DIRECTED DESIGN OF EXPERIMENTS FOR VALIDATING PROBABILITY OF DETECTION CAPABILITY OF NDE SYSTEMS (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 observance 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.

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

INTRODUCTION

Directed design of experiments for probability of detection (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 also yields a quantitative measure of the lower confidence bound (value). 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 (values) of the observed estimated POH. The binomial distribution has been used previously for determining POD by observation of occurrences. Prior work[1, 2] 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 [1, 2, 3] method, in combination with validation that the POD is increasing with flaw size, in order to meet the requirements of MSFC-STD-1249 [3] and NASA-STD-(I)-5009 [4]. 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 [5] is estimated or predicted. It is noted that the combined statistical procedures described here require further validation by Monte Carlo simulation or similar tests.

DOEPOD CONCEPTS

DOEPOD is based on the application of the binomial distribution to a set of flaws that have been grouped into classes, where each class has a width. The classes are allowed to vary in width and start at 0.001" and increase by 0.001" increments. Classes start at the largest flaws and move toward the smallest flaws. Flaw size is referred to throughout the subsequent text as a "class length". Class length is used here in order 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 flaws in the data set. The largest flaw in any class width group is assigned as the identifier of the group. The next moving class width group is identified by decrementing the upper and lower class lengths by 0.001". DOEPOD evaluates the lower confidence bound obtained from any class width group. If the 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 or greater level. DOEPOD provides requirements for obtaining 90/95 POD at a flaw size. Directed DOEPOD also 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 KEY DEFINITIONS

X_{Best_LCL}	Class length exhibiting the maximized lower confidence bound (LCL).
X_{pod}	Class length at which the LCL is 0.90 or greater (90/95 POD).
$X_{poh=1}$	There are no Misses 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 samples, 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 samples in a test, and requirement (1) is satisfied. The definition of similar flaws remains vague and good engineering judgment must be made. Observations are made independently and do not depend on the result of the previous test and requirement (2) is satisfied. Weighting functions are not explored here, but will be addressed in subsequent presentations on DOEPOD. 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.

If the true POH is the same for each outcome, then the probability of observing X Hits after N trials, when the binomial distribution describes the behavior of the count variable X, is given by $POH_N(X)$. Example observations are shown as open circles in Figure 2. There are conditions or constraints that are made on the DOEPOD analysis and data interpretation that assists in assuring that the probability is sufficiently similar over the class widths of interest.

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}, 0.0147", and larger, the observed estimated POH is 1.0 (100%), and at 0.0147" the lower confidence bound (LCL, filled triangle) is 0.912. The class width for the estimated POH at 0.0147" is 0.004" 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 0.004" at a class length of 0.0147". 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 at 100% for all class lengths above 0.0147".

For class lengths below 0.0107" 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. DOEPOD uses estimated POH less than 100% for guidance for further sample selection or for identifying optional 90/95 POD class lengths. If the guidance is executed successfully, and the observed lower confidence bound is equal to or greater than 0.9, then it is proposed here that validation of the inspection capability may be obtained. The presence of mixed true POH existing within the class widths used are progressively minimized at the validation and larger class lengths by increased observations of Hits. Since, DOEPOD requires validation that 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 requirement (4) is approximated. A curve estimating POH is shown in Figure 3. This estimated POH curve is the chi-square best fit to a log-odds [5] model and is not part of the DOEPOD analysis, however, the curve is displayed for visualization only and not for supporting system validations.

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 from [6]

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

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

where α is the required confidence level (95%) and $F_{\alpha}(f_1,f_2)$ is obtained from tables of the F-distribution [7]. For the procedure and flaw size in this example, and at a 95% confidence level, if LCL = 0.9, then the following statement applys: "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."

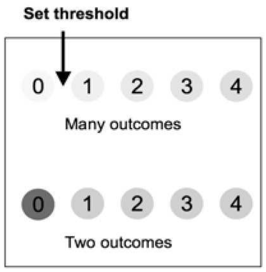

FIGURE 1. Binomialization of test data.

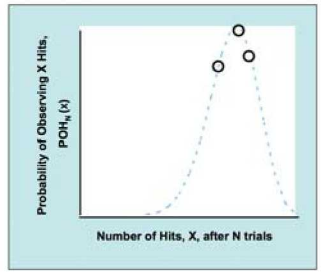

FIGURE 2. Probability of observing X Hits after N trials.

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. The differences in the cases are described later. Due to manuscript limits not all Cases are shown or here. Case 1 is the best case and is shown in Figure 3. 90/95 X_{pod} is reached at a class length, and there are Misses only below X_{pod} (i.e., estimated POH =1 everywhere greater than X_{pod}). Further validation is still required in order to verify that the POD is actually increasing with increasing class length. The DOEPOD recommendations are 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 need to be included in the DOEPOD analysis. Case 1 must be achieved before validation of the inspection system can occur. It is noted here that other approaches to validate that 90/95 POD (or greater) also exists for flaws larger than X_{pod} are being explored. Including the of addition of 27 flaws at equally distributed class lengths between X_{pod} and X_L, exclusively, grouping of flaws by number, and procedures for using good engineering judgment supported by data obtained from similar systems.

Case 2 is the most interesting case and is shown in the Figure 4. In this case, 90/95 X_{pod} is reached at a class length. There are Misses below X_{pod} and some Misses above X_{pod}. Since Misses exist at class lengths, X_i, above X_{pod}, then these greater lengths need to be validated. The DOEPOD recommendations are listed as two options that may be executed. Successful execution of the recommendations will transition this Case 2 to Case 1. The recommendations are: (a) add samples of class length X_i where estimated POH<1 (Figure 4, Table A). Starting from largest class length, X_i, and work toward small class lengths until reaching an acceptable X_{pod} or reaching X_{pod}, or (b) add samples of class length X_i where estimated POH=1 (Figure 4, Table B) and accept a larger X_{pod} class length at any of the X_i . This acceptance is valid as long as any existing larger class lengths where estimated POH<1 are shown [via (a) above] to be at 90/95 X_{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, DOEPOD recommendations are to satisfy the smallest X_{pod} in Figure 4, Table B that is greater than the largest X_{pod} in Figure 4, Table A, and/or the largest X_{pod} in Figure 4, Table A. There is a caution when adding samples to an already existing data set. It is recommend that, when adding samples to an existing set that the inspection of the entire set of samples be done before performing a DOEPOD analysis. DOEPOD Analysis Summary and Recommendations of Cases 4, 5, 6, and 7 are shown in

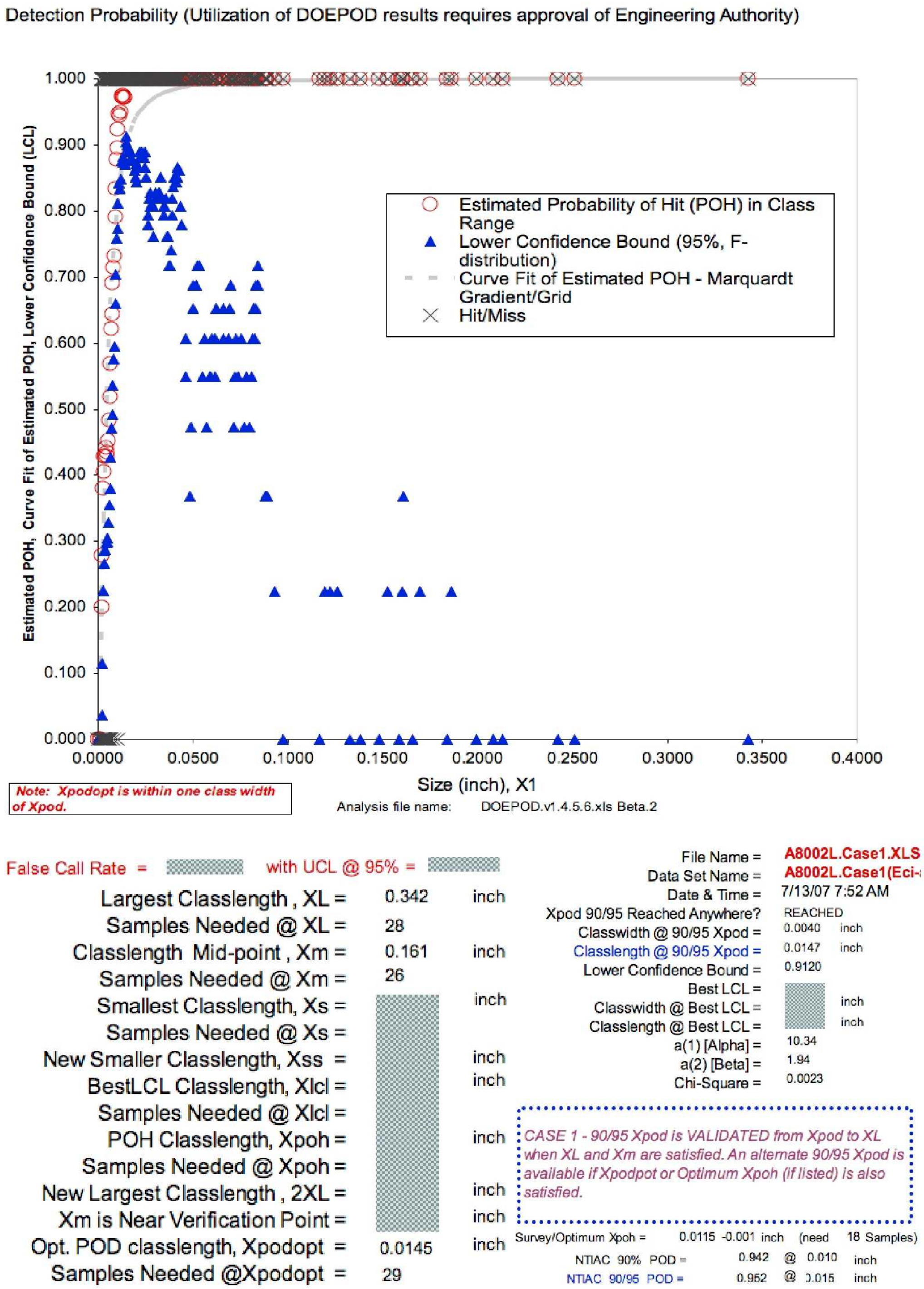

FIGURE 3. Case 1 example output format of DOEPOD analysis.

Table C, and an example analysis for Case 6 is shown in Figure 5. Survey data sets have an insufficient number of samples for unconstrained class width optimization. DOEPOD recommendations are to add samples at Survey/Optimum X_{poh} and X_L.

DOEPOD FALSE CALL ANALYSIS

False Calls are handled similarly except the upper confidence limit is used. Test samples with no flaws present should be included in the DOEPOD data set 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,

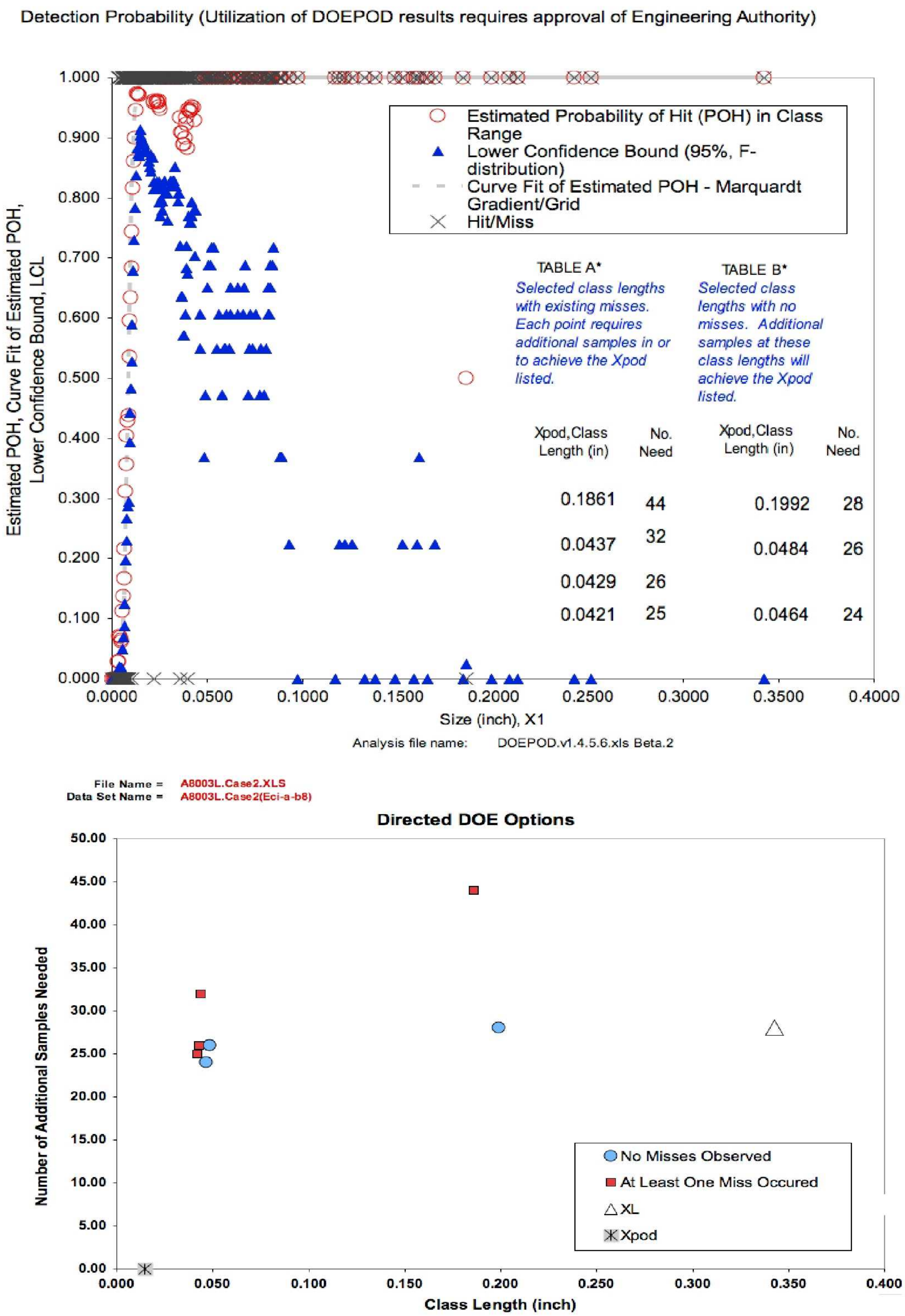

FIGURE 4. Case 2 example of DOEPOD analysis recommendations.

Table C. DOEPOD Analysis Summary and Recommendations of Cases 1, 2, 4, 5, 6, and 7.

	Is 90/95 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 $X_L/3$?	DOEPOD Analysis Summary and Recommendations
Case #1	●	●	●	NA	Validation at X_{pod} has been reached. Recommendation: If not established, validation at large- and mid-flaw sizes is required.
Case #2	●	●	○	NA	Validation at X_{pod} has been reached, however, there are Misses above X_{pod}. Recommendation: Additional validation at identified flaw sizes is required.
Case #4	○	●	●	NA	Validation has not been reached. Recommendation: Increase number of samples at $X_{POH=1}$ or X_{Best_LCL}.
Case #5	○	●	○	●	Validation has not been reached and there are Misses above X_{Best_LCL}. Recommendation: Increasethe number of samples at $X_{POH=1}$.
Case #6	○	●	○	○	Validation has not been reached. The POH is fluctuating above X_{Best_LCL} and X_{poh} is greater than $X_L/3$. The inspection system is unstable for the flaw size range analyzed. Recommendation: Increase the flaw size range by a factor of two.
Case #7	○	○	NA	NA	Validation has not been reached. The inspection system is unstable for the entire flaw size range analyzed. Recommendation: The inspection system may not be appropriate or increase the flaw size range by a factor of two.

● = YES ○ = NO NA = NOT APPLICABLE

Detection Probability (Utilization of DOEPOD results requires approval of Engineering Authority)

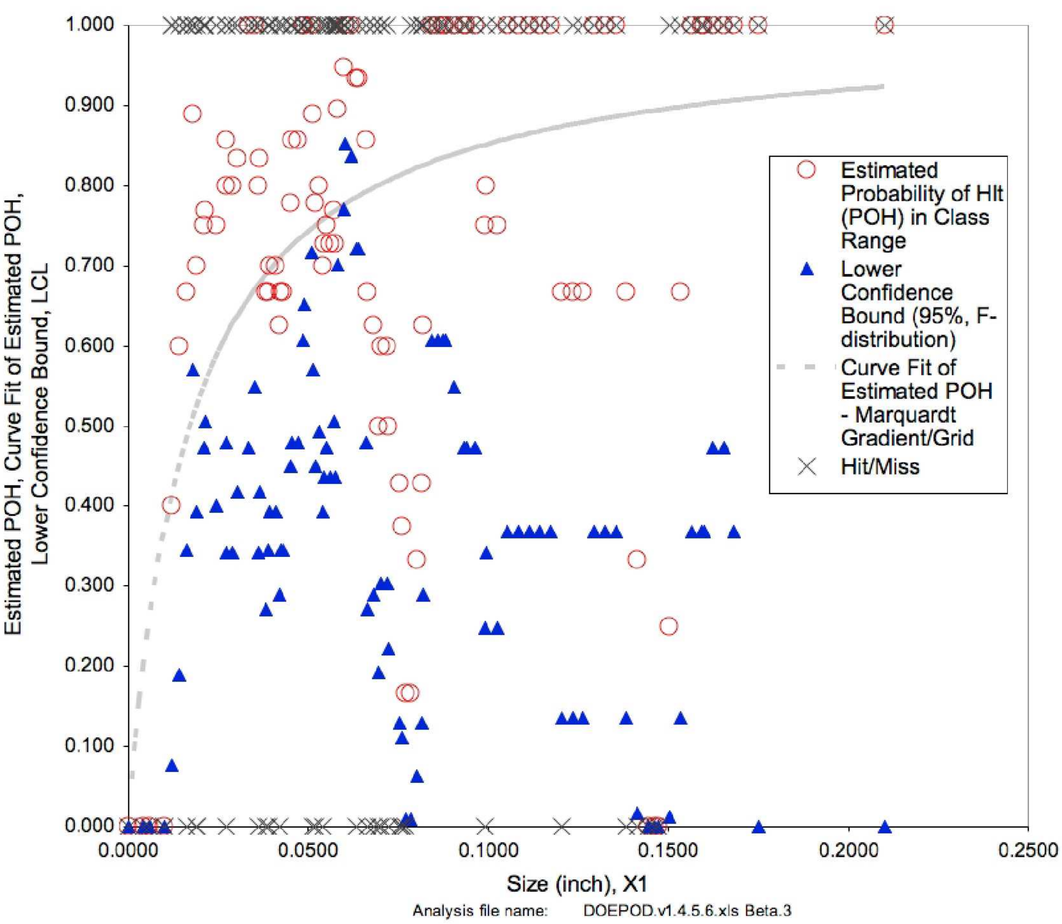

FIGURE 5. Case 6 example of DOEPOD analysis.

90/95 X_{pod} may be reached at cost of increasing false call rate. False calls should not be accepted as is without first addressing the cause of the false call and identifying procedures to remove false calls. The estimated false call rate is given by,

$$FalseCallRate = \frac{Numberof\ FalseCalls\ (X)}{Numberof\ FalseCall\ Opportunites\ (N)} \tag{3}$$

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

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

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."

SUMMARY

In summary, the following have been presented; the concept for binomialization of test data, the process for determining observed probability of Hit (estimated POH) and associated confidence bounds, the utilization of moving class width to group flaws and for flaw class width optimization, the classification of POD Cases and directed actions or requirements needed to validate inspection systems, and the false call rate and confidence bounds. Future work includes distribution of the DOEPOD software for Beta testing, interfacing DOEPOD with validated software implementations of MIL-HDBK-1823 and model assisted POD approaches as companion tools, comparisons with other POD methodologies, and addressing very limited data sets when 90/95 X_{POD} can never be reached, and communicating those risks.

ACKNOWLEDGEMENTS

The author wishes to thank Ward D. Rummel of D&W Enterprises, for his unwavering support, encouragement, recommendations, and testing of DOEPOD software, and William Q. Meeker of Iowa State University for reviewing the manuscript and clarifying statistical concepts utilized in DOEPOD.

REFERENCES

[1] B. G. W. Yee, et al, Assessment of NDE Reliability Data, NASA-CR-134991, Oct. 1976
[2] Ward D. Rummel, Recommended Practice for Demonstration of Nondestructive (NDE) Reliability on Aircraft Production Parts, Materials Evaluation, vol 40, August 1982
[3] Standard NDE Guidelines and Requirements for Fracture Control Programs, MSFC-STD-1249, National Aeronautics and Space Administration, September 1985.
[4] NDE Requirements for Fracture Critical Metallic Components, NASA-STD-(I)-5009, September 2006
[5] NDE Capabilities Data Book, 3rd ed., Nov. 1997, NTIAC DB-97-02
[6] Statistical Theory with Engineering Applications, A. Hald, 1952
[7] Metallic Materials and Elements for Aerospace Vehicle Structures, MIL-HDBK-5H, Table 9.6.4.9, pg 9-251, December 1998.