPLES** section.). Return to Step two (2).

              X_{podopt} Option*: Note that X_{podopt} may be very near the existing X_{pod}. X_{podopt} samples added are approximately uniformly space within the range X_{podopt} − Classwidth @90/95 POD to X_{podopt}, inclusively. Return to Step two (2).*

Step 8)   Instructions if CASE 1, CASE 1+, CASE 1#, or CASE 1* is not reached for System Validation then generate samples[5], **if listed**, and execute instructions in the following priority:

a. Satisfy the sample requirements for the greater of the extended flaw size (identified by the large flaw validation failure) or 2XL. When extending the flaw range due to large flaw validation failure, there are to be a minimum of 25 large flaws equally spaced in size between X_{pod} and the extended flaw size. When adding flaws at 2XL, the flaws added are approximately uniformly spaced within the range X_L to 2*XL*. Go to Step two (2) above.

b. Satisfy the missing large flaw sizes identified by the large flaw validation failure (see specific flaw sizes in the Large Flaw Validation table of the "Analysis Data" sheet, columns CE - CG and rows 2-30). Go to Step two (2) above.

c. Satisfy the sample requirements of the smallest X_{POD} in Table B that is greater than the largest X_{POD} in Table A, and/or the largest X_{POD} in Table

---

[5] Subsets of samples may be used before returning to Step two (2), the risk here is primarily cost. Directed DOEPOD is driven by the observed lower confidence bound, so it is important to execute the Directed DOEPOD program whenever a Miss is observed in order to receive updated instructions for sample requirements.

A (The flaws added are approximately uniformly spaced within the range Table X_{POD} – Classwidth @ X_{POD} to Table X_{POD}, inclusively). Go to Step two (2) above.

d. Satisfy the sample requirements for the larger of X_{LCL}, X_{POH}, or Survey/Optimum X_{POH}. The flaws added are approximately uniformly spaced within the range X_{LCL} – Classwidth @ Best LCL to X_{LCL}, and X_{POH} – Classwidth @ X_{POH} to X_{POH} for X_{LCL} and X_{POH}, respectively. The Survey/Optimum X_{poh} flaws added are contained within the size and tolerance range range listed in the analysis chart (See the Survey Set in the **CASE EXAMPLES** section for more details.). See note below. Go to Step two (2) above.

Note: X_{LCL} X_{POH}, or Survey/Optimum X_{POH} are all equally valid flaw insertion sizes. X_{LCL} exhibits the best lower confidence bound. X_{POH} or Survey/Optimum X_{POH} may have lower confidence bounds, so adding flaws at these sizes is at a higher risk.

e. Satisfy the sample requirements for X_M. Add at least one flaw at X_M. Additional flaws may optionally be added and are approximately uniformly spaced within the range X_M – Classwidth @ Best LCL(or Classwidth @ X_{POD}) to X_M. Note, X_M requirements may be automatically satisfied by previous flaw additions so that more than one flaw may not be required. Go to Step two (2) above.

f. Satisfy the sample requirements for X_L. Add at least one flaw at X_L. Additional flaws may optionally be added and are approximately uniformly spaced within the range X_L – Classwidth @ Best LCL(or Classwidth @ X_{POD}) to X_L. Note, X_L requirements may be automatically satisfied by previous flaw additions so that more than one flaw may not be required. Go to Step two (2) above.

Step 9)   Instructions for supporting inspector qualification:

Identify the test to be for validation of inspector capability (**ADVANCED DOEPOD INSTRUCTIONS** section.

The inspection system must be validated at the target 90/95 POD flaw size, X_{pod} , before performing inspector qualification tests.

The minimum number of flawed sites for inspector qualification is 29 flaws of the target flaw size and 5 flaws larger than the target flaw size.

It is required that there be a minimum of 84 unflawed inspection sites.

a. There are to be 5 flaws larger than the X_{pod} flaw size. The flaws are to be equally spaced between X_{pod} and include the largest flaw expected, X_L. Additional flaws in this size range may be included to provide for an inspector missing a large flaw.

b. If CASE 1 is reached with no false call warnings, and the DOEPOD analysis is for qualification of inspectors, then there is no further action, since the inspection system should have already been validated for 90/95 POD or greater at X_{pod} and for larger flaws. The inspector qualification level will be at the observed 90/95 POD at X_{pod} for the inspector.

c. If CASE 1+, is reached with no false call warnings, and all highlighted Misses are explained and resolved, and the DOEPOD analysis is for qualification of inspectors, then there is no further action, since the inspection system should have already been validated for 90/95 POD or greater at X_{pod} and for larger flaws. The inspector qualification level will be at the observed 90/95 POD at X_{pod} for the inspector.

d. If CASE 1# is reached, then there are not enough large flaws in the test set. See specific flaw sizes in the Large Flaw Validation table of the "Analysis Data" sheet, columns CE - CG and rows 2-30. Any false call warnings are to be explained and resolved. The inspector fails qualification due to inadequate test set up. Follow your Standard's instructions for retesting requirements.

e. If CASE 1* is reached, then there are not enough large flaws in the test set. See specific flaw sizes in the Large Flaw Validation table of the "Analysis Data" sheet, columns CE - CG and rows 2-30. Any false call warnings and all highlighted Misses are to be explained and resolved. The inspector fails qualification due to inadequate test set up. Follow your Standard's instructions for retesting requirements.

f. If CASE 1, CASE 1+, CASE 1#, or CASE 1* is reached with a CONDITIONAL PASS then it is required that the examiner justify and explain all the listed specific conditions of the presence of false calls, the upper confidence bound of the false call rate is too high, there are no false call opportunities, there is an insufficient number or range of large flaws, and the large flaws are poorly distributed in size.

g. If any other case is reached then the inspector fails qualification. Follow your Standard's instructions for retraining and retesting requirements.

ADVANCED DOEPOD INSTRUCTIONS

Disabling the Class Width Optimization

During special operations of DOEPOD it may be useful to fix the class width so that class width optimization does not occur. In order to disable the class width optimization, the user is to change the cells in the "Analysis Data" sheet that indicates "auto" to "noauto" AND to enter the fixed class width (in the Analysis Data" sheet, Column I, Rows 8 and 9) as shown below.

Original run with class width optimization ON:

<div style="margin-left:4em">

Moving Class Width by Size =   **-0.0010**
Auto Increase Class Width =   **auto**

</div>

Modified run with class width optimization OFF and analysis will use 0.039" for the moving classwidth:

<div style="margin-left:4em">

Moving Class Width by Size =   **-0.0039**
Auto Increase Class Width =   **noauto**

</div>

All data sets in the DATA folder will now be run with the fixed moving classwidth listed. No other parameter changes on the "Data Analysis" sheet are supported at this time.

Survey results are not available when user sets or fixes the class width.

Enabling the Maximum Likelihood Estimation (MLE) Analysis

Maximum Likelihood Estimate of POD using a two parameter statistical model. The MLE is included in DOEPOD as a user request for comparison. The included method is that of the NDE Capabilities Data Book, 3rd ed., Nov. 1997, NTIAC DB-97-02, DoD. *The use of MLE estimated POD is not recommend unless a full validation of the estimated POD is performed (see Generazio, E. R., Interrelationships Between Receiver/Relative Operating Characteristics Display, Binomial, Logit, and Bayes' Rule Probability of Detection Methodologies, NASA-TM-2014-21818, April 2014. The MLE method detailed in NASA-TM-2014-21818 is preferred*

The Maximum Likelihood Evaluation (MLE) may be enabled by making an entry (e.g., "yes") in column V, row 2 in the template "Data.xls" file. This entry will apply only to the data set in which it occurs. Disabling the Maximum Likelihood Evaluation is done by deleting the same entry.

 Do not execute MLE analysis

 Execute MLE analysis

Listing Unit Label

The unit labels may be listed in the template in column V row 3. The default unit is inches shown below.

U	V
MLE Analysis? (no = "any entry" or yes = blank) =	
Units? =	inch
Faster processing? (no = blank or yes = any entry) =	
Is there already a validated 90/95 POD flaw size that is smaller than the largest flaw in the this data set? (no = blank or yes = enter the flaw size) =	
Inspector Qualification? (no = blank or yes = any entry) =	
Maximum flaw size allowed =	

Units are in inches

To change the units, enter unit label in column V row 3. An example is shown below for class lengths that are in in^2 areas.

U	V
MLE Analysis? (no = "any entry" or yes = blank) =	
Units? =	in^2
Faster processing? (no = blank or yes = any entry) =	
Is there already a validated 90/95 POD flaw size that is smaller than the largest flaw in the this data set? (no = blank or yes = enter the flaw size) =	
Inspector Qualification? (no = blank or yes = any entry) =	
Maximum flaw size allowed =	

Units are in square inches

Inhibiting Screen Updates for Faster Processing

The screen update may be inhibited for faster processing. The default is to allow screen updates. The screen updates may be inhibited by any entry in column V row 4 as shown below.

U	V
MLE Analysis? (no = "any entry" or yes = blank) =	
Units? =	inch
Faster processing? (no = blank or yes = any entry) =	yes
Is there already a validated 90/95 POD flaw size that is smaller than the largest flaw in the this data set? (no = blank or yes = enter the flaw size) =	
Inspector Qualification? (no = blank or yes = any entry) =	
Maximum flaw size allowed =	

Screen updates inhibited for fast processing speed.

Identifying Inspector Qualification

In order to identify if this is for validating the capability of the inspector (Inspector Qualification). Enter any value in column V, row 6.

U	V
MLE Analysis? (no = "any entry" or yes = blank) =	
Units? =	inch
Faster processing? (no = blank or yes = any entry) =	
Is there already a validated 90/95 POD flaw size that is smaller than the largest flaw in the this data set? (no = blank or yes = enter the flaw size) =	
Inspector Qualification? (no = blank or yes = any entry) =	
Maximum flaw size allowed =	

Validation is not for Inspector Qualification

U	V
MLE Analysis? (no = "any entry" or yes = blank) =	
Units? =	inch
Faster processing? (no = blank or yes = any entry) =	
Is there already a validated 90/95 POD flaw size that is smaller than the largest flaw in the this data set? (no = blank or yes = enter the flaw size) =	
Inspector Qualification? (no = blank or yes = any entry) =	yes
Maximum flaw size allowed =	

Validation is for Inspector Qualification

Setting the Maximum Flaw Size Allowed

The maximum flaw size is the largest flaw that is expected to occur in the component. Typically, this is the maximum flaw size that could occur in the specimen due to the

structural configuration. For example, if the DOEPOD analysis is by flaw length, this value may be the length of the entire sample, or if the DOEPOD analysis is by depth, then this value may be the sample thickness, etc. If this entry is not provided and (3) three times the X_{POD} flaw size is greater than the maximum flaw size, then large flaw validation will fail and indicated requirements for flaws sizes that are greater than the specimen dimensions. Entering the maximum flaw size allows for the large flaw validation analysis to be constrained so as not to go beyond the maximum flaw size.

In order to identify a maximum flaw size. Enter maximum flaw size in column V, row 7.

U	V
MLE Analysis? (no = "any entry" or yes = blank) =	
Units? =	inch
Faster processing? (no = blank or yes = any entry) =	
Is there already a validated 90/95 POD flaw size that is smaller than the largest flaw in the this data set? (no = blank or yes = enter the flaw size) =	
Inspector Qualification? (no = blank or yes = any entry) =	
Maximum flaw size allowed =	

No restriction on maximum flaw size.

U	V
MLE Analysis? (no = "any entry" or yes = blank) =	
Units? =	inch
Faster processing? (no = blank or yes = any entry) =	
Is there already a validated 90/95 POD flaw size that is smaller than the largest flaw in the this data set? (no = blank or yes = enter the flaw size) =	
Inspector Qualification? (no = blank or yes = any entry) =	
Maximum flaw size allowed =	0.06

Maximum flaw size is 0.060".

VALIDATION AT LARGER CLASS LENGTHS

Validation at larger class lengths is required in order to demonstrate that the POD is increasing with increasing class length. The initial DOEPOD recommendations from prior DOEPOD releases were to increase or add samples at the largest class length, X_L, and at a recommended mid-point class length, X_m. The X_m is also dependent on the physics of the inspection system. For example, if a differential eddy current probe system is being evaluated and if the class lengths are greater than the eddy current footprint, then there is a possibility that the POD will decrease when the flaw size is greater than the eddy current footprint. These larger class lengths needed to be included in the DOEPOD analysis. DOEPOD v.1.0 now addresses this issue by requiring specific large flaw sizes to be included in the analysis.

Grouping of flaws by number is allowed as long as the four requirements for using binomial statistics are met. POH should not be varying within the class width group. This is expected to be approximately true when 90/95 POD at X_{POD} exist, where POH for class lengths greater than X_{POD} are expected to be near 1.0. All other Cases may have varying POH, however, the effect of varying POH for these cases is to prohibit 90/95 POD at X_{POD} from being established.

Grouping of flaws by number is executed by DOEPOD for CASE 1, CASE 1+, CASE 1#, CASE 1*, and CASE 2 when 90/95 POD at X_{POD} exists. Class groupings by number may combine up to 76 samples in the class group. X_p identifies the minimum class length at which all class lengths greater or equal to X_p have met or exceeded 90/95 POD when grouped by number of samples. The range of class lengths that have met or exceeded 90/95 POD is shown as a shaded horizontal bar which extends from the class length, X_p, to the largest class length to X_L. The presence of X_p[6] is used to support validation of 90/95 POD for all class lengths at and above X_p.

In addition to validation by number of large class lengths, DOEPOD verifies that there are at least 25 large class lengths equally distributed above the X_{pod} class length, and that theses large class lengths extend to at least three times the X_{pod} class length. (See **DOEPOD VALIDATION** section.) If there are sufficient number of large class lengths, the large class length validation may be completed or additional large class lengths of different sizes are further required. DOEPOD expects the size distribution of large class lengths to have a coefficient of variation (CV) between 0.33 – 0.51. The large flaw validation table in the "Analysis Data" sheet, columns CE - CG and rows 2-30 lists the recommended class lengths that are needed to complete the large class length validation. The expected large class length average, standard deviation, and coefficient of variation are shown in the "Analysis Data" sheet, columns CH – CJ, rows 3-5. The typical listing below indicates that 18 large flaws are needed (in RED) within the flaw size ranges listed. Once these samples are added, the coefficient of variation will be approximately 0.3491 indicating an acceptable distribution of large class lengths.

---

[6] X_p supercedes any LCL above X_{POD} since the maximum number of Misses, in any one group above X_{POD}, is three (3) when grouping by number. In this case, 90/95 POD at the X_p class length is conservative. The user may accept X_{POD} as the 90/95 POD class length, if all LCL values between X_{POD} and X_P are at or above 0.90 and when all Misses in this class length range are understood, explained, and resolved.

CE	CF	CG
LARGE FLAW VALIDATION		
NEEDED LARGE FLAWS =	**18**	
Flaw Size Must be Greater Than:	Flaw Size Must be Less Than or Equal to:	Available Flaw Sizes
0.2015	0.2264	0.2030
0.2264	0.2513	0.2270
0.2513	0.2762	0.2530
0.2762	0.3011	0.2910
0.3011	0.3261	0.3220
0.3261	0.3510	NEEDED
0.3510	0.3759	0.3720
0.3759	0.4008	NEEDED
0.4008	0.4257	NEEDED
0.4257	0.4507	NEEDED
0.4507	0.4756	NEEDED
0.4756	0.5005	NEEDED
0.5005	0.5254	NEEDED
0.5254	0.5503	NEEDED
0.5503	0.5753	NEEDED
0.5753	0.6002	NEEDED
0.6002	0.6251	NEEDED
0.6251	0.6500	NEEDED
0.6500	0.6749	NEEDED
0.6749	0.6999	NEEDED
0.6999	0.7248	NEEDED
0.7248	0.7497	NEEDED
0.7497	0.7746	NEEDED
0.7746	0.7995	NEEDED
0.7995	0.8245	0.8120

CH	CI	CJ
	EXPECTED	
Average	Standard Deviation	Cooefficient of Variation (CV)
0.5254	0.1834	0.3491

CONNECTING VALIDATED 90/95 POD RESULTS FROM PRIOR TESTS

A validated preexisting 90/95 POD flaw size greater than X_{pod} may be included in the DOEPOD analysis, in order to assist with validating that 90/95 POD exists for large flaws. A preexisting 90/95 POD flaw size does not need to be from a DOEPOD analyses. However, preexisting results must have been validated to show that 90/95 POD or greater exists for all flaw sizes at and greater than the preexisting 90/95 POD flaw size to be included. These preexisting POD results must be from use of a similar inspection system using similar procedures for inspections on components that are similar in structure having similar materials and flaw types. Similar is not defined here, but good engineering judgment is to be made to assure that similar describes nearly identical.

Use of a preexisting 90/95 POD flaw size is not permitted for inspector qualification.

Include a validated preexisting 90/95 POD flaw size by entering the 90/95 POD flaw size in column V, row 5 as indicated in the template "Data.xls" file. The preexisting 90/95 POD flaw size is to be less than the largest flaw size in the current data set. The units of the entry must be the same as that used for the flaw size data in column B. Below are examples showing no preexisting 90/95 POD flaw size is available and where a validated preexisting 90/95 POD flaw size of 0.050" is to be included in the DOEPOD analysis.

◇	U	V
5	Is there already a validated 90/95 POD flaw size that is smaller than the largest flaw in the this data set? (no = blank or yes = enter the flaw size) =	

There is no preexisting 90/95 POD flaw size.

◇	U	V
5	Is there already a validated 90/95 POD flaw size that is smaller than the largest flaw in the this data set? (no = blank or yes = enter the flaw size) =	0.050

Yes, there is preexisting 90/95 POD flaw size and include this validated 0.050" 90/95 POD flaw size.

As an example, if there preexisting data validating that 90/95 POD exists for 0.050" and larger flaw sizes for the CASE 1# example shown in Figure 6c, then this information is used to satisfy the missing large flaw size. The CASE 1# moves to CASE 1 as shown Figure 15.
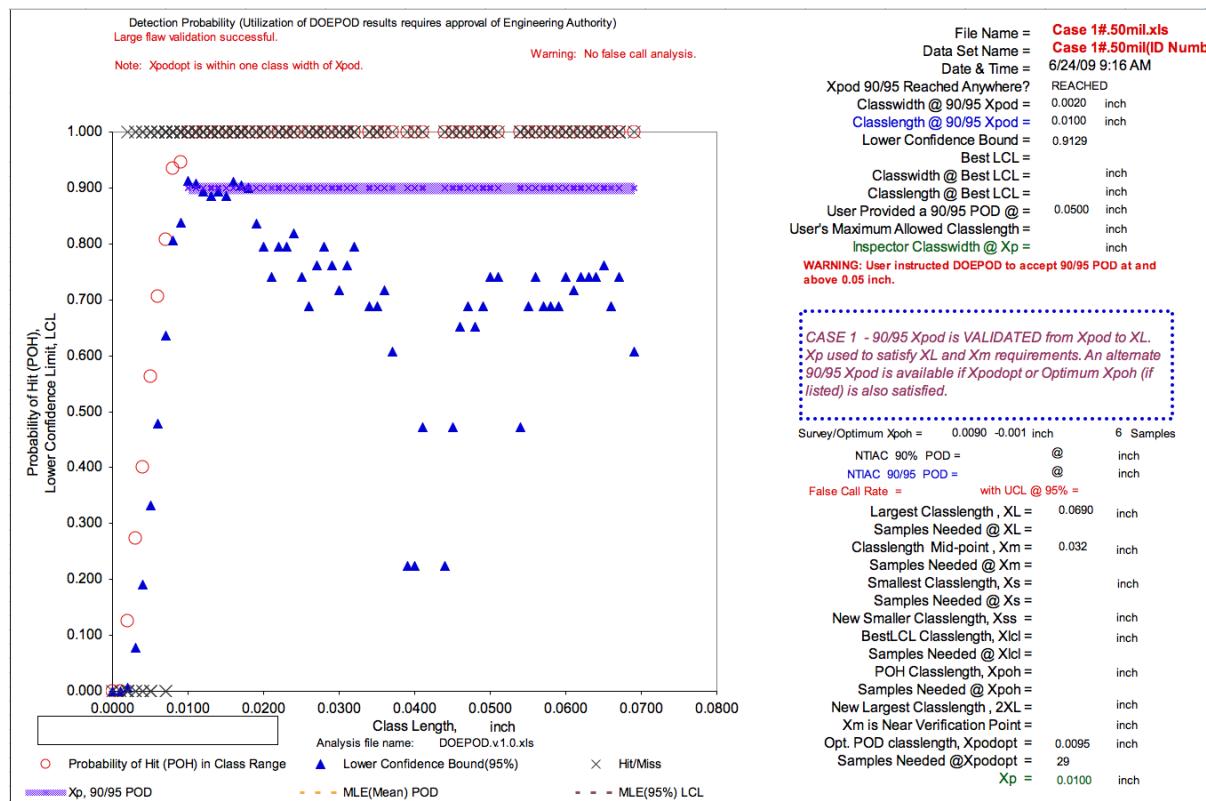
FIGURE 15. Example of CASE 1# moved to CASE 1

The specific utilization of the preexisting 90/95 POD data may be seen in the "Analysis Data" sheet, column CG, rows 1 - 30, where the flaw sizes needed and provided by the preexisting 90/95 POD data are listed in blue color as shown below.

	CE	CF	CG
	LARGE FLAW VALIDATION*		
1			
2	**NEEDED LARGE FLAWS =**		**0**
3			
4	Flaw Size Must be Greater Than:	Flaw Size Must be Less Than or Equal to:	Available Flaw Sizes
5	inch	inch	inch
6	0.0112	0.0135	0.0120
7	0.0135	0.0159	0.0140
8	0.0159	0.0183	0.0160
9	0.0183	0.0206	0.0190
10	0.0206	0.0230	0.0210
11	0.0230	0.0253	0.0230
12	0.0253	0.0277	0.0260
13	0.0277	0.0301	0.0280
14	0.0301	0.0324	0.0310
15	0.0324	0.0348	0.0340
16	0.0348	0.0371	0.0350
17	0.0371	0.0395	0.0390
18	0.0395	0.0419	0.0400
19	0.0419	0.0442	0.0440
20	0.0442	0.0466	0.0450
21	0.0466	0.0489	0.0470
22	0.0489	0.0513	0.0490
23	0.0513	0.0537	**PROVIDED**
24	0.0537	0.0560	0.0540
25	0.0560	0.0584	0.0570
26	0.0584	0.0607	0.0590
27	0.0607	0.0631	0.0610
28	0.0631	0.0655	0.0640
29	0.0655	0.0678	0.0660
30	0.0678	0.0702	0.0690

FALSE CALL ANALYSIS

The presence of false calls artificially increases the estimated probability of detection. The probability enhancement occurs when the process producing false call occurs simultaneously in the presence of a flawed specimen site. That is, when there are false calls, the probability that a Hit is a true Hit is no longer 1.0. An extreme example occurs when an inspector guesses Hits for all inspection sites. As a result, it erroneously appears that the inspector is able to find all flaws required, and with a high probability. Warnings are required to indicate the level of this condition.

For a narrow class width, the number of true Hits may be estimated by the relation[7],

(Number of Observed Hits)(1-UCL) = Estimated Number of True Hits,

where UCL = upper confidence bound of the false call rate at 95% confidence.

For example, using 29 flawed specimens, if UCL = 0.03448 (3.45%) (or one false call out of 29 trials) with 29 observed Hits, the estimated number of true Hits is 28, yielding the estimated POH =28/29 or 0.965 with LCL = 0.896, or 90/95 POD at X_{pod} is not reached for this group even though 29 Hits have been observed. This is in contrast to when the UCL = 0.0 and the estimated POH =29/29 or 1.0 with LCL= 0.9, so that 90/95 X_{pod} is reached for this group.

The presence of false calls affects the entire range of possible classlengths. DOEPOD yields a warning when the upper confidence bound of the false call rate exceeds 0.03448. **The observed 90/95 POD at X_{pod}, when the upper confidence bound of the false call rate**

---

[7] The functional for of this relationship is dependent on how the false call rate is determined, and here it is assumed that the false call rate is determined independently.

exceeds 0.03448, is no longer valid. DOEPOD does not adjust the estimated POH or it's companion lower confidence bound in order to account for false calls. It is left to the user to explain and resolve false calls first.

If the false call rate is 0.0 and the upper confidence bound of the false call rate exceeds 0.003448 then there is an insufficient number of blank test specimens. A minimum of 84 blank specimens or blank inspection sites with no false calls are required to reach an acceptable upper confidence bound of the false call rate of 0.03448.

There are two methods for Including False Calls. **Method (2) is the preferred method** for beginning users of DOEPOD.

Method (1): For test samples or inspection sites with no flaw present, enter flaw size of 0.00001" in column B as a Hit (100, false call) or Miss (0, not a false call) in column D, or a Signal Amplitude in column Q with a Signal Threshold so that the Signal Threshold is used to determine if the 0.00001" entry is a Hit or Miss. Below are examples of a false call for the unflawed sample A1.

A	B	C	D	Q	R
ID Number	**CRACK SIZE**	**DEPTH**	**HIT/MISS (0 or 100)**	**Signal Amplitude Measured (Arbitrary Units)**	**SIGNAL TREASHOLD**
A1	0.00001			0.04	0.03
A2	0.00001			0.02	

ID Number	**CRACK SIZE**	**DEPTH**	**HIT/MISS (0 or 100)**	**Signal Amplitude Measured (Arbitrary Units)**	**SIGNAL TREASHOLD**
A1	0.00001		100		
A2	0.00001		0		

Method (2): Enter any or all of the following parameters in the false call data input table, columns S and T, rows 2-5 of the "Data.xls" template: number false calls and number of false call opportunities (e.g., number of blank samples or number of blank inspection sites), or the linear or area covered by each inspection (inspection window), total inspection length or total inspection area. Entries by this method will superceded entries include in (1) directly above. The example entries below, all represent 100 false call opportunities with 3 observed false calls.

S	T
False Call	**False Call**
Non-Imaging Inspections: Total length of inspection region =	
Imaging Inspections: Total area of inpsection region =	
Length or area covered during each inspection opportunity =	
Number of discrete false call opportunities =	100
Number of discrete false calls =	3

S	T
False Call	**False Call**
Non-Imaging Inspections: Total length of inspection region =	200
Imaging Inspections: Total area of inpsection region =	
Length or area covered during each inspection opportunity =	2
Number of discrete false call opportunities =	
Number of discrete false calls =	3

S	T
False Call	**False Call**
Non-Imaging Inspections: Total length of inspection region =	
Imaging Inspections: Total area of inpsection region =	200
Length or area covered during each inspection opportunity =	2
Number of discrete false call opportunities =	
Number of discrete false calls =	3

If there are no false call opportunities listed via method (1) or (2) above, then a false call analysis is not performed, and the DOEPOD POD results are subject to this uncertainty and a warning is indicated.

False calls and false call opportunities entered by method (2) above are not considered to be part of the total number of samples, and therefore there is no maximum number of false call opportunities when using method (2). However, if method (1) above is used to enter false calls and false call opportunities, then the total number of flawed samples and false call opportunities can not exceed 1999.

If the total inspection region (total length or total area) is provided, then DOEPOD automatically adjusts the false call opportunities to account for the presence of real flaw lengths or areas.

If 90/95 POD at X_{pod} is reached and the inspection widow is not provided, then DOEPOD will use the X_{pod} class length to determine the inspection widow.

If the user supplies contradictory information such as providing both length and area window zone values, or total inspection lengths or areas that are less than the actual total lengths or areas of all flaws, then the false call analysis is not executed.

A false call data summary is listed in rows 45 - 49 and columns H and I of the "Analysis Data" sheet. The example summary shown below includes: the total length or area over which false calls may occur, length or area per inspection (i.e., inspection window or zone), the number of false call opportunities, and number of false calls. An estimate of the false call rate and the upper confidence of the false call rate is listed in the output charts.

◇	G	H	I
45		Total False Call Opportunity Length =	50.203
46		Total False Call Opportunity Area =	
47		Length or Area per Inspection =	0.105
48		Number of False Call Opp's =	478
49		Number of False Calls =	3

VERIFICATION BY MANUAL CALCULATIONS

The numerical POD results of DOEPOD may also be verified by use of manual calculations. Verifying the X_{pod} value is done simple by applying equations 1, 3, and 4 shown in the "Design of Experiments for Validating Probability of Detection Capability of NDE Sysytems (DOEPOD) and for Qualification of Inspectors" document in the references section of this manual. By including the number of trials and Hits from the flaws within the range X_{pod} to X_{pod} -Classlength @ X_{pod}, 90/95 POD will be observed. Similarly, by including increasing numbers of flaws (starting at the largest flaws size) at and below X_p, in the same expressions, 90/95 POD will be observed. The above process is similarly followed for any POH flaws size grouping listed. The numerical false call results of DOEPOD may also be verified by using the number of false call opportunities and number of false calls shown on "Analysis Data" sheet in the table at columns H-I, rows 48 and 49.

REFERENCE & VALIDATION DOCUMENTS

E.R. Generazio, Interrelationships Between Receiver/Relative Operating Characteristics Display, Binomial, Logit, and Bayes' Rule Probability of Detection Methodologies, NASA/TM–2014-218183, April 2014

E.R. Generazio, Validating Design of Experiments for Determining Probability of Detection Capability for Fracture Critical Applications , Materials Evaluation, Vol 69, No. 12, pgs 1399-1407, December 2011

E.R. Generazio, Design of Experiments for Validating Probability of Detection Capability of NDT Systems and for Qualification of Inspectors, Materials Evaluation, Vol 67, No 6, pgs 730-738, June 2009

DESIGN OF EXPERIMENTS FOR VALIDATING PROBABILITY OF DETECTION CAPABILITY OF NDE SYSTEMS (DOEPOD) AND FOR QUALIFICATION OF INSPECTORS

E. R. Generazio[1]

[1]National Aeronautics and Space Administration, Hampton, VA 23681

ABSTRACT. The capability of an inspection system is established by applications of various methodologies to determine the probability of detection (POD). One accepted metric of an adequate inspection system is that there is 95% confidence that the POD is greater than 90% (90/95 POD). Design of experiments for validating probability of detection capability of nondestructive evaluation (NDE) systems (DOEPOD) is a diagnostic tool providing detailed analysis of POD test data, guidance on establishing data distribution requirements, and resolving test issues. DOEPOD demands utilization of observation of occurrences. The DOEPOD capability has been developed to provide an efficient and accurate methodology that yields observed POD and confidence bounds for both Hit-Miss or signal amplitude testing. DOEPOD does not assume prescribed POD logarithmic or similar functions with assumed adequacy over a wide range of flaw sizes and inspection system technologies, so that multi-parameter curve fitting or model optimization approaches to generate a POD curve are not required. DOEPOD applications for supporting inspector qualifications are discussed.

OVERVIEW

The capability of an inspection system is established by applications of various methodologies to determine the probability of detection (POD). One accepted metric of an adequate inspection system is that there is 95% confidence that the POD is greater than 90% (90/95 POD). Design of experiments for validating probability of detection capability of nondestructive evaluation (NDE) systems (DOEPOD) is a methodology that is implemented via software to serve as a diagnostic tool providing detailed analysis of POD test data, guidance on establishing data distribution requirements, and resolving test issues. DOEPOD demands utilization of observance of occurrences. The DOEPOD capability has been developed to provide an efficient and accurate methodology that yields observed POD and confidence bounds for both Hit-Miss or signal amplitude testing. DOEPOD does not assume prescribed POD logarithmic or similar functions with assumed adequacy over a wide range of flaw sizes and inspection system technologies, so that multi-parameter curve fitting or model optimization approaches to generate a POD curve are not required. DOEPOD applications for supporting inspector qualifications is included.

DOEPOD utilizes the concept of "point estimate Probability of a Hit" (POH) at any flaw size (Generazio, 2008). That is, the number of Hits observed per set of specimens exhibiting flaws of similar characteristics (e.g., flaw lengths). The determination of estimated POH at any selected flaw size is a measured or observed quantitative value between zero and one, and knowledge of the estimated POH also yields a quantitative measure of the lower confidence bound. This process is statistically referred to as "observation of occurrences" and is distinct from use of functional forms that predict probability of detection (POD). The driving parameters of DOEPOD are the observed

estimated POH and the lower confidence bounds of the observed estimated POH. Flaw size is referred to throughout the subsequent text as a "class length" for length, depth, area, etc.

The binomial distribution has been used previously for determining POD by observation of occurrences. Prior work (Yee, 1976, Rummel, 1982) used a selection of arrangements for grouping flaws of similar characteristics. Yee (1976) used smoothing optimized probability and overlapping sixty point methods, grouped by number of flaws into a class and by cumulative sums of fixed flaw size class intervals, while Rummel (1982) used fixed class widths. These binomial approaches have lead to the acceptance of using the 29 out of 29 (29/29) point estimate (Yee, 1976, Rummel, 1982, MSFC-STD1249) method, in combination with validation that the POD is increasing with flaw size, to meet the requirements of MSFC-STD-1249 and NASA-STD-5009.   DOEPOD extends work in binomial applications for POD by adding the concept of lower confidence bound maximization as the driver for establishing that there is 95% confidence that the POD is greater than 90% (90/95 POD). DOEPOD satisfies the requirement for critical applications where validation of inspection systems, individual procedures, and operators are required even when a predicted POD curve (NTIAC, 1997) is estimated. Inspection processes and procedures are to be fixed and under control before applying DOEPOD analysis.

DOEPOD follows a series of defined processes to evaluate inspection data that is placed in the user friendly data template files. Details of the processes used are identified in the references at the end of the manual. During operation DOEPOD statistically evaluates the inspection data and identifies the data sets as being a specific case from a particular class of data set classes. The classes range from CASE 1 to CASE 7, referring to fully validated at a 90/95 POD level to extremely far from validation, respectively. Once this class or CASE is known, DOEPOD identifies a series of ordered steps, that if pursued successfully, will lead to full validation.

In addition to validating inspection systems, DOEPOD provides support for the qualification of inspectors. DOEPOD includes the capability to evaluate false call rates for both linear and area inspection windows, and to validate the connection of DOEPOD POD results with other POD results obtained from other previous testing.

DOEPOD KEY DEFINITIONS

|	
C_L	Class length (flaw size)
C_W	Class width (width of the moving class; all flaws within the range C_L to C_L - C_W, inclusively, are grouped together)
LCL	Lower confidence bound of POH @ 95% confidence
X_{Best_LCL}	Class length exhibiting the maximum or "best" LCL, (Best LCL). X_{Best_LCL} is determined by increasing the moving class width until a maximum LCL is obtained.
X_i	Class length of the ith flaw
X_L	Class length of the largest flaw in the data set
X_{pod}	Class length at which the LCL is 0.90 or greater, 90/95 POD
$X_{poh=1}$, X_{POH}	There are no Misses above this class length, and POH = 1 above this class length.

USE OF BINOMIAL STATISTICS

There are four requirements that need to be met in order to determine if a statistical variable is described by a binomial distribution: (1) The number of specimens, N, is to be

fixed, (2) Each observation (or trial) is independent, (3) Each observation represents one of two outcomes (Hit or Miss), and (4) The true probably of Hit (POH) is the same for each possible outcome.

Since flaws of similar characteristics are grouped together, there is a fixed number of specimens in a test, and requirement (1) is satisfied. The definition of similar flaws remains vague and good engineering judgment must be made. Observations (inspections) are made independently and do not depend on the result of the previous test and requirement (2) is satisfied. DOEPOD reduces amplitude signal information to Hit or Miss data satisfying requirement (3). Information is suppressed when reducing analog data to Hit or Miss data and this suppression is acceptable since DOEPOD is not designed for flaw sizing. A concept for converting signal amplitude information to Hit or Miss information is shown in Figure 1. The numbers and shading in Figure 1 may refer to flaw sizes or signal amplitude. The top row indicates that there are many outcomes from signal amplitude data (shading). Once an amplitude threshold is set, all flaws above the threshold have the same probability as being observed as a Hit, and all flaws below the threshold are observed as a Miss. By setting a signal amplitude threshold, compatibility with binomial statistics is assured and requirement (3) is now satisfied. It is noted that false calls may also occur, and these false calls are neither Hits or Misses. However, DOEPOD verifies that the false call rate is sufficiently small as not to affect the utilization of binomial statistics. Meeting requirement (4) will be discussed later.

If the true POH is the same for each outcome, then the probability of observing X Hits after N trials is given by $POH_N(X)$, when the binomial distribution describes the behavior of the count variable X. Example observations are shown as open circles in Figure 1.
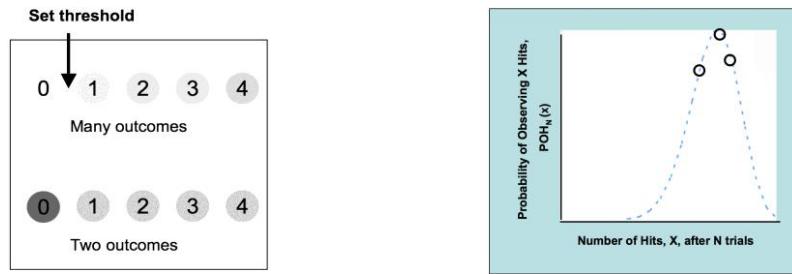
FIGURE 1. Binomialization of test data and probability of observing X Hits after N trials.

DETERMINATION OF CONFIDENCE BOUND FOR POD

Conservative lower confidence bounds for a binomial proportion are given by Equation (1). For example and using identical flaws, with X = 59 hits after N = 61 trials, yields the estimated POH (point estimate) =59/61 = 0.97 (the observed frequency), and the lower confidence bound, LCL , may be obtained (Hald, 1952),

$$LCL = \frac{X}{X+(N-X+1)F_\alpha(f_1,f_2)} \;,\; F_\alpha(f_1,f_2)=2.25 \begin{Bmatrix} f_1 = 2(N-X+1)=6 \\ f_2 = 2X = 118 \end{Bmatrix} \tag{1}$$

$$LCL = 0.9 \;(0.897 \; rounded \; for \; discussion \; purposes) \tag{2}$$

where α is the required confidence level (95%) and $F_\alpha(f_1,f_2)$ is obtained from tables of the F-distribution (MIL-HDBK-5H, 1998). For the procedure and flaw size in this example, and at a 95% confidence level, if LCL = 0.9, then the following statement applies: "This confidence bound procedure has a probability of at least 0.95 to give a lower bound for the 90% POH point that exceeds true (unknown) 90% POH point."

DOEPOD CONCEPTS

DOEPOD is based on the application of the binomial distribution to a set of flaws that have been grouped into size classes, where each class has a width. The classes are allowed to vary in width and start at 0.001" and increase in width by 0.001" increments. Classes start at the largest flaw and move toward the smallest flaw. Class length is used here to represent the flaws features of interest to allow for flaw depth, shape, volume, etc, to be used as the inspection criteria. The first class width group is assigned to the largest flaw in the data set.   The largest flaw in any class width group is assigned as the identifier of the group. DOEPOD evaluates the probability of Hit (POH) lower confidence bound (LCL) obtained from the flaw data within class width group. The next moving class width group is determined by decrementing the upper and lower class lengths bounding the class width group by 0.001". In this manner the class of uniform width is moved. DOEPOD again evaluates the POH and LCL obtained from the flaw data within class width group. This process continues until the smallest flaw is contained in the moving class width group. The class width is increased by 0.001" and the specimens are regrouped using the larger class width and starts at the largest flaw size. DOEPOD again evaluates the POH and LCL obtained from the flaw data within the larger class width group. This larger class width group is again decremented (moved) as before until the smallest flaw is contained in the class width

group. This process continues for all flaw sizes and class widths until all the flaws are eventually contained within one wide class width group.

If a lower confidence bound does (does not) equal or exceeds 0.90 at any class width, then there does (does not) exists a grouping of flaws detected at the 90/95 POD, X_{pod}. If 90/95 POD at X_{pod} exists, then DOEPOD requires further validation that the POD increases with flaw size (this increase is not assumed a priori) within the range of flaw sizes for which the results are valid. DOEPOD addresses validation at large flaw sizes by using two sequentially applied analyses. The first analysis is to apply binomial statistics to groups of numbers of large flaws with sizes greater than X_{pod}. Grouping of flaws by number (Yee, 1976, Rummel, 1982) is allowed as long as the four requirements for using binomial statistics are met. POH should not be varying within the class width group. This is expected to be approximately true when 90/95 POD at X_{pod} exists, where POH for class lengths greater than X_{pod} are expected to be near 1.0. All other cases may have varying POH, however, the effect of varying POH for these cases is to prohibit 90/95 POD, X_{pod}, from being established. The second analysis is to evaluate the number and distribution of flaws with sizes greater than X_{pod}. POD is also dependent on the physics of the inspection system. For example, if a differential eddy current probe system is being evaluated and if the class lengths are greater than the eddy current footprint, then there is a possibility that the POD will decrease when the flaw size is greater than the eddy current footprint. These class lengths need to be included in the DOEPOD analysis. DOEPOD addresses this issue by requiring specific large flaw sizes to be included in the analysis.

Grouping of large flaws by number is executed by DOEPOD when 90/95 POD at X_{pod} exists anywhere in the data set. Class groupings by number may combine up to 76 large flaws in the class group. X_p identifies the minimum class length at which all class lengths greater or equal to X_p have met or exceeded 90/95 POD when grouped by number of large flaws. The range of class lengths that have met or exceeded 90/95 POD is shown as a shaded horizontal bar in the following figures and extends from the class length, X_p, to the largest class length to X_L, the largest flaw in the data set. The presence of X_p is used to support validation of 90/95 POD for all class lengths above X_p. If $X_p = X_{pod}$, then there is 90/95 POD initial validation at X_{pod}. The phrase "initial validation" is used to indicate that the validation is not complete for large flaws.

Recent Monte Carlo results [Generazio, 2009] have shown that, in addition to the initial validation by number of large class lengths, that there must be at least 25 large flaws (class lengths) approximately uniformly distributed above the X_{pod} class length. DOEPOD requires that theses large class lengths extend to at least three times the X_{pod} class length, and this range, X_{pod} to $3*X_{pod}$ is defined here to be the minimum range of class lengths required for successful validation.   The probability of the DOEPOD procedures successfully (POS) identifying scenarios where the POD is decreasing above X_{pod} is shown in Figure 2 for two data sets labeled as D1002BD and A6003H (NTIAC, 1997).

Probability of Success in Determining if the POD of Large Flaws is less than 90/95 POD
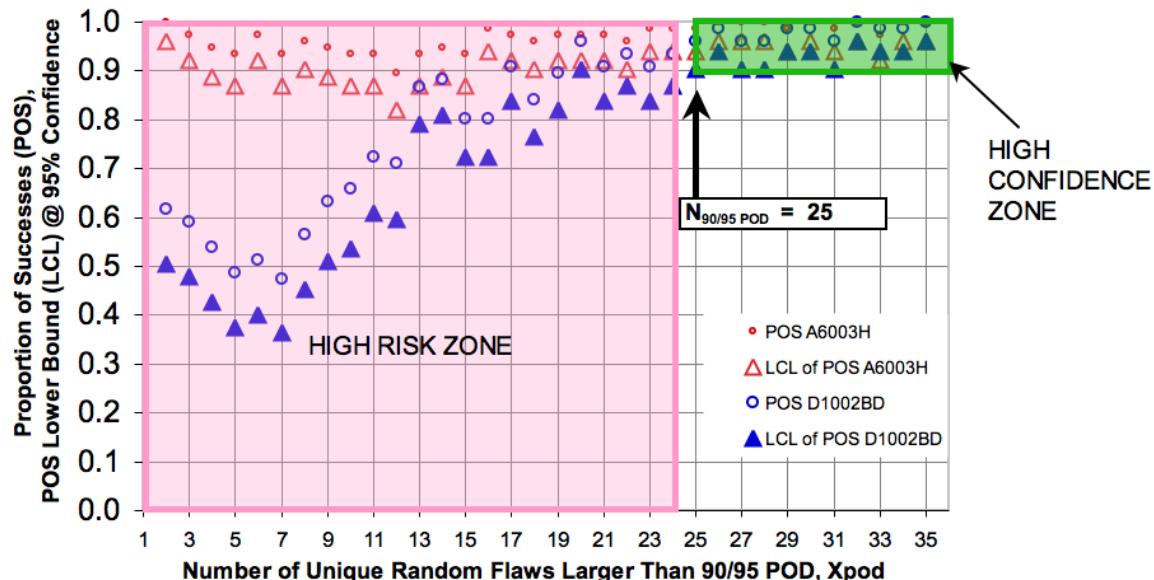
Figure 2. Monte Carlo results showing the minimum number of larger flaws, $N_{90/95\ POD} = 25$, required to demonstrate that there is a 90/95 Probability of Success (POS) in determining if $POD_{Large\ Flaws} < 90/95$ POD. HIGH CONFIDENCE ZONE: The number of flaws (with sizes larger than the 90/95 POD flaw size) required to demonstrate that 90/95 (or greater) POD exists for flaw sizes larger than the 90/95 POD flaw size, X_{pod}. Required when new NDE or enhanced NDE technologies are being evaluated. HIGH RISK ZONE: 90/95 POD for flaw sizes larger than the 90/95 POD flaw size, X_{pod}, is not established by adding theses number of flaws with sizes larger than the 90/95 POD flaw size. This number of larger flaws may be accepted, with justification, when conventional or derivative NDE technologies are being evaluated.

    Meeting the requirement that the true probably of Hit (POH) is the same for each possible outcome may now be addressed. Figure 3 is an example of an abbreviated output of the DOEPOD analysis. The open circles refer to the observed estimated POH. At $X_{pod} = 0.010$", and larger, the observed estimated POH (open circles) is 1.0 (100%), and at 0.010" the lower confidence bound (LCL, filled triangle) is 0.9129. The class width for the estimated POH at 0.010" is 0.002" and this class width is rather small. The interpretation here is that the true POH is similar, i.e., 100%, within the narrow class width of 0.002" at a class length of 0.010". If the true POH was not similar within the class width then the estimated POH would be expected to be less than 100%. Also, note that the estimated POH is observed at 100% for all class lengths above 0.010".

    For class lengths below 0.010" there is a rapidly decreasing estimated POH with decreasing class length. A caution exists for this region when the estimated POH is less than 100%. The estimated POH and the lower confidence bound may be from a group of flaws for which the true POH is varying within the class. Data where the estimated POH is less than 100% are initially used for guidance only with the understanding that binomial statistics requirements may be violated to some extent. If the estimated POH is less than 100%, then DOEPOD uses POH for guidance for specimen selection or for identifying optional class lengths that may be added to achieve 90/95 POD. If the guidance is executed successfully, and the observed lower confidence bound is equal to or greater than 0.9, then validation of the inspection capability may be obtained. The presence of mixed (varying) true POH,

existing within the class widths used, is progressively minimized at the validation and larger class lengths by increased observations of Hits. Since, DOEPOD requires validation that the estimated POH increases with class length, then the presence of mixed true POH within a class yields a conservative value of estimated POH. This reasserts the validity of using a binomial distribution in these cases. By using Hit-Miss, or signal amplitude data with a companion threshold, and while constraining the binomial statistical interpretation of the estimated POH and the lower confidence bound to be applicable only to the validation class length and larger class lengths, the binomial statistics requirement (4) is approximated.

The predicted POD and it's 95% lower confidence bounds as determined by the maximum likelihood estimation (MLE) POD method (NTIAC, 1997) are also shown as upper and lower dashed curves, respectively, in Figure 3. Predicted POD curves are shown here for solely for comparison with DOEPOD POH and LCL observations. These predicted POD curves are dependant on math models, that may not be adequate, and are not used in the DOEPOD analysis.

DOEPOD FALSE CALL ANALYSIS

False calls are handled similarly except the upper confidence limit is used. Test specimens or inspection sites with no flaws present should be included in all POD tests for determination of false call rate and the upper confidence bound of the false call rate at 95% confidence. There is a warning present when allowing unresolved false calls, specifically, 90/95 POD, X_{pod}, may be reached at cost of increasing false call rate. False calls should not be accepted without first addressing the cause of the false call and identifying procedures to remove false calls. The estimated false call rate is given by,

$$False\ Call\ Rate = \frac{Number\ of\ False\ Calls\ (X)}{Number\ of\ False\ Call\ Opportunities\ (N)} \tag{3}$$

And the upper confidence bound, UCL, is given by,

$$UCL = \frac{(X+1)\,F_\alpha(f_1,f_2)}{(N-X)+(X+1)\,F_\alpha(f_1,f_2)}\ , \qquad \left\{\begin{matrix}f_1 = 2(X+1)\\ f_2 = 2(N-X)\end{matrix}\right\} \tag{4}$$

where α is the required confidence level (95%) and $F_\alpha(f_1,f_2)$ is obtained from tables of the F-distribution. The companion statement that is obtained on false calls is, "This confidence bound procedure has a probability of at least 0.95 to give an upper bound for the UCL false call rate point that is equal or less than the true (unknown) UCL false call rate point." UCL's greater than 3.44% indicate an excessive false call rate, and is observed when there is one false call per 29 trials and element (3) of the binomial statistical requirements is violated.
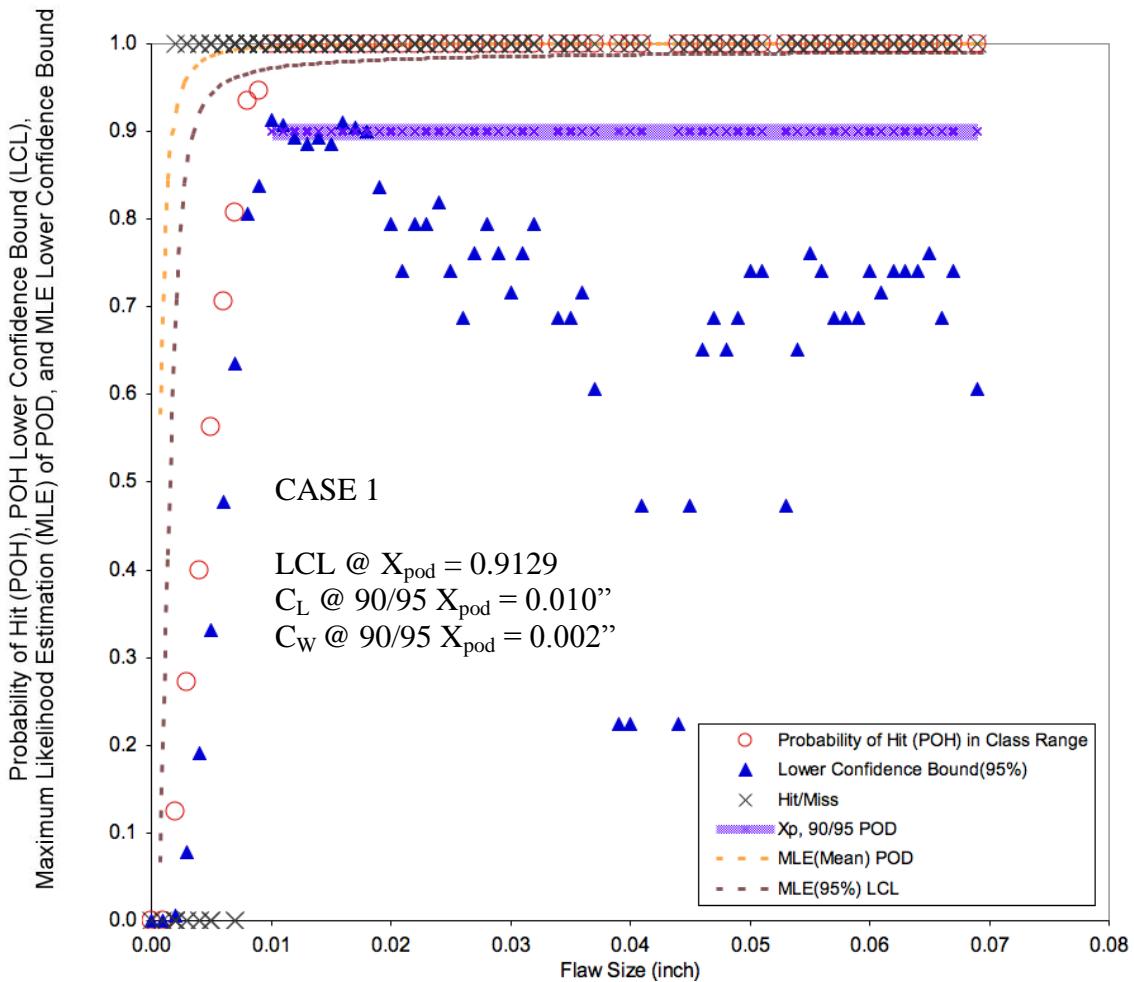
FIGURE 3. CASE 1 example of DOEPOD analysis. Probability of Hit (POH),   POH Lower Confidence Bound (LCL), Maximum Likelihood Estimation (MLE) of predicted POD and MLE Lower Confidence Bound versus flaw size.

DOEPOD CASE EXAMPLES FOR SYSTEMS VALIDATION

DOEPOD classifies the POD data as being one of seven different cases. The cases are identified as CASE 1, 2, 4, 5, 6, 7, and Survey Data sets. During development of DOEPOD the number of unique cases was not known, and CASE 0 (all Hits) and CASE 3 (multiple flaws sizes where 90/95 POD is observed for a fixed class width) are now included in CASE 1 and 2, respectively. CASE 1 is the only case exhibiting full validation. CASE 1 has three sister cases (not shown), CASE 1+, CASE 1#, CASE 1* that indicate specific reasons why the full validation CASE 1 has not occurred. The differences in the cases are highlighted in Table I.

CASE 1 is the best case and is shown in Figure 3.   There is an adequate distribution of flaws at X_{pod} and there is a sufficient number of well distributed large flaws above the X_{pod} flaw size. 90/95 POD is reached at a class length, X_{pod} , and there are Misses only below X_{pod} and full validation is demonstrated when any false call warnings are addressed. Note that the point estimate lower confidence values (solid triangles) for flaw sizes greater than X_{pod} are scattered and below 0.90. This decrease in the lower confidence values is due to the very limited number of flaws at these large flaws sizes and is typically observed in most data sets.

CASE 1+ requires Misses above X_{pod} to be explained and resolved, and any false call warnings addressed before validation is achieved. There is an adequate distribution of flaws at X_{pod} and there is a sufficient number of well distributed large flaws above the X_{pod} flaw size.

CASE 1# requires further validation at flaw sizes greater than X_{pod} by adding specified large flaws at the sizes identified. There is an adequate distribution of flaws at X_{pod}, however, there is an insufficient number of well distributed large flaws above X_{pod} flaw size.

CASE 1* requires further validation at flaw sizes greater than X_{pod} by adding specified large flaws at the sizes identified and requires Misses above X_{pod} to be explained and resolved. There is an adequate distribution of flaws at X_{pod}, however, there is an insufficient number of well distributed large flaws above X_{pod} flaw size.

CASE 2 is the most interesting case and is shown in the Figure 4. In this case, 90/95 POD is reached at a class length, X_{pod}. There is an adequate distribution of flaws at X_{pod} however, there are too many Misses above X_{pod}. The number of flaws with sizes greater than X_{pod} needs to be increased. There are Misses below X_{pod} and excessive Misses above X_{pod}. Therefore, the 90/95 POD at X_{pod} can not be accepted as a validation flaw size. The term excessive is used here since the binomial analysis by number of flaws yields a Best LCL less than 0.90 for large flaws. Since excessive Misses exist at class lengths, X_i, above X_{pod}, then these greater lengths need to be validated by adding more test data. The DOEPOD recommendations are listed as two options that may be executed to establish an acceptable and generally larger 90/95 POD flaw size. Successful execution of the recommendations will transition this CASE 2 to CASE 1. Option 1 is to add specimens of class length X_i where POH<1 (Figure 5, TABLE A). Starting from largest class length, X_i, and work toward small class lengths until reaching an new acceptable larger X_{pod} or reaching X_{pod}. Option 2 is to add specimens of class length X_i where POH=1 (Figure 5, TABLE B, below), and accept a larger X_{pod} class length at the X_i selected. This acceptance is valid as long as any class lengths larger than the new X_{pod} class length where POH<1 are shown [via Option 1 above] to be at 90/95 POD or greater. Acceptance of a larger X_{pod} is not necessarily the ultimate X_{pod} capability of the inspection system, but rather the current demonstrated capability of the inspection system. It is also important to recognize that by introducing additional data an acceptable or larger X_{pod} may never be obtained. In summary, the initial DOEPOD recommendations for CASE 2 are to satisfy the smallest X_{pod} in Table B that is greater than the largest X_{pod} in Table A, and/or the largest X_{pod} in Table A.

FIGURE 4. CASE 2 example of DOEPOD analysis recommendations. Probability of Hit (POH), POH Lower Confidence Bound (LCL), Maximum Likelihood Estimation (MLE) of predicted POD and MLE Lower Confidence Bound versus flaw size for data set D8001(3)L


TABLE A* Selected class lengths with existing misses. Each point requires additional samples in or to achieve the Xpod listed.		TABLE B* Selected class lengths with no misses. Additional samples at these class lengths will achieve the Xpod listed.	
Xpod,Class Length	No. Need	Xpod,Class Length	No. Need
0.5930	73	0.6120	26
0.5730	31	0.5780	20
0.5720	34	0.5760	17
0.5710	37		

FIGURE 5. Screen image of A and B tables for the CASE 2 data set shown in Figure 4.

CASE 4 is similar to CASE 1 except that 90/95 POD at X_{pod} is not reached anywhere as shown in Figure 6. There is an inadequate number of flaws with similar sizes, therefore,

the number of flaws needs to be increased. The best lower confidence bound, Best LCL, is below 0.9 for the best class width group. There are no Misses at or greater than the $X_{Best\ LCL}$ class length, or within the class width group exhibiting the best LCL, $X_{Best\ LCL}$. This is a well behaved data set as defined by the absence of Misses above $X_{Best\ LCL}$. The DOEPOD recommendations are to add specimens of $X_{Best\ LCL}$ or X_{POH} in class length in order to achieve 90/95 X_{pod} at $X_{Best\ LCL}$ or X_{POH}, respectively.
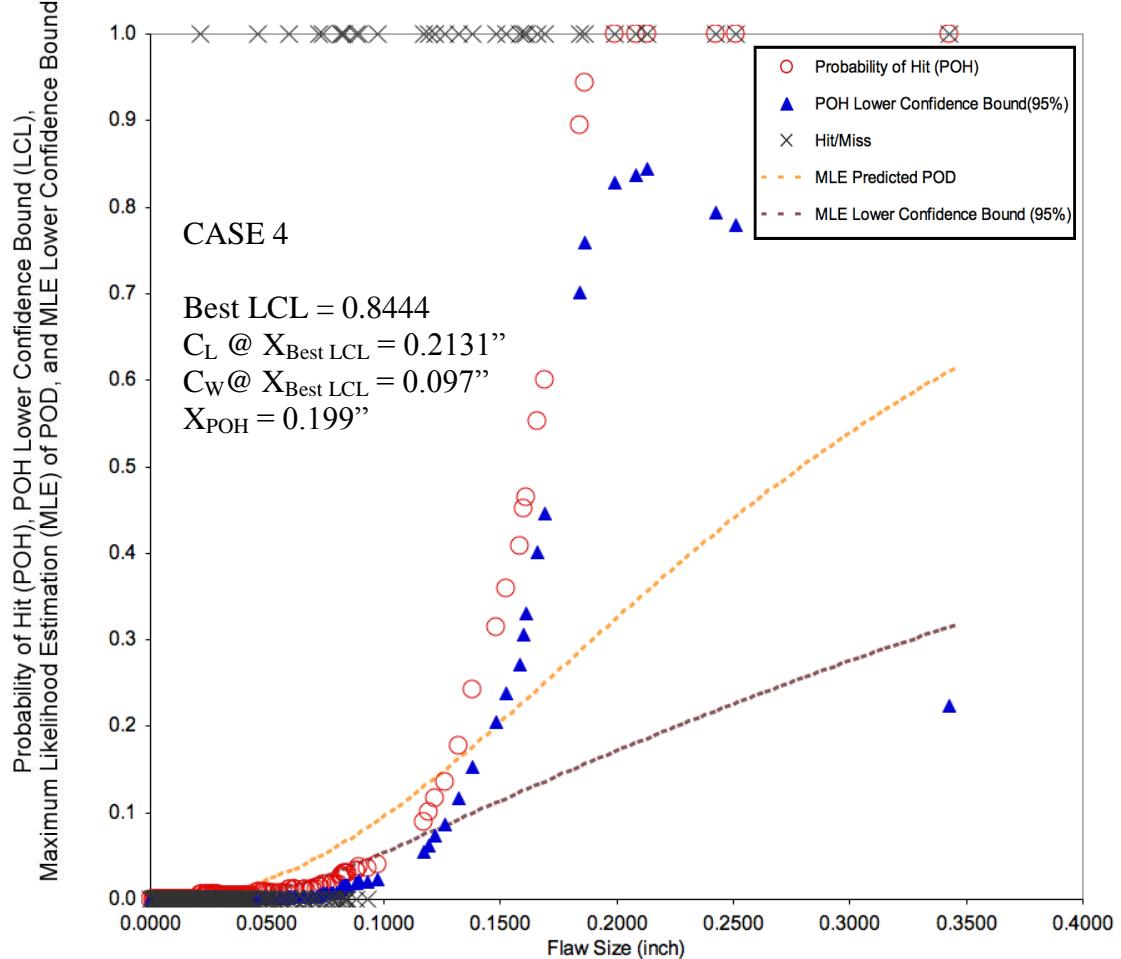

Figure 6. CASE 4 example of DOEPOD analysis. Probability of Hit (POH), POH Lower Confidence Bound (LCL), Maximum Likelihood Estimation (MLE) of predicted POD and MLE Lower Confidence Bound versus flaw size.

CASE 5 is similar to CASE 2 except 90/95 POD at X_{pod} is not reached anywhere as shown in Figure 7. The POH is well behaved for flaw sizes at and above X_{POH}, therefore, the number of flaws with sizes at X_{POH} needs to be increased. There is an inadequate number of flaws at $X_{Best\ LCL}$ and there are misses above $X_{Best\ LCL}$. There are Misses at or greater than the class length $X_{Best\ LCL}$ or within the $X_{Best\ LCL}$ class width group. There exists a class length, $X_{POH=1}$, above which there are no Misses. There are no Misses for class lengths equal to greater than $X_L /3$ (i.e., $X_{POH=1} \leq X_L /3$). $X_{POH=1} \leq X_L /3$ so that POH is not fluctuating at larger class lengths. DOEPOD recommendations are to use $X_{POH=1}$ as the trial X_{pod} by adding specimens at $X_{POH=1}$.
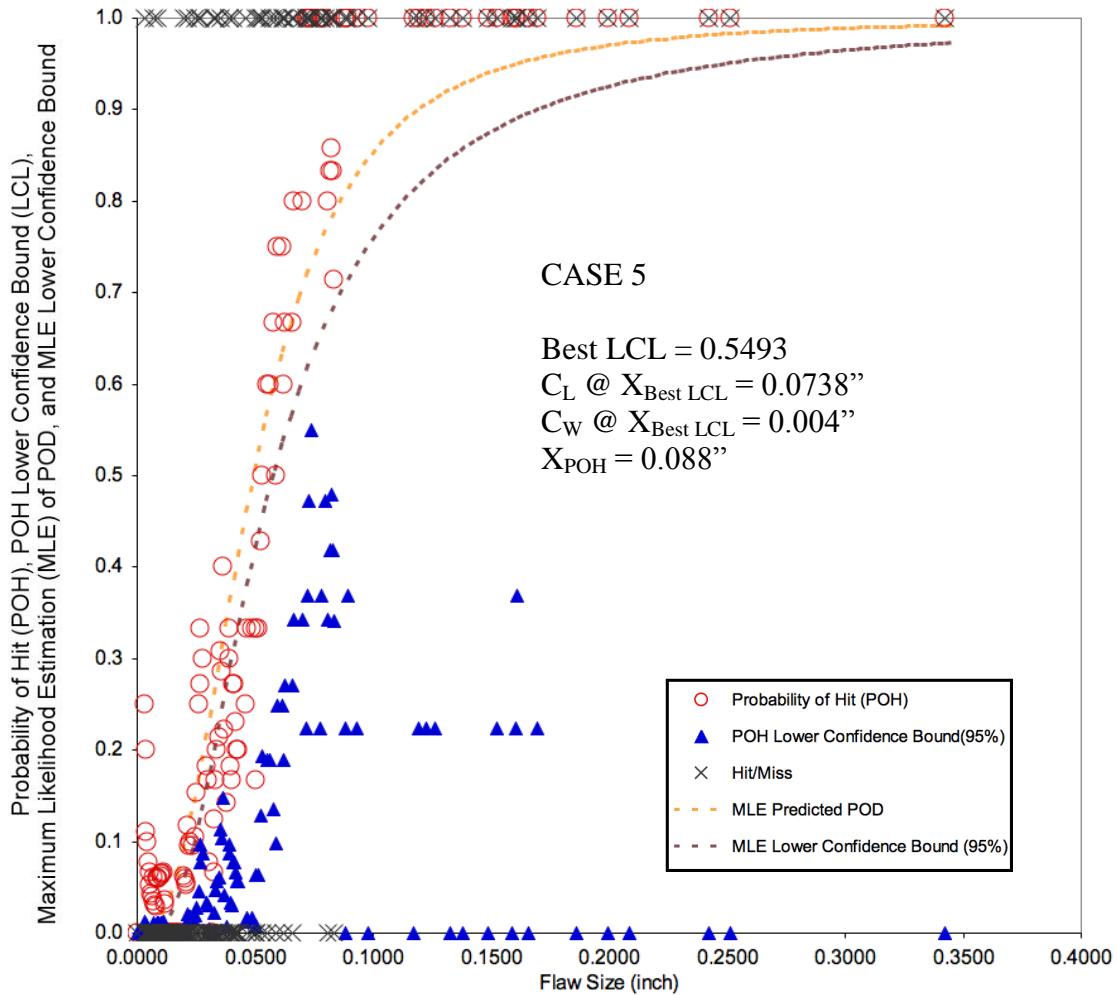
Figure 7. CASE 5 example of DOEPOD analysis. Probability of Hit (POH), POH Lower Confidence Bound (LCL), Maximum Likelihood Estimation (MLE) of predicted POD and MLE Lower Confidence Bound versus flaw size.

CASE 6 is similar to CASE 5, 90/95 POD at X_{POD} is not reached anywhere as shown in Figure 8. The POH is fluctuating throughout a considerable range of the flaw sizes used, therefore, the range of flaw sizes needs to be increased. The Best LCL is below 0.9 for the best class width group. There are Misses at $X_{Best\ LCL}$ or within the $X_{Best\ LCL}$ class width group or at class lengths greater than class length $X_{Best\ LCL}$. There exists a class length, $X_{POH=1}$, above which there are no Misses. There are are Misses for class lengths greater than $X_L/3$ (i.e., $X_{POH=1} > X_L/3$). $X_{POH=1} > X_L/3$ so that POH is fluctuating. Since POH is fluctuating at large class lengths, there is a need to expand current range of X_L. The DOEPOD recommendations are to add specimens with class lengths of $2X_L$ or greater, and add specimens at $X_{POH=1}$.
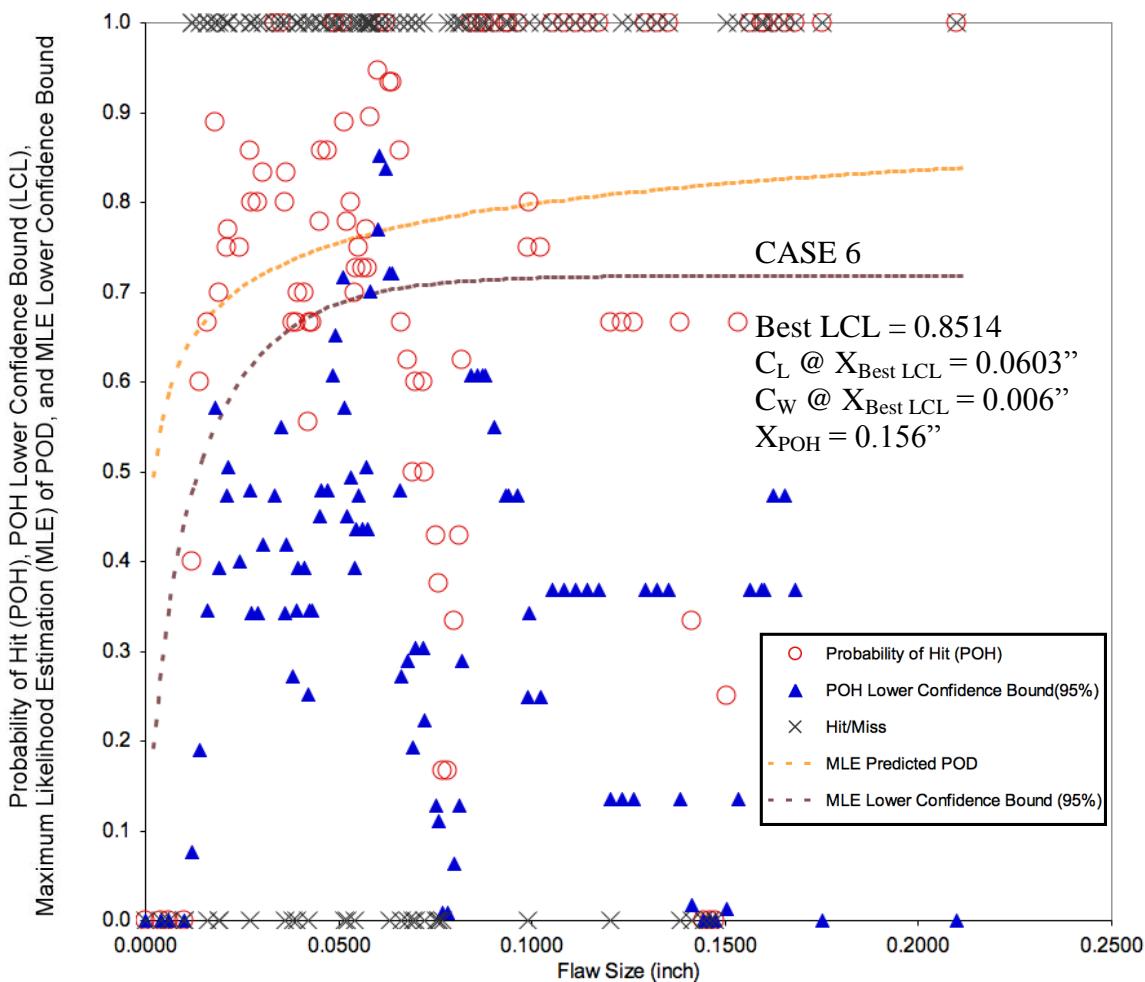
Figure 8. CASE 6 example of DOEPOD analysis. Probability of Hit (POH), POH Lower Confidence Bound (LCL), Maximum Likelihood Estimation (MLE) of predicted POD and MLE Lower Confidence Bound versus flaw size.

CASE 7 (not shown) is similar to CASE 6, 90/95 POD at X_{POD} is not reached anywhere. The POH is fluctuating throughout the entire range of flaw sizes used, therefore, the range of flaw sizes needs to be increased. The Best LCL is below 0.9 for the best class width group. There are Misses at $X_{Best LCL}$ or within the $X_{Best LCL}$ class width group or at class lengths greater than class length $X_{Best LCL}$. There does <u>not</u> exist a class length, $X_{POH=1}$, above which there are no Misses. POH is fluctuating or there may be no Hits anywhere. DOEPOD recommendations are that inspection system may not be appropriate for meeting inspection criteria, or there is a need to expand current range of X_L by adding specimens with class lengths of $2X_L$ or greater.

DOEPOD serves a tool for optimizing the flaw size distribution requirements when analyzing Survey Data Sets. DOEPOD identifies Survey Data Sets when there is an insufficient number of specimens for unconstrained class width optimization as shown in Figure 9. This occurs when the optimized class width exceeds 1/3 X_L <u>and</u> X_{pod} has not been reached. The class width optimization has determined that there is a survey class width for which the smallest $X_{POH=1}$ class length is identified. For survey data sets the optimization procedure that maximizes LCL by increasing class width is superceded. Here, $X_{Best LCL}$ is identified for survey data sets by determining the maximum C_W at X_{POH} for which there are no Misses within the grouping. DOEPOD recommendations are to add samples in the range X_{POH} to $X_{POH} - C_W$, inclusively.

Figure 9. SURVEY example of DOEPOD analysis. Probability of Hit (POH), POH Lower Confidence Bound (LCL), Maximum Likelihood Estimation (MLE) of predicted POD, and MLE Lower Confidence Bound versus flaw size.

Case	Is 90/95 POD at X_{pod} reached? (i.e., lower confidence bound, X_{Best_LCL}, is equal to or greater than 0.9)	Does $X_{POH=1}$ exist?	Is POH = 1 everywhere greater than X_{Best_LCL}?	Is X_{pooh} less than or equal to $X_L/3$?	Large flaw validation complete?	DOEPOD Analysis Summary and Recommendations
CASE 1	YES	YES	YES	NA	YES	90/95 POD at X_{pod} has been reached. Actions: Address any false call warnings.
CASE 1+	YES	YES	NO	NA	YES	90/95 POD at X_{pod} has been reached. Actions: Misses above Xpod need to be explained and resolved. Address any false call warnings.
CASE 1#	YES	YES	YES	NA	NO	90/95 POD at X_{pod} has been reached. Actions: Further validation at flaw sizes greater than Xpod is required. Add large flaws. Address any false call warnings.
CASE 1*	YES	YES	NO	NA	NO	90/95 POD at X_{pod} has been reached. Actions: Further validation at flaw sizes greater than Xpod is required. Add large flaws. Misses above Xpod need to be explained and resolved. Address any false call warnings.
CASE 2	YES	YES	NO	NA	NA	90/95 POD at X_{pod} has been reached, however, there are an excessive number Misses above X_{pod}. Actions: Additional validation at identified flaw sizes is required. Add flaws per instructions.
CASE 4	NO	YES	YES	NA	NA	90/95 POD at X_{pod} has not been reached. Actions: Increase number of flaws at $X_{POH=1}$ or X_{Best_LCL}.
CASE 5	NO	YES	NO	YES	NA	90/95 POD at X_{pod} has not been reached and there are Misses above X_{Best_LCL}. Actions: Increase the number of flaws at $X_{POH=1}$.
CASE 6	NO	YES	NO	NO	NA	90/95 POD at X_{pod} has not been reached. The POH is fluctuating above X_{Best_LCL} and X_{pooh} is greater than X/3. The inspection system is unstable for the flaw size range analyzed. Actions: Increase the flaw size range by a factor of two.
CASE 7	NO	NO	NA	NA	NA	90/95 POD at X_{pod} has not been reached. The inspection system is unstable for the entire flaw size range analyzed. Actions: The inspection system may not be appropriate or increase the flaw size range by a factor of two.
SURVEY CASES	NO	YES	NA	NA	NA	The optimized class width exceeds 1/3 XL and X_{pod} has not been reached. The class width optimization has determined that there is a class width for which the smallest $X_{POH=1}$ class length is identified. Actions: Add flaws at Survey/Optimum X_{POH}

○ = YES   ○ = NO   NA = NOT APPLICABLE

Table I. Summary of all CASES and actions.

INSPECTOR QUALIFICATION

DOEPOD analysis may be applied to evaluate the capability of inspectors. This is similar to validating that the inspection system meets the inspection requirements, except that the requirement for validation at large flaws is not strictly required as it is already included in the systems validation. The 90/95 POD capability of the inspection system must be demonstrated first, by obtaining CASE 1 or CASE 1+ with inspection processes and procedures fixed and under control, before asking inspectors to demonstrate their inspection capability using the inspection system. There are situations where critically large flaws have been missed by inspectors even though the inspection system had a demonstrate capability to find large flaws. Since human factors plays an important and possibly large role here, it is good engineering practice to include large flaws in the sample set when performing inspector qualification. It is recommended, as a minimum, that 29 unique flaws at the target flaw size, X_{pod}, and 5 equally spaced unique larger flaws, along with a minimum of 84 false call opportunity sites, be included in all inspector qualification tests. The largest flaw size should be size of the largest flaw expected in the component. Ideally the number of large flaws is to be 25 in order to strictly assure that the inspector is capable of demonstrating 90/95 POD over the entire expected flaw size range. A minimum of five flaws is reasonably set by experience of current industry qualification test practices, and is solely established by good engineering judgment. POD testing for qualifying inspectors is only one element of inspector qualification. Other elements included in inspector qualification are calibration, adherence to procedures, visual acuity, etc.

COMPARISON BETWEEN THE OBSERVED POD FROM DOEPOD AND THE PREDICTED POD FROM MAXIMUM LIKELIHOOD ESTIMATION (MLE) METHOD

It is important to make comparisons of DOEPOD results with prior POD methods. Figure 4 shows a comparison between predicted POD obtained from MLE and the observed POD obtained from DOEPOD. The DOEPOD 90/95 POD flaw size (upper most solid triangle) is at 0.164" for the data set shown in Figure 4. Although 90/95 POD is observed at a point, 0.164", DOEPOD identifies this data set as a CASE 2 with excessive Misses indicating an oscillating POH for flaws sizes greater than 0.164". Therefore this observed 90/95 POD flaw size can not be accepted as the validation flaw size, and DOEPOD recommends attempting to achieve 90/95 POD at 0.612" via Table B (Figure 5). In contrast, the MLE curve fitting procedure shows the predicted POD (upper dashed curve) increasing for all flaws sizes. However, the presence of 10 Missed (out of 62 opportunities) large flaws above 0.510" in the original data set, makes the MLE predicted 90/95 POD flaw size of 0.513" questionable. This highlights that predictive POD math models may be inadequate for NDE systems.

DOEPOD YIELDS CONSERVATIVE 90/95 POD FLAW SIZE

A DOEPOD analysis of the 437 data sets in the NTIAC NDE Capabilities Data Book, 1997, was performed. There are only 4 data sets for which the DOEPOD 90/95 POD flaw sizes are non-conservative with respect to the MLE predicted 90/95 POD flaw sizes. A close examination (Generazio, 2009) of these four data sets reveals that the MLE math model (NTIAC, 1997) is inadequate and does not represent the observed data. DOEPOD yields an observed conservative value of the 90/95 POD flaw size with respect to MLE predicted 90/95

POD flaw size, except when the math model[8] of the predicted POD does not fit the observed data well.

SUMMARY

The design of experiments for validating the probability of detection (POD) capability of inspection systems (DOEPOD) and for supporting the qualification of inspectors is presented. The statistical and test procedures are discussed and include the concept for binomialization of test data, the process for determining observed probability of Hit (estimated POH) and associated lower confidence bounds, the utilization of moving class width to group flaws and class width optimization, the classification of POD data sets into cases and directed actions or requirements needed to validate inspection systems, and the determination of false call rate and upper confidence bounds. DOEPOD is shown to be a diagnostic tool providing detailed analysis of POD test data, guidance on establishing data distribution requirements, and resolving test issues. A comparison of DOEPOD analysis with maximum likelihood estimation (MLE) POD methodology highlights the conservativeness of DOEPOD results and the need for validating the adequacy of math models used in MLE methods to predict POD for NDE systems. The POD test specimen requirements supporting inspection system validation and inspector qualification are established.

ACKNOWLEDGEMENTS

REFERENCES

E. R. Generazio, Directed Design of Experiments for Validating Probability of Detection Capability of NDE Systems (DOEPOD), 34th Annual Review of Progress in Quantitative Nondestructive Evaluation, Golden, Colorado, July 22-27, 2007; Review of Progress in Quantitative Nondestructive Evaluation, Volume 975, pgs 1693-1700, American Institute of Physics, Melville, New York, 2008; NASA patent disclosure LAR-17656-1

E.R. Generazio, Validating Design of Experiments for Determining Probability of Detection Capability (DOEPOD), Submitted, Journal of Nondestructive Evaluation, Springer, New York, New York, January 6, 2009.

Standard NDE Guidelines and Requirements for Fracture Control Programs, MSFC-STD-1249, National Aeronautics and Space Administration, September 1985.

Metallic Materials and Elements for Aerospace Vehicle Structures, MIL-HDBK-5H, Table 9.6.4.9, pg 9-251, December 1998.

Nondestructive Evaluation Requirements for Fracture Critical Metallic Components, NASA-STD-5009, April 2008

---

[8] Goodness-of-Fit for evaluating adequacy of math models is not presented in NTIAC NDE Capabilities Data Book, 1997 or Mil-HDBK-1823.

Nondestructive Evaluation (NDE) Capabilities Data Book, 3rd ed., Nov. 1997, NTIAC DB-97-02

Ward D. Rummel, Recommended Practice for Demonstration of Nondestructive (NDE) Reliability on Aircraft Production Parts, Materials Evaluation, vol 40, August 1982

Statistical Theory with Engineering Applications, A. Hald, 1952

B. G. W. Yee, et al, Assessment of NDE Reliability Data, NASA-CR-134991, Oct. 1976

Validating Design of Experiments for Determining Probability of Detection Capability (DOEPOD)

E. R. Generazio[1]

[1]National Aeronautics and Space Administration, Hampton, VA 23681

ABSTRACT. The capability of an inspection system is established by applications of various methodologies to determine the probability of detection (POD). One accepted metric of an adequate inspection system is that there is 95% confidence that the POD is greater than 90% (90/95 POD). Directed design of experiments for probability of detection (DOEPOD) has been developed to provide an efficient and accurate methodology that yields observed POD and confidence bounds for both Hit-Miss or signal amplitude testing. Specifically, DOEPOD demands utilization of observation of occurrences. Directed DOEPOD does not assume prescribed POD logarithmic or similar functions with assumed adequacy over a wide range of flaw sizes and inspection system technologies, so that multi-parameter curve fitting or model optimization approaches to generate a POD curve are not required. This work provides validation of the DOEPOD methodology.

 Keywords: Probability of Detection, POD, NDE, NDI, NDT, Nondestructive
PACS: 02.50.Cw, 81.70.–q

INTRODUCTION

Recently is was reported [1] that Design of Experiments for Determining Probability of Detection Capability (DOEPOD) methodology provided a unique perspective on understanding probability of detection data. Specifically, it was reported the probability of detection (POD) data falls into a series of classes or cases. The identification of cases allows development of an intuitive understanding that provides guidance on qualifying nondestructive inspection technologies. This work provides validation of the DOEPOD methodology and extends the validation range from the 90/95 POD flaw size to larger flaw sizes. A DOEPOD analysis of hundreds of POD data sets is performed to validate the conservativeness of DOEPOD results. Monte Carlo testing, using randomly selected larger flaws, is performed to further validate DOEPOD results for larger flaws.

BACKGROUND

DOEPOD utilizes the concept of "point estimate Probability of a Hit" (POH) at any flaw size. That is, the number of Hits observed per set of samples exhibiting flaws of similar characteristics (e.g., flaw lengths). The determination of estimated POH at any selected flaw size is a measured or observed quantitative value between zero and one, and knowledge of estimated POH yields a quantitative measure of the lower confidence bound. This process is statistically referred to as "observation of occurrences" and is distinct from use of functional forms that predict probability of detection (POD). The driving parameters of DOEPOD are the observed estimated POH and the lower confidence bounds of the observed estimated POH. The binomial distribution has been used previously for determining POD by observation of occurrences. Prior work [2, 3] used a selection of arrangements for grouping flaws of similar characteristics. Yee (1976) used smoothing optimized probability and

overlapping sixty point methods, grouped by number of flaws into a class and by cumulative sums of fixed flaw size class intervals, while Rummel (1982) used fixed class widths. These binomial approaches have lead to the acceptance of using the 29 out of 29 (29/29) point estimate [2, 3, 4] method, in combination with validation that the POD is increasing with flaw size, in order to meet the requirements of MSFC-STD-1249 [4] and NASA-STD-5009 [5]. DOEPOD extends work in binomial applications for POD by adding the concept of lower confidence bound maximization as the driver for establishing 90/95 POD. DOEPOD satisfies the requirement for critical applications where validation of inspection systems, individual procedures, and operators are required even when a full POD curve [6] is estimated or predicted. It was noted in prior work [1] that the combined statistical procedures of DOEPOD required further validation by Monte Carlo simulation or similar tests.

DOEPOD EXTENDED FOR LARGE FLAW VALIDATION

Grouping of flaws by number [3] is allowed as long as the four requirements for using binomial statistics are met [1]. In order to meet one of the requirements, POH should not be varying within the flaw size grouping. This is expected to be approximately true when 90/95 POD flaw size, X_{pod}, exists, and where POH, for flaw sizes greater than the 90/95 POD flaw size, is expected to be near 1.0. If POH is varying for large flaw sizes, then the effect of varying POH is to prohibit 90/95 POD from being established for large flaws.

Grouping of large flaws by number [2] is executed by DOEPOD when X_{POD} exists. Flaw groupings may combine up to 76 adjacent flaws in the group. X_p in the following charts identifies the minimum flaw size at which all flaw sizes greater or equal to X_p have met or exceeded 90/95 POD when grouped by number of samples. The range of flaw sizes that have met or exceeded 90/95 POD is shown as a shaded horizontal bar which extends from the flaw size, X_p, to the largest flaw size, X_L. The presence of X_p is used to support validation of 90/95 POD for all flaw sizes at and above X_p. If $X_p = X_{pod}$, then there is 90/95 POD initial validation [1] near the midpoint class length, X_m, and X_L, and if the dependence on the physics of the inspection system is evaluated and determined not to be an issue, then this removes the requirement for 29/29 or similar demonstration at mid-flaw size, and at the largest flaw size. It will be shown here that this is a necessary but not sufficient requirement for validating that 90/95 POD or greater also exists for flaw sizes greater than X_{pod}.

VALIDATING DOEPOD

The following describes the DOEPOD validation testing performed to demonstrate that DOEPOD identification of X_{pod}, the 90/95 POD flaw size, without false call or large flaw warnings and with explanation and resolution of any Misses above X_{pod}, qualifies that the inspection system is adequate and that there is 95% confidence that the POD is greater than 90% (90/95 POD) at and above X_{pod}. Specifically, the DOEPOD validation testing is to demonstrate that 29/29 or similar observation testing on uniquely different flaws yields a conservative value of the 90/95 POD flaw size with respect to maximum likelihood estimation (MLE) predictive POD procedures used in the NTIAC NDE Capabilities Data Book [6]. The MLE procedures used to generate the NTIAC NDE Capabilities Data Book are based on Mil-HDBK-1823 [7] requirements.

There are two phases to the validation testing. Phase 1 is to establish that DOEPOD analysis yields an observed 90/95 POD flaw size that is conservative (i.e., a larger flaw size at 90/95 POD) with respect to MLE predicted 90/95 POD. Phase 2 is to validate DOEPOD procedures for establishing 90/95 POD or better is observed for flaw sizes above 90/95 POD flaw size.

Phase 1: Validate X_{pod} is a Conservative Value

DOEPOD was run on all 437 POD data sets in NTIAC NDE Capabilities Data Book. DOEPOD identified whether 29/29 (or equivalent, e.g., 45/46, 59/61, etc…) 90/95 POD flaw size exists in the data sets, and validates that the POD is observed at 90/95 POD or greater for flaws larger than 90/95 POD flaw size. 153 of the 437 data sets are identified to have 90/95 POD reached by both methods: (1) DOEPOD observations (CASE 1, CASE 1*), and (2) by MLE prediction. These are the data sets that may be compared, and for 145 of the 153 data sets, DOEPOD yields an observed 90/95 POD flaw size that is conservative (i.e., larger flaw size at 90/95 POD) with respect to MLE predicted 90/95 POD flaw size. Therefore, DOEPOD yields a conservative value of the 90/95 POD flaw size for 95% of the data sets contained in the NTIAC NDE Capabilities Data Book. There are 8 out of 153 data sets, for which DOEPOD yields an observed 90/95 POD flaw size that is at least 15% smaller (i.e., less conservative) with respect to MLE predicted 90/95 POD flaw size. The 15% difference is chosen to define and quantify a significant difference between DOEPOD and MLE 90/95 POD flaw sizes.

A careful examination of the 8 data sets identifies both data integrity issues and inadequacy of the MLE model. One of the 8 data sets has erroneous analysis in the NTIAC NDE Capabilities Data Book. When the MLE analysis is corrected, DOEPOD yields an observed 90/95 POD flaw size that is conservative with respect to MLE predicted 90/95 POD flaw size. One of the 8 data sets contains mixed sample thicknesses for an analysis by crack depth to thickness ratio. Comparisons of this data set with other data sets analyzed by either crack length or crack depth is not appropriate for this validation. There are 2 data sets of the 8, for which the the MLE predicted POD 90/95 flaw size is outside the range of the actual flaw sizes in the data set. Use of the MLE 90/95 flaw size for these two data sets without supporting test data near the predicted 90/95 POD flaw size, is not good engineering judgment. This highlights and warns of the predictive nature of the MLE curve fit procedures.

As a result, there are only 4 data sets out of 437, where DOEPOD yields an observed 90/95 POD flaw size that is at least 15% smaller (i.e., less conservative) with respect to MLE predicted 90/95 POD flaw size.

Further evaluation of the 4 data sets exhibiting an apparent but observed non-conservative 90/95 POD with respect to MLE predicted 90/95 POD, reveals that the MLE estimation POD curve fitting approach does not fit the probability of Hit proportions (POH) of the observed data very well. This lack of fit is quantitatively identified by large variances between the MLE predicted POD and the observed probability of Hits proportions (POH). A quantitative comparison between good and poor curve fits is discussed below.

The 4 data sets where DOEPOD yields an observed 90/95 flaw size that is at least 15% smaller (i.e., less conservative) with respect to MLE curve fitted predicted 90/95 POD flaw size are listed below along with their companion variances:

Data Set	Variance
D7002L	0.0329
D7001L	0.0353
CA003(3)L	0.0174
G2001L	0.0440

D7002L analysis results shown in Figure 1 highlight the rather poor fit of the MLE prediction POD function (upper dashed curve), as measured by the variance of 0.0329, to the observe proportions (POH) (open circles). Here DOEPOD identifies and observed 90/95 POD flaw size (upper most solid triangle) at 0.066" in comparison to the MLE predicted 90/95 POD point of 0.165". The proportions are from flaws all having sizes within 0.020" of each other so these grouped flaws are similar in size.
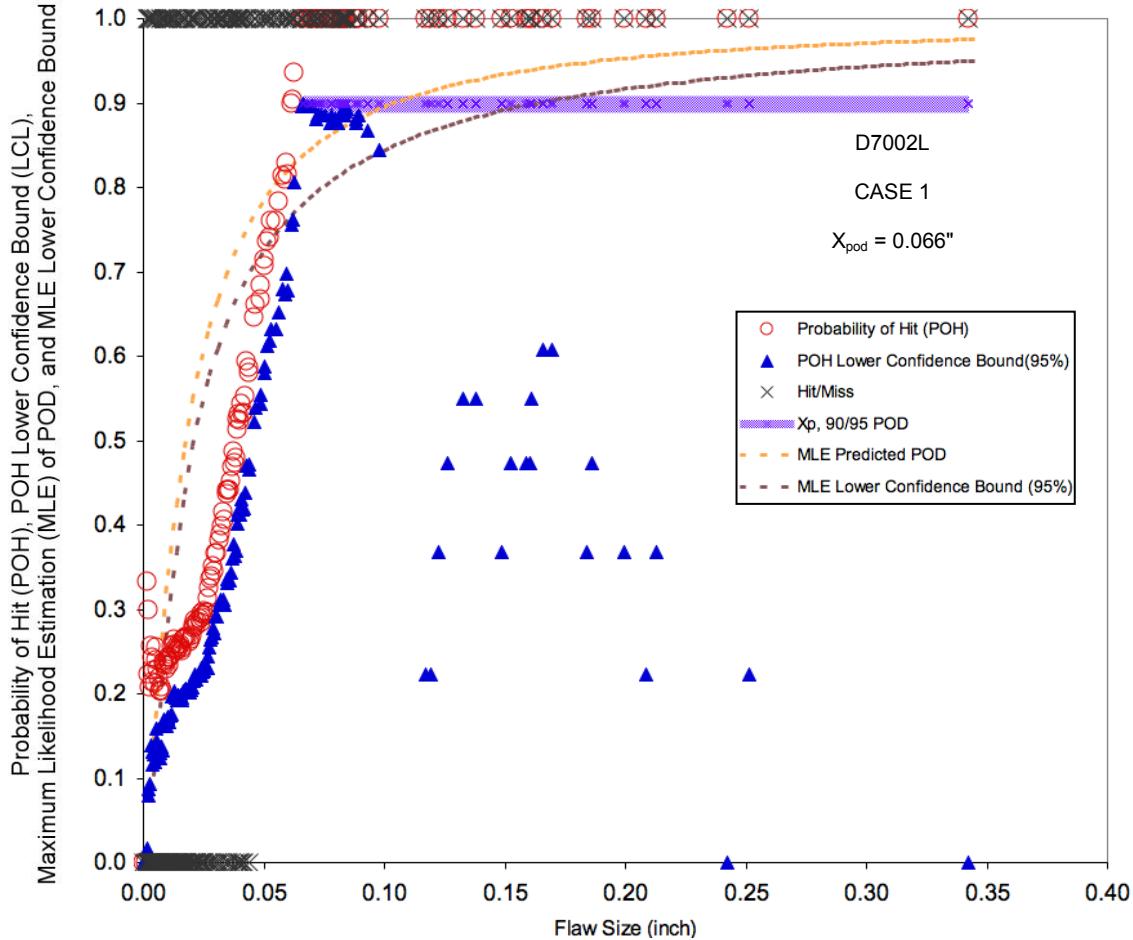
FIGURE 1. Probability of Hit (POH), POH Lower Confidence Bound (LCL), Maximum Likelihood Estimation (MLE) of POD and MLE Lower Confidence Bound versus flaw size for data set D7002L.

A8002L results are shown in Figure 2 for comparison, where the variance of the MLE function is small at 0.0064, and the MLE predicted POD function (upper dashed curve) tracks the observed proportions (POH) well. Here DOEPOD identifies and observed 90/95 POD flaw size (upper most solid triangle) at 0.0147" in comparison to the MLE predicted 90/95 POD flaw size of 0.015". In this example, the DOEPOD 90/95 POD flaw size closely matches the MLE predicted 90/95 POD flaw size.
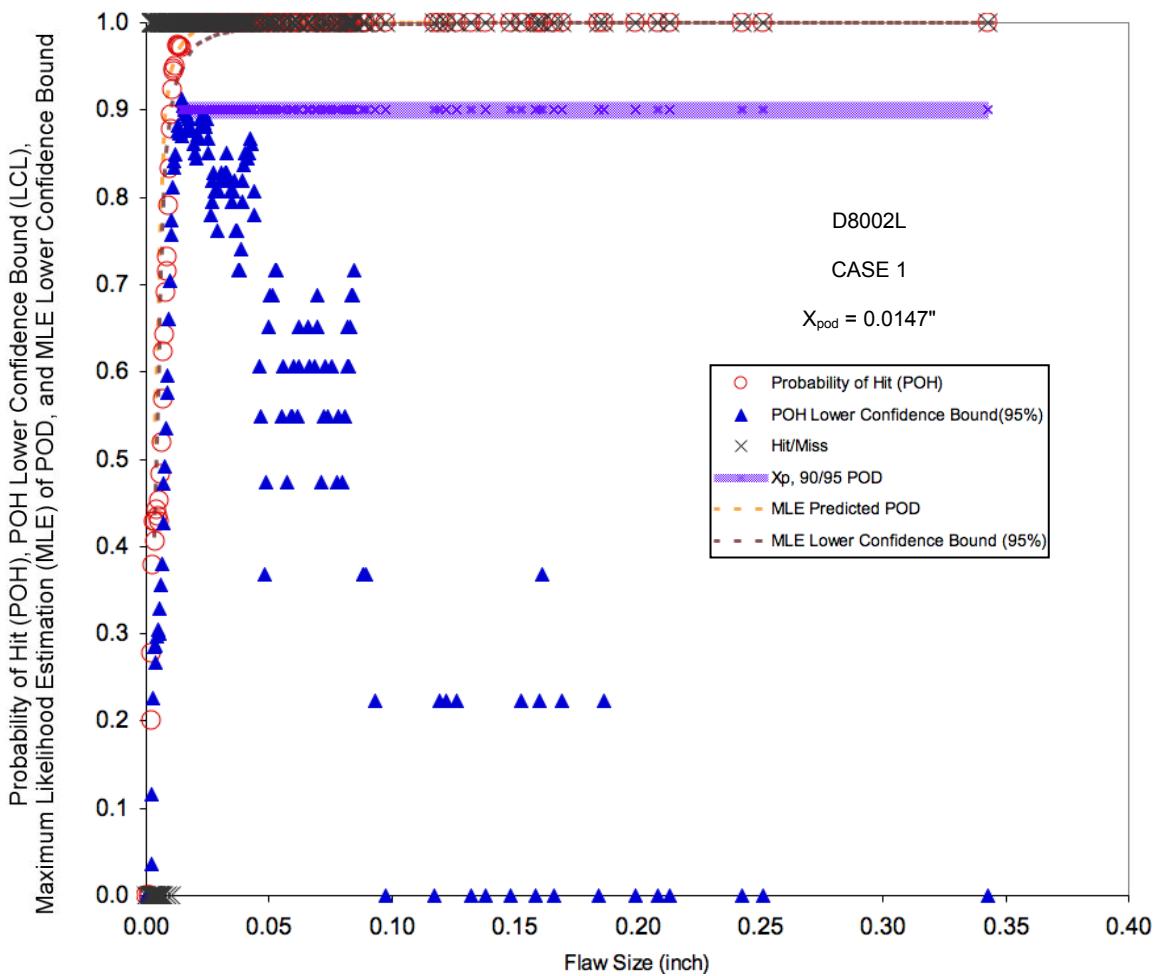
FIGURE 2. Probability of Hit (POH), POH Lower Confidence Bound (LCL), Maximum Likelihood Estimation (MLE) of POD and MLE Lower Confidence Bound versus flaw size for data set A8002L.

Summarizing the above Phase 1 results. DOPOD's use of 29/29 or equivalent to determine the 90/95 POD flaw size, yields an observed conservative value of the 90/95 POD flaw size with respect to MLE predicted 90/95 POD flaw size, except when the math model[9] of the predicted POD does not fit the observed data well. This exception occurred in 4 of 153 data sets in the NTIAC NDE Capabilities Data Book, 1997 where 90/95 POD is reached by both DOEPOD 29/29 or equivalent observations and by MLE predicted POD. The MLE predicted POD math model is inadequate for at least these 4 data sets.

Phase 2: Validate 90/95 POD or Better Exists for Large Flaws

One important aspect of relying on the 90/95 POD flaw size as determined by an isolated 29/29 or equivalent test is that it still remains unknown whether the POD is increasing with increasing flaw sizes above the identified 90/95 POD flaw size. This needs to be evaluated. The fundamental question to be answered is: "If only the tested flaws are

---

[9] Goodness-of-Fit for evaluating adequacy of math models is not presented in NTIAC NDE Capabilities Data Book, 1997 or Mil-HDBK-1823.

those that lead to the 90/95 POD flaw size, X_{pod}, what additional flaws are needed to assure that the POD is also at or greater than 90/95 POD for larger flaws?"

The following describes Phase 2 validation of DOEPOD procedures demonstrating that POD is increasing with flaw size yields a conservative value of the 90/95 POD flaw size.

It is possible that 29/29 or equivalent testing may be validated only at one point, so that further validation is required to verify that the POD is actually increasing with flaw size. DOEPOD identifies the possible presence of this scenario as CASE 2, where further evaluation is needed for flaws larger than the 90/95 POD flaw size. In prior work [1] it was proposed that validation at larger flaw size may be performed by at least three different methods. The first method is to repeat the 29/29 or equivalent testing at two additional flaws sizes: (1) at the largest flaw size, and (2) at a flaw size midway between X_{pod} and the largest flaw size. The second method was to include the of addition of 27 flaws at equally distributed class lengths between X_{pod} and largest flaw size of the test set, and subsequently grouping of flaws by number. The third method is the development of procedures for using good engineering judgment supported by data obtained from similar systems. There is also a caution noted here when identifying flaw sizes for all POD studies. Selection of flaw sizes may be dependant on physics of the inspection system. For example, if a differential eddy current probe system is being evaluated and if the flaw sizes are greater than the eddy current footprint, then there is a possibility that the POD will decrease when the flaw size is greater than the eddy current footprint. These larger flaw sizes need to be included in the POD test.

There are 46 CASE 2 data sets out of 437 POD data sets in NTIAC NDE Capabilities Data Book where further evaluation is needed to validate that the POD is actually increasing with flaw sizes above the 90/95 POD flaw size.

Only 12 of the 46 CASE 2 data sets yield an observed 90/95 POD flaw size that is at least 15% smaller (i.e., less conservative) with respect to MLE predicted 90/95 POD flaw size. These 12 data sets represent possible data samplings for which a 29/29 or similar testing may result in an apparent 90/95 POD flaw size that is not conservative with respect to MLE predicted 90/95 POD flaw size, and that the POD may not be increasing with larger flaw sizes. That is, if the initial specimen or data set is a selected subset of the entire specimen or data set, then an apparent 90/95 POD flaw size that is non-conservative, with respect to MLE predicted 90/95 POD flaw size, may be obtained. The issue is, what if only these selected specimens where generated and tested, then the results of the test on the larger flaws remains unknown, and unknown risk is introduced.

The 12 data sets where this scenario occurs are A3001BL, A6003H, B1003AL, C6003AL, C8001(3)D, CE011(6)D, CE011(6)L, D1002BD, D8001(3)L, D8003(3)D, D8003(3)L, and DC002(3)D.

This risk is highlighted in the next two figures. The DOEPOD and MLE analyses of the original D8001(3)L full data set is shown in Figure 3.
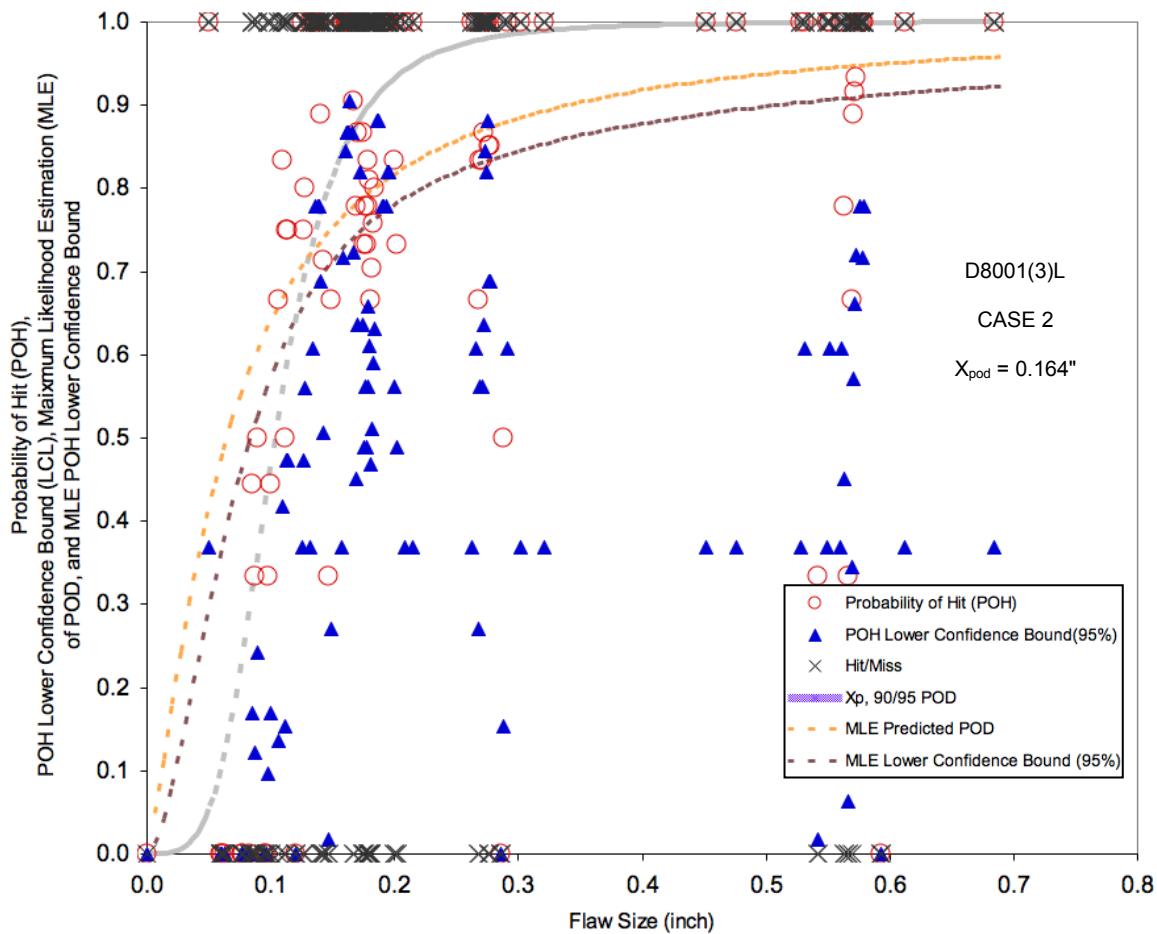
FIGURE 3. Probability of Hit (POH), POH Lower Confidence Bound (LCL), Maximum Likelihood Estimation (MLE) of POD and MLE Lower Confidence Bound versus flaw size for data set D8001(3).

The DOEPOD 90/95 POD flaw size (upper most solid triangle) for this full data set is 0.164". In contrast, by selecting a small sampling consisting of a subset of the original D8001(3)L data one may obtain an identical 90/95 POD flaw size as shown in Figure 4.
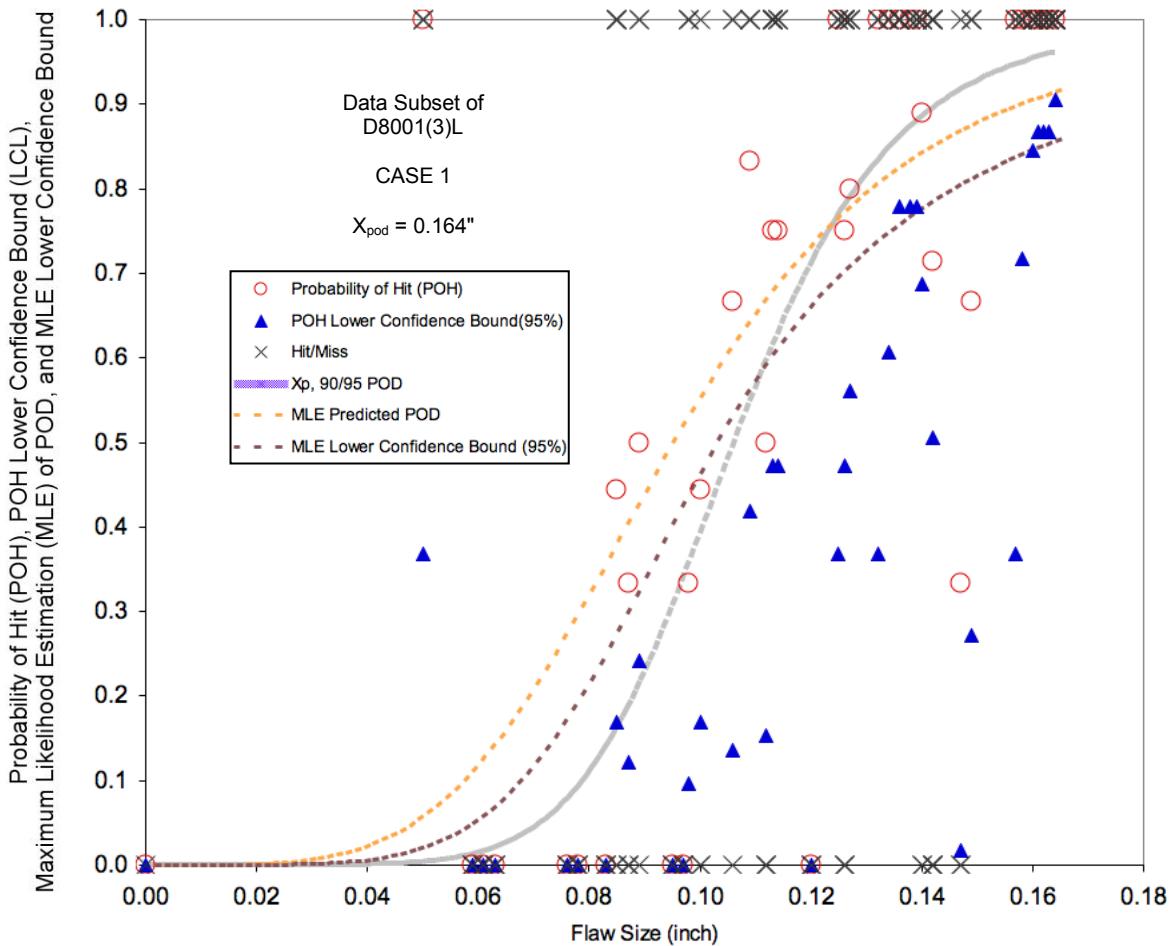
FIGURE 4. Probability of Hit (POH), POH Lower Confidence Bound (LCL), Maximum Likelihood Estimation (MLE) of POD and MLE Lower Confidence Bound versus flaw size for a subset of data in D8001(3).

The DOEPOD 90/95 POD flaw size (upper most solid triangle) is at 0.164" for both the data sets shown in Figures 3 and 4. The MLE curve fitting procedure shows the predicted POD (upper dashed curve) increasing for all flaws sizes and for both data sets, however, the presence of 10 Missed (out of 62 opportunities) large flaws above 0.510" in the original data set, makes this MLE predicted POD questionable. This also highlights that predictive POD math models may be inadequate.

This information now provides us with guidance on how to proceed in validating that the POD is actually increasing with flaw sizes greater than the 90/95 POD flaw size. First, 90/95 POD must be reached at some flaw size. Second, a range of flaw sizes above the 90/95 POD point needs to be included in the data set. Third, the predictive POD models should not be relied upon for demonstrating that the POD is increasing with flaw size above the 90/95 POD point. That is, the adequacy of the predictive model is not assured.

A challenge presents itself in identifying what range and number of flaw sizes need to be evaluated above the 90/95 POD flaw size. So returning to the fundamental question: "If only the generated tested flaws are those that lead to the 90/95 POD flaw size, X_{pod}, (e.g., a typical 29/29 POD test) what additional flaws are needed to assure that the POD is also at or greater than the 90/95 POD for larger flaws?" Here again we can rely upon existing data sets to provide guidance. The guidance is obtained by Monte Carlo testing of DOEPOD procedures. The testing domain is the data from identified files in the NTIAC NDE Capabilities Data Book. Input files for DOEPOD analysis are randomly generated from the

domain files. DOEPOD analysis is performed on the individual data files.   The individual analysis results are aggregated into a final result.

Initial DOEPOD work [1] required a 29/29 or similar demonstration at a mid-point between the 90/95 POD flaw size and the largest flaw size, and a 29/29 or similar demonstration the largest flaws size of the range to be validated. This approach does add two additional flaw sizes for which 90/95 POD or greater is to be observed, but it does not allow for comparison with existing data where there are limited samples at the mid-point and largest flaw size. This does not mean that the mid-point and largest flaw size demonstrations are inadequate, rather that these demonstrations do not allow for direct comparison with existing data sets. However, direct comparison may be made with existing data sets by performing a Monte Carlo test that utilizes existing observed data in a random manner to simulate the lack of a priori knowledge of Hit or Miss information.

The number of flaws required with flaw sizes greater than the 90/95 POD flaw size is unknown. Therefore, it is appropriate to establish what the probability and confidence are that the POD is actually increasing with larger flaws when the number of large flaws and their sizes is varied. In other words, what number and sizes of large flaws are needed to demonstrate that CASE 1 does or does not exist?

In order to perform this Monte Carlo test, a series of randomly generated files are required where the number of flaws having sizes greater than the 90/95 POD flaw size is allowed to increase from 2 to 35. The number range is arbitrary where the actual number required is, at this point, unknown.

There are some important constraints on the domain data files. The domain data files must have a sufficient number of samples available above the 90/95 POD flaw size so that uniquely different samples can be selected. There should be no Misses at the largest flaw size. By DOEPOD design, CASE 1 can never occur when there is a Miss at the largest flaw size, so that these data sets are excluded. Using the above constraints, there are two (2) original CASE 2 data files from which to generate random data files for this simulation. They are files labeled as A6003H and D1002BD.

The first Monte Carlo data set is generated by randomly selecting 1 sample having a flaw size greater than the 90/95 POD flaw size. The largest flaws in the data set is also included to define the range of validation. This completed Monte Carlo data set now contains all the original flaw sizes up to the 90/95 POD flaw size and exactly 1 additional randomly selected flaw larger than the 90/95 POD flaw size, and the largest flaw in the data set. 76 complete random individual Monte Carlo data sets are generated by repeating the process 76 times. This comprises one complete set of randomly generated input data files. The process is continued for 2, 3, 4, … , 34 randomly selected flaws sizes to yield a total 2584 randomly generated input data files. A total of 5168 random data files are generated for the A6003H and D1002BD data sets.

There are two possible outcomes from the DOEPOD analysis of the randomly generated files. Either the DOEPOD analysis yields a CASE 1 (without conditions, to yield a 90/95 POD validation at flaw sizes greater than the 90/95 POD flaw size) or it does not. Here the presence of CASE 1 without conditions represents a failure of the DOEPOD analysis system for either of original CASE 2 data sets A6003H and D1002BD, and represents added risk. It is noted here that the proportion given by the ratio of (Number of Misses)/ (Number of Available Large Flaws) in A6003H and D1002BD data sets are similar at 0.10 and 0.11, respectively. Therefore, during the random selection of one large flaws from either data sets there is approximately 90% chance that a Hit will be selected, and with an observed Hit, CASE 1 may be observed. Conversely, there is a 10% chance that a Miss will be selected, and with an observed Miss, CASE 1 will never be observed.

The original D1002BD CASE 2 and A6003H CASE 2 data sets are shown in Figures 5 and 6, respectively,
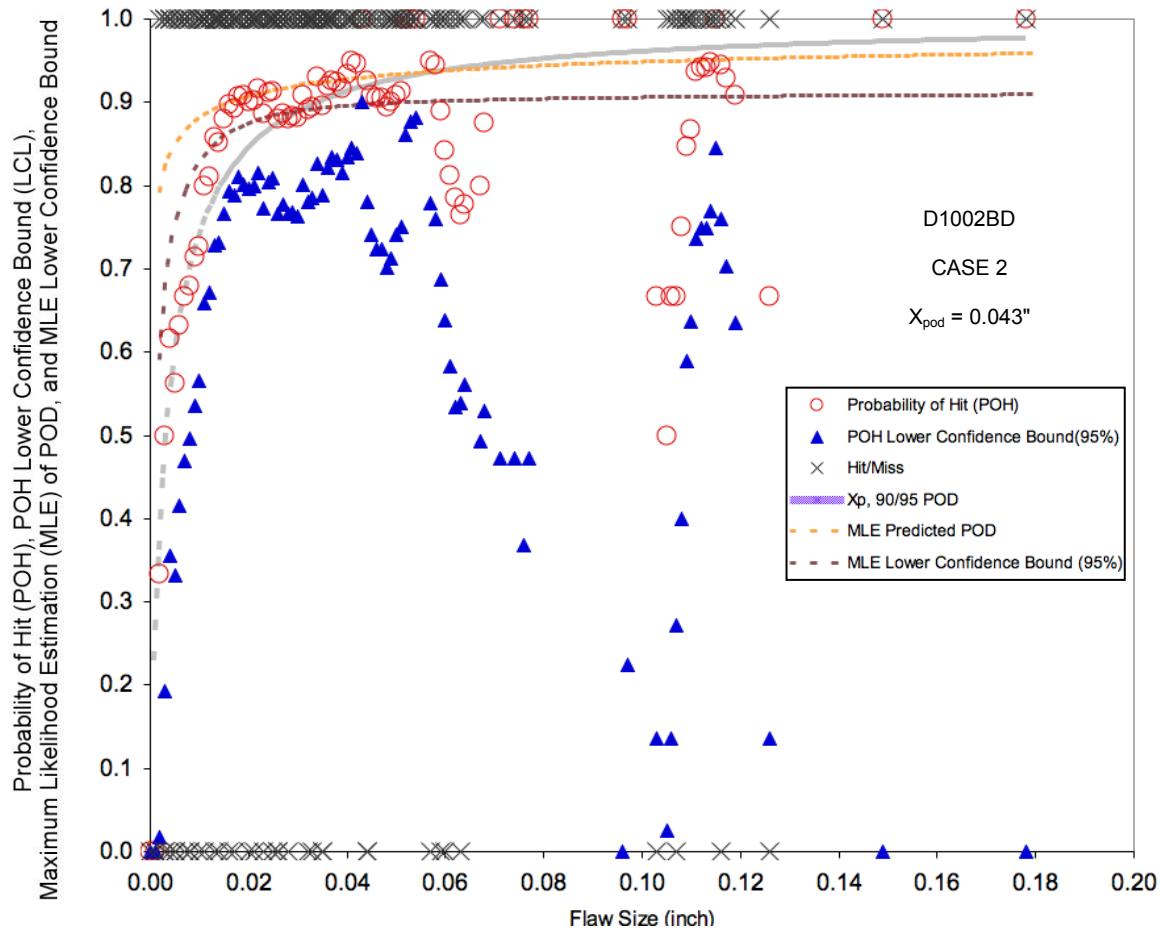
FIGURE 5. Probability of Hit (POH), POH Lower Confidence Bound (LCL), Maximum Likelihood Estimation (MLE) of POD and MLE Lower Confidence Bound versus flaw size for data set D1002BD.
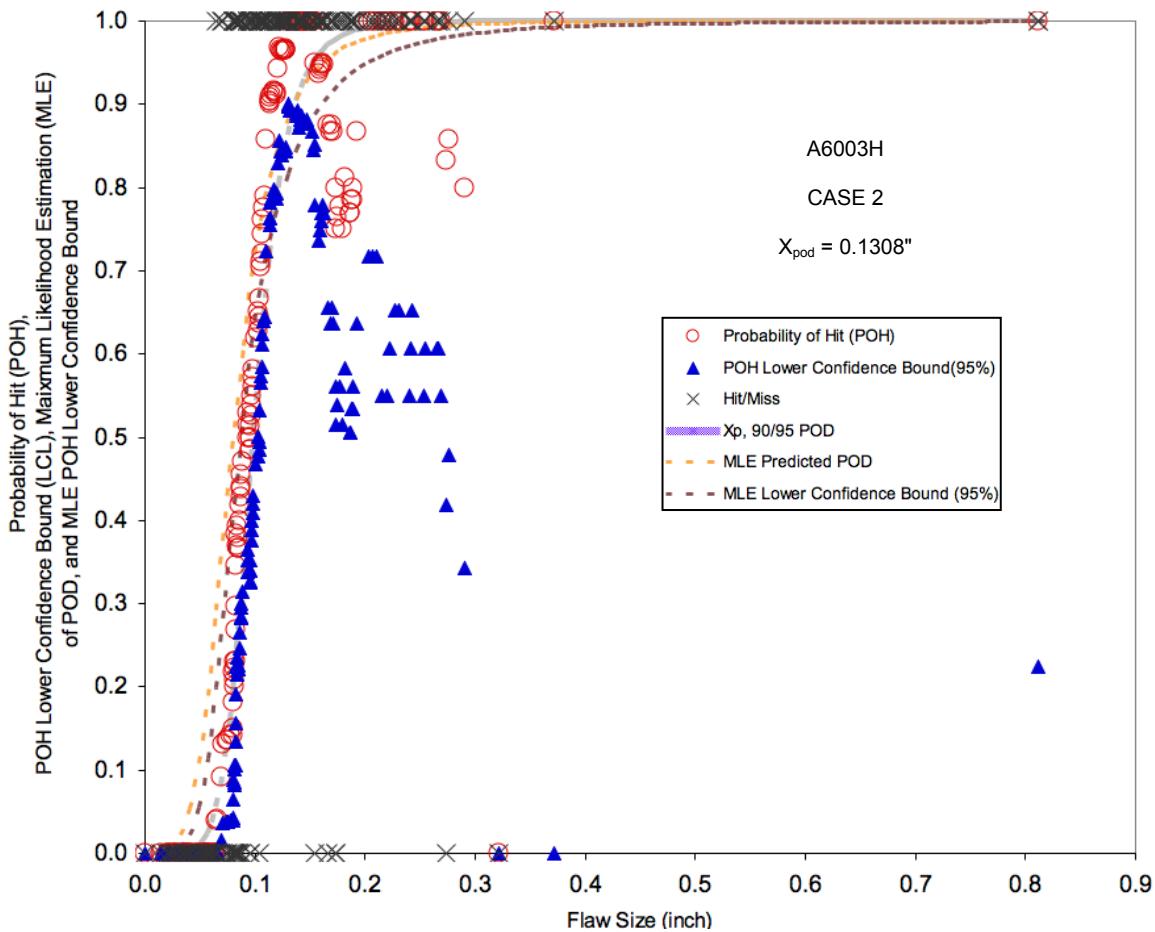
FIGURE 6. Probability of Hit (POH), POH Lower Confidence Bound (LCL), Maximum Likelihood Estimation (MLE) of POD and MLE Lower Confidence Bound versus flaw size for data set A6003H.

A typical random data set generated from the original D1002B data is shown in Figure 7 for trial number 68 when only 20 larger flaws are randomly selected for this trial. DOEPOD yields a CASE 1* and is a success, i.e., not CASE 1, since there are conditions on CASE 1* that limit the validation at large flaw sizes. The conditions are that Misses must be explained and resolved before validation at large flaw sizes is accepted.
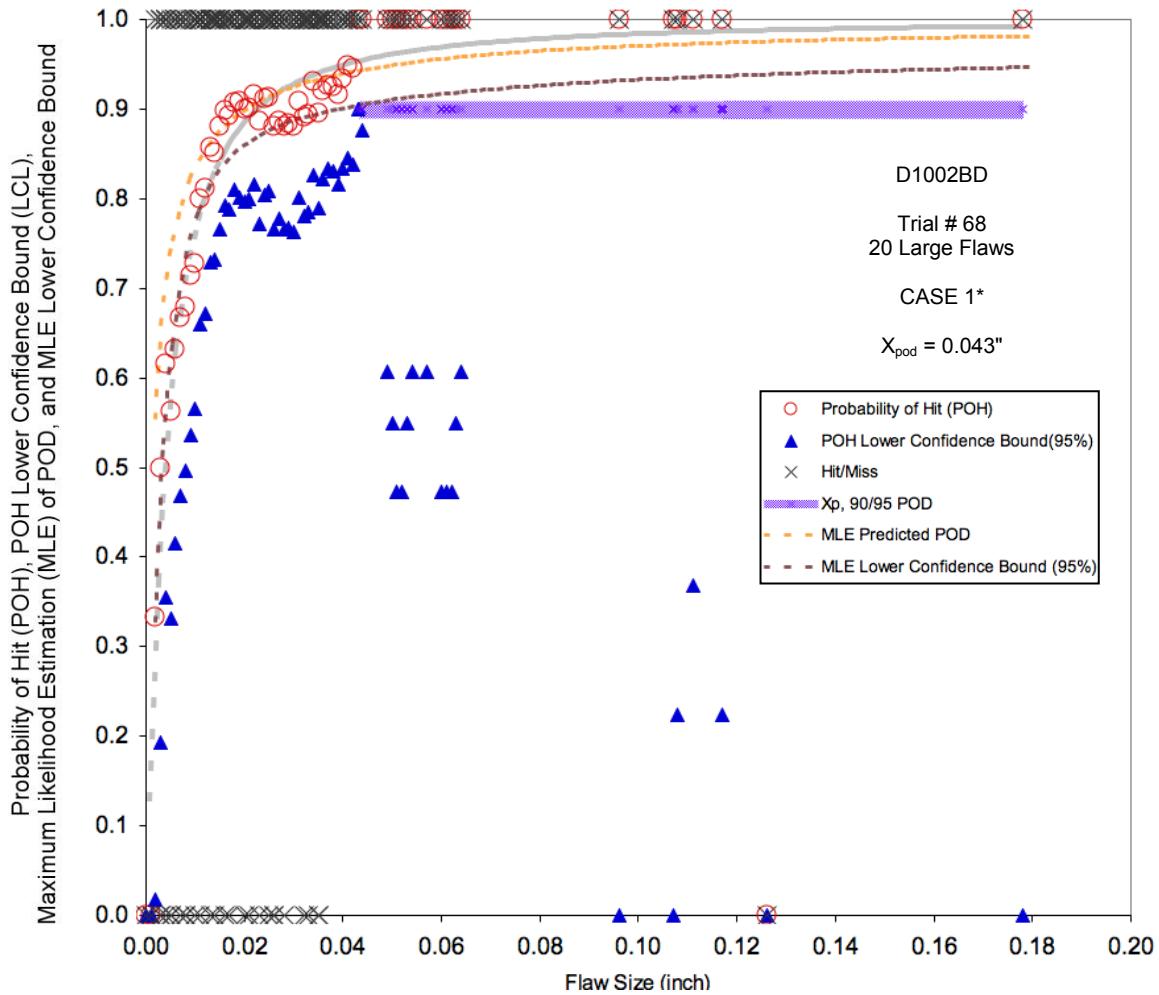
FIGURE 7. Trial #68 with 20 random large flaws from data set D1002BD. Probability of Hit (POH), POH Lower Confidence Bound (LCL), Maximum Likelihood Estimation (MLE) of POD and MLE Lower Confidence Bound versus flaw size.

In contrast, another typical random data set generated from the original D1002BD data is shown in Figure 8 for trial number 65 when only 20 larger flaws are randomly selected for this trial. DOEPOD yields a CASE 1 and is a failure, since there are no conditions on CASE 1 that limit the validation at large flaw sizes. Note the absence of Misses in this random data set above the the 90/95 POD flaw size, 0.043". This trial represents added risk where the random data selected from the original CASE 2 data set yields a CASE 1, and for this reason DOEPOD Prerelease v.1.0 fails. That is DOEPOD Prerelease v.1.0 fails to identify any difficulty in detecting large flaws, even when 20 large flaws are included in the analysis.
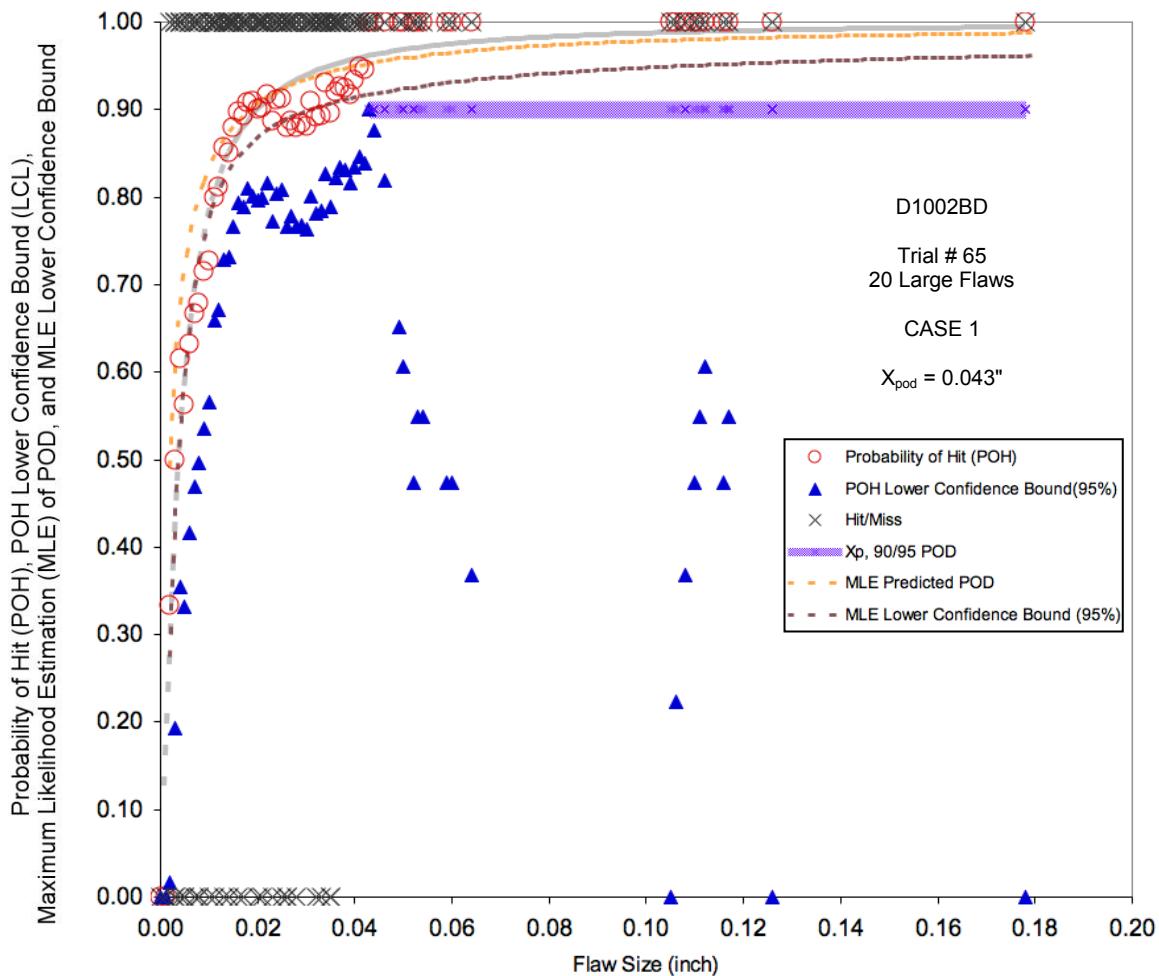
FIGURE 8. Trial #65 with 20 random large flaws from data set D1002BD. Probability of Hit (POH), POH Lower Confidence Bound (LCL), Maximum Likelihood Estimation (MLE) of POD and MLE Lower Confidence Bound versus flaw size.

## AGGREGATING THE INDIVIDUAL DOEPOD ANALYSIS RESULTS

The individual DOEPOD analysis results are aggregated into a final result by evaluating the probability of success (POS) that DOEPOD properly identifies that the POD at flaw sizes larger than the 90/95 POD flaw size is less than 90/95 POD (i.e., $POD_{Large\ Flaws} <$ 90/95 POD).

POS is estimated by applying binomial statistics to the results of each data set having the same number of randomly selected and unique large flaws. As stated earlier, the use of binomial statistics requires that four elements be true if a statistical variable is described by a binomial distribution: (1) The number of trials, N, is to be fixed. N = 76, is the number of runs of DOEPOD. (2) Each observation, i.e., DOEPOD analysis result on a randomly generated data set, is independent, (3) Each observation (DOEPOD analysis result) represents one of two outcomes (success or fail). Any results other than CASE 1 is a success and CASE 1 is a failure, and (4) The true Probably of Success (POS) that DOEPOD identifies a success (i.e., not CASE 1) is the same for each possible outcome. The POS is expected to increase as the number of randomly selected flaws is increased for fixed N. The selection of more than one unique large flaw from the entire set of large flaws represents a stochastic process. The large flaws are unique and only allowed to be selected once during each random selection of large flaws that are needed to create a data set. That is, as large

flaws are selected, there are a fewer number of large flaws from which to choose. The requirement that POS is the same for each outcome remain satisfied when the number of large flaws chosen is fixed for each of N trials.

In this Monte Carlo test there are 76 data sets with the same number of randomly selected flaws for each of the original 2 data sets (A6003H, D1002BD), or 76 trials with either a fail (CASE 1) or a success (not CASE 1). The ratio of (success)/ (number of trials) is a proportion and is an estimate of the probability that DOEPOD is successful in identifying the occurrence of $POD_{Large\ Flaws} < 90/95\ POD$, where the lower bound (LCL) on POS at 95% confidence is also determined. A 90/95 POS indicates that there is a 95% chance that the true POS is greater than the 90%.

A summary of the DOEPOD analysis for both sets of 2584 random data files is shown in Figure 9. The POS exhibits a different structure between the two data sets, and this is expected since the distribution of large flaw sizes between the two data sets are different.
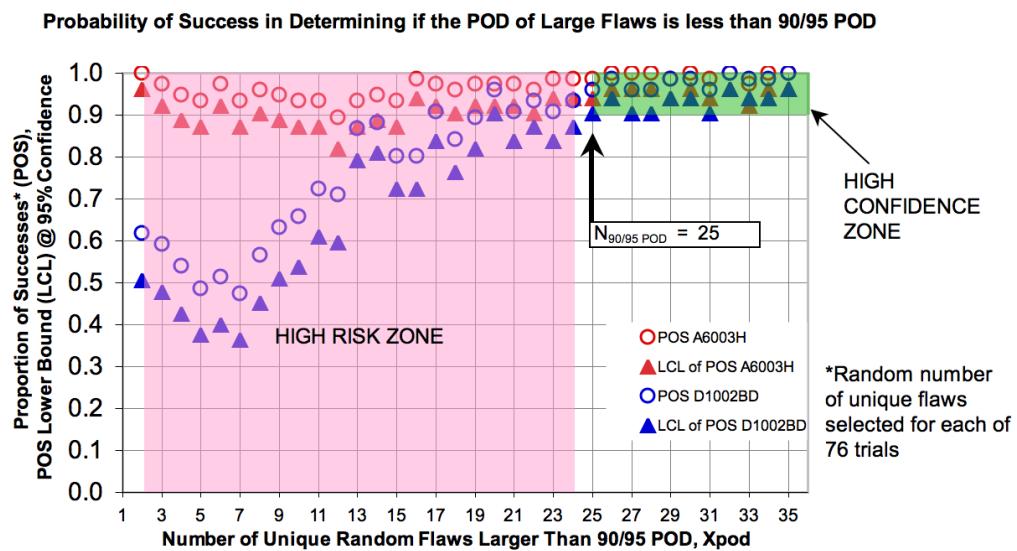
Probability of Success in Determining if the POD of Large Flaws is less than 90/95 POD

HIGH CONFIDENCE ZONE: The number of flaws (with sizes larger than the 90/95 POD flaw size) required to demonstrate that 90/95 POD (or greater) exists for flaw sizes larger than the 90/95 POD flaw size, Xpod. Required when new NDI or enhanced NDI technologies are being evaluated.

HIGH RISK ZONE: 90/95 POD for flaw sizes larger than 90/95 POD flaw size, Xpod, is not established by adding theses number of flaws with sizes larger than the 90/95 POD flaw size. The number of larger flaws may be accepted, with justification, when conventional or derivative NDI technologies are being evaluated.

N90/95 POD: The minimum number of larger flaws (N90/95 POD = 25) required to demonstrate that there is a 90/95 Probability of Success (POS) in determining if PODLarge Flaws < 90/95 POD.

FIGURE 9. Summary of the DOEPOD analysis for both sets of 2584 random data files.

Adding 25 or more random flaws with flaw sizes exceeding the 90/95 POD flaw size yields a 90/95 probability of success (POS) that DOEPOD will identify the occurrence of $POD_{Large\ Flaws} < 90/95\ POD$. Adding 25 ($N_{90/95\ POD} = 25$) or more random and unique flaws with flaw sizes exceeding the 90/95 POD flaw size represents a successful large flaw evaluation test in the HIGH CONFIDENCE ZONE shown in the Figure 9. This test should be considered as mandatory for all evaluations of new or enhanced NDI technologies.

In contrast, adding less than 25 random flaws with flaw sizes exceeding the 90/95 POD flaw size yields a LCL/95 probability of success (POS) that DOEPOD will identify the occurrence of $POD_{Large\ Flaws} < 90/95\ POD$. Since LCL may be less than 0.90, and this represents added risk (HIGH RISK ZONE) shown in the figure above. Therefore, adding less than 25 random flaws with flaw sizes exceeding the 90/95 POD flaw size should only be considered when justification is provided and when evaluating conventional or derivative NDI technologies.

There is a different POS trend observed between the A6003H and D1002BD data sets. The origin of the difference is identified by examination of the distribution of large flaw sizes. A6003H has large flaws grouped together and not uniformly spaced above the 90/95 POD flaw size. The D1002BD data set exhibits a fairly uniform distribution of large flaws distributed above the 90/95 POD flaw size.

In order to provide the most general stringent test for validation of large flaws, it is appropriate to identify the data sets similar to D1002BD as the preferred large flaw size distribution. That is, flaws above the 90/95 POD flaw size need to be uniformly distributed in sizes between the 90/95 POD flaw size and largest flaw size. The definition of "uniformly" is subjective, however, the coefficient of variation, CV, may be used to test for degree of the uniformity distribution. CV is the ratio of the standard deviation of large flaw sizes greater than the 90/95 POD flaw size to the mean of the large flaw sizes greater than the 90/95 POD flaw size,

$$Coefficient\ of\ Variation,\ CV \equiv \frac{Standard\ Deviation\ of\ Large\ Flaws\ Sizes}{Mean\ of\ Large\ Flaws\ Sizes}$$

DOEPOD provides guidance on the acceptable values of CV. Optimum is defined here to have large flaws with sizes equally spaced from the 90/95 POD flaw size, X_{pod}, to the largest flaws size, X_L. Data sets with a CV less than 0.33 are not sufficiently uniform and exhibit narrow groupings of flaws. When uniformly space flaws are considered, a CV of 0.34 is identified as the optimum for D1002BD, while the actual CV for this data set is 0.39. Large flaws with a CV greater than 0.56 are not sufficiently uniform and exhibit skewed groupings of flaws. This CV is observed for the data set A6003H, while the optimum CV when considering uniformly spaced large flaws for this data set is 0.40. An examination of the entire of the data files in the NTIAC Capabilities Data Book yields the optimum CV to be in the range 0.337 – 0.506.

The requirements for 25 uniformly distributed large flaws yielding a CV in the range of 0.33 – 0.51 for these large flaws is added as a requirement to reach CASE 1 in DOEPOD Prerelease v.1.0.3 This requirement assures a 90/95 POS or greater will be identified if it exists and therefore, DOEPOD Prerelease v.1.0.3 is only allowed to identify CASE 1 when in the high confidence zone in the chart above.

Summarizing the above Phase 2 results. A minimum of 25 randomly selected uniquely different flaw sizes larger than the 90/95 POD flaw size, that span the range to the largest flaw size, are required for validating that 90/95 POD exists in the range from the 90/95 POD flaw size to the largest flaw size. DOEPOD Prerelease v.1.0 is upgraded to v.1.0.3 in order to assure validation that the POD is observed at 90/95 POD or greater for flaws larger than 90/95 POD flaw size.

SUMMARY

It has been shown that the DOEPOD analysis methodology always yields a conservative value of the 90/95 POD flaw size, when it is observed and compared to predicted MLE POD flaw size. Including 25 or more random flaws with uniformly spaced flaw sizes exceeding the 90/95 POD flaw size is required to determine if the 90/95 POD or greater also exists for flaws exceeding the 90/95 POD flaw size. DOEPOD Prerelease v.1.0.3 is validated to provide a conservative value of the 90/95 POD flaw size, and further validates when the POD is increasing with flaw sizes greater than the 90/95 POD flaw size.

[1] E.R. Generazio, Design of Experiments for Validating Probability of Detection Capability of NDT Systems and for Qualification of Inspectors, Materials Evaluation, Vol 67, No 6, pgs 730-738, June 2009; 34th Annual Review of Progress in Quantitative Nondestructive Evaluation, July 2007, and NASA patent disclosure LAR-17656-1

[2] B. G. W. Yee, et al, Assessment of NDE Reliability Data, NASA-CR-134991, Oct. 1976

[3] Ward D. Rummel, Recommended Practice for Demonstration of Nondestructive (NDE) Reliability on Aircraft Production Parts, Materials Evaluation, vol 40, August 1982

[4] Standard NDE Guidelines and Requirements for Fracture Control Programs, MSFC-STD-1249, National Aeronautics and Space Administration, September 1985.

[5] NASA-STD-5009, "Nondestructive Evaluation Requirements for Fracture Critical Metallic Components", April 7, 2008.

[6] NDE Capabilities Data Book, 3rd ed., Nov. 1997, NTIAC DB-97-02

[7] Mil-HDBK-1823 Nondestructive Evaluation System Reliability, 30 April 1999

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. REPORT DATE *(DD-MM-YYYY)*	2. REPORT TYPE	3. DATES COVERED *(From - To)*
01- 03 - 2015	Technical Memorandum	

4. TITLE AND SUBTITLE

Directed Design of Experiments for Validating Probability of Detection Capability of NDE Systems (DOEPOD) Manual v.1.2

5a. CONTRACT NUMBER

5b. GRANT NUMBER

5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)

Generazio, Edward R.

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER

724297.40.44.07

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)

NASA Langley Research Center
Hampton, VA 23681-2199

8. PERFORMING ORGANIZATION REPORT NUMBER

L-20520

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)

National Aeronautics and Space Administration
Washington, DC 20546-0001

10. SPONSOR/MONITOR'S ACRONYM(S)

NASA

11. SPONSOR/MONITOR'S REPORT NUMBER(S)

NASA-TM-2015-218696

12. DISTRIBUTION/AVAILABILITY STATEMENT

Unclassified - Unlimited
Subject Category 38
Availability: NASA STI Program (757) 864-9658

13. SUPPLEMENTARY NOTES

14. ABSTRACT

Directed Design of Experiments for Validating Probability of Detection Capability of NDE Systems (DOEPOD)Manual v.1.2 The capability of an inspection system is established by applications of various methodologies to determine the probability of detection (POD). One accepted metric of an adequate inspection system is that there is 95% confidence that the POD is greater than 90% (90/95 POD). Design of experiments for validating probability of detection capability of nondestructive evaluation (NDE) systems (DOEPOD) is a methodology that is implemented via software to serve as a diagnostic tool providing detailed analysis of POD test data, guidance on establishing data distribution requirements, and resolving test issues. DOEPOD demands utilization of observance of occurrences. The DOEPOD capability has been developed to provide an efficient and accurate methodology that yields observed POD and confidence bounds for both Hit-Miss or signal amplitude testing. DOEPOD does not assume prescribed POD logarithmic or similar functions with assumed adequacy over a wide range of flaw sizes and inspection system technologies, so that multi-parameter curve fitting or model optimization approaches to generate a POD curve are not required. DOEPOD applications for supporting inspector qualifications is included.

15. SUBJECT TERMS

Binomial; Likelihood; NDE; NDI; NDT; Nondestructive; POD; Probability; Probability of Detection; Quality

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE			STI Help Desk (email: help@sti.nasa.gov)
U	U	U	UU	89	19b. TELEPHONE NUMBER *(Include area code)* (757) 864-9658

Standard Form 298 (Rev. 8-98)
Prescribed by ANSI Std. Z39.18