

# Optimizing Probability of Detection Point Estimate Demonstration

Ajay M. Koshti, D. Sc. PE  
NASA Johnson Space Center, Houston

March 25-29, 2017

**SPIE Smart Structures/NDE 2017**  
**Portland, OR, U. S. A.**

- Acknowledgements
- References
- Introduction
- Background, Two Types of POD Datasets
- Three Cases of POD Datasets and POD Objectives
- Case 1: a set of 29 flaw sizes that are uniformly spaced with median and mean at  $a_{90}$
- Case 2: Probability of passing demonstration equals 5% and 50%
- Case 3: Comparison of various  $a_{90/95}$  estimates from simulated data
- Conclusions



# Acknowledgements

---

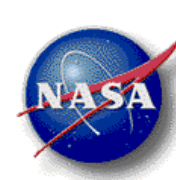


- David Stanley (NASA Johnson Space Center),
  - Charles Nichols (NASA White Sands Test Facility), and
  - Michael Suits (NASA Marshall Space Flight Center)
- provided valuable review comments to this work.

- [1] “Nondestructive Evaluation System Reliability Assessment,” MIL-HDBK-1823A, United State of America, Department of Defense Handbook, (April 2009).
- [2] Annis, C., “mh1823” POD Software (v4.2.5),” Statistical Engineering, [www.statisticalengineering.com](http://www.statisticalengineering.com), (Oct. 2014).
- [3] “Standard Practice for Probability of Detection for Hit/Miss Data,” ASTM E 2862-12, ASTM International.
- [4] Koshti, A. M., “Modeling the X-ray Process and X-ray Flaw Size Parameter for POD Studies,” Proc. SPIE 9063, Nondestructive Characterization for Composite Materials, Aerospace Engineering, Civil Infrastructure, and Homeland Security, (2014).
- [5] Koshti, A. M., “Simulating the X-ray Image Contrast to Set-up Techniques with Desired Flaw Detectability,” Structural Health Monitoring and Inspection of Advanced Materials, Aerospace, and Civil Infrastructure, Proc. of SPIE Vol. 9437, (2015).
- [6] Koshti, A. M., “Methods and Systems for Characterization of an Anomaly Using Infrared Flash Thermography,” U.S. Patent US8577120 B1, (Nov. 2013).

- Probability of detection (POD) analysis is used in assessing reliably detectable flaw size in nondestructive evaluation (NDE).
- MIL-HDBK-1823<sup>1</sup> and associated mh1823<sup>2</sup> POD software gives most common methods of POD analysis.
- Real flaws such as cracks and crack-like flaws are desired to be detected using these NDE methods.
- A reliably detectable crack size is required for safe life analysis of fracture critical parts.
- The paper provides discussion on optimizing probability of detection (POD) demonstration experiments using Point Estimate Method. POD Point estimate method is used by NASA for qualifying special NDE procedures. The point estimate method uses binomial distribution for probability density. Normally, a set of 29 flaws of same size within some tolerance are used in the demonstration.
- The optimization is performed to provide acceptable value for probability of passing demonstration (PPD) and achieving acceptable value for probability of false (POF) calls while keeping the flaw sizes in the set as small as possible.

- Traditionally largest flaw size in the set is considered to be a conservative estimate of the flaw size with minimum 90% probability and 95% confidence.
- The Point Estimate flaw size is denoted as  $a_{90/95PE}$ .
- The paper investigates relationship between range of flaw sizes in relation to  $a_{90}$ , i.e. 90% probability flaw size, to provide a desired PPD.
- The range of flaw sizes is expressed as a proportion of the standard deviation of the probability density distribution.
- Difference between median or average of the 29 flaws and  $a_{90}$  is also expressed as a proportion of standard deviation of the probability density distribution.
- In general, it is concluded that, if probability of detection increases with flaw size, average of 29 flaw sizes would always be larger than or equal to  $a_{90}$  and is an acceptable measure of  $a_{90/95PE}$ .
- If NDE technique has sufficient sensitivity and signal-to-noise ratio,
  - then the 29 flaw-set can be optimized to meet requirements of minimum required PPD,
  - maximum allowable POF,
  - requirements on flaw size tolerance about mean flaw size and
  - flaw size detectability requirements.
  - The paper provides procedure for optimizing flaw sizes in the point estimate demonstration flaw-set.



# Background, Two Types of POD Datasets



- MIL-HDBK-1823<sup>1</sup> and associated mh1823<sup>2</sup> software cover two types of datasets.
- First type of dataset is signal response  $\hat{a}$  (read as a-hat) versus flaw size “ $a$ ”.
  - The  $\hat{a}$  (y-axis) versus “ $a$ ” (x-axis) data may be transformed using logarithm function along appropriate axes, if needed, to create linear correlation around the decision threshold,  $\hat{a}_{decision}$ .
  - A generalized linear model (GLM) is fitted to the transformed data for analysis.
  - Here, noise data is taken separately to define noise distribution.
  - Noise is same as signal response from part where there is no flaw.
  - Noise data is used to determine false call rate or probability of false calls (POF).
- Second type of dataset is called hit-miss data, which contains flaw size and corresponding detection result i.e. hit or miss.
  - Hit has numerical value of 1 and miss has numerical value of 0.
  - Here, false call data is noted to determine false call rate using Clopper-Pearson binomial distribution function.
  - Normally, POD increases with flaw size and POF decreases with flaw size.
  - POF value shall be within certain limit to prevent adverse impact on cost and schedule.
  - ASTM E 2862<sup>3</sup> also provides the hit-miss POD data analysis method that is consistent with MIL-HDBK-1823.



# Binomial Point Estimation of POD

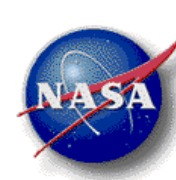


- Binomial point estimation of POD requires detection of all 29 flaws of same size out of 29 flaw detection opportunities, i.e. 29/29.
- Point estimation method is described by Rummel<sup>16</sup>. Another variation is detecting 45 flaws out of 46 flaw detection opportunities. We will only work on the 29/29 case here.
- These, hit-miss versus detection opportunity numbers come from the cumulative binomial distribution, which provides confidence value for given number of hits, opportunities and probability.
- This function is also given in Microsoft Excel. is calculated in Excel as follows,
  - BINOMDIST (28, 29, 0.9, TRUE) = 0.9529. For 29 hits in 29 opportunities for POD of 0.9, a confidence of 0.95
  - BINOMDIST (28, 29, 0.97637, TRUE) = 0.5 For 29 hits in 29 opportunities for POD of 0.9763, a confidence of 0.5
- ASTM E 2862 gives the **Clopper-Pearson binomial method** for constructing confidence intervals for proportions. It can be used for POD (lower bound) and POF (upper bound) confidence level e.g. 50%, 90% and 95%. The Clopper-Pearson upper 100(1- $\alpha$ ) % confidence bound for POF,  $p$  is,

$$P_U = \left\{ 1 + \frac{n-x}{(x+1) \cdot F_{(1-\alpha, 2x+2, 2n-2x)}} \right\}^{-1}, \quad (1)$$

- where  $x$  is misses,  $n$  is opportunities,  $\alpha$  is confidence level, and  $F(1-\alpha, 2x+2, 2n-2x)$  is the F-statistics with degrees of freedom  $(2x+2, 2n-2x)$  and  $P[F < F(1-\alpha, 2x+2, 2n-2x)] = 1-\alpha$ . This method is consistent with that used in MIL-HDBK-1823.



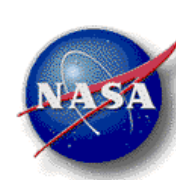


# Optimizing Point Estimate Method



- Usually, hardest-to-detect flaw configuration is used in the 29 flaw-set to envelope other less severe flaw detection cases.
- Point estimate method assures that  $a_{90/95PE} > a_{90}$  and normally the margin is assumed to be very large.
- But as  $a_{90/95PE}$  comes closer to true  $a_{90}$ , it becomes difficult to pass the demonstration i.e. Probability of passing demonstration (PPD) becomes lower. Moreover, the false call rate may become unacceptable.
- **Since, true  $a_{90}$  is mostly a characteristic of an NDE technique, the point estimate demonstration flaw sizes should take into account  $a_{90}$  for the technique if optimal  $a_{90/95PE}$  is desired.**
- Following requirements are needed to be met in optimizing a point estimate flaw-set.
  - 1. Probability of passing demonstration (PPD) test shall be more than a set value (i.e.  $\geq 50\%$ ).
  - 2. Probability of false (POF) calls shall be lower than a set value (i.e. 2%).
  - 3. Flaw size range shall be practical to achieve as flaw manufacturing process has inherent tolerance on flaw size.
  - 4.  $a_{90/95PE}$  shall be less than or equal to the design requirement flaw size (e.g. safe life initial flaw size).
  - 5. Optimal flaw size is based on following rules. If above requirements can be met, then smallest  $a_{90/95PE}$  flaw size is desirable. If only some of the above requirements can be met, then choose smallest  $a_{90/95PE}$  flaw size that meets some of the chosen requirements e.g. design requirement, PPD or POF.

- This approach is based on hypothesis that  $\hat{a}$  versus “ $a$ ” curve-fit POD,  $\hat{a}$  versus “ $a$ ” mh1823 POD, hit-miss mh1823 POD approaches provide lower  $a_{90/95}$  flaw sizes compared to point estimate POD.
- Therefore, if these methods are used to determinate POD curves, perform noise analysis, choose decision threshold, and perform POF analysis, then this information can be used to optimize the point estimate 29 flaw-set.
- We would first check the hypothesis that average flaw size of the 29 flaws in the point estimate set provides  $a_{90,95PE}$  that is greater than true  $a_{90}$  of the data model and  $a_{90/95}$  of POD analysis models which assume that POD increases with flaw size.
- We would use mh1823  $\hat{a}$  versus “ $a$ ” analysis or hit-miss analysis to generate the POD curve. POD curves for the method are determined using a flaw-set per MIL-HDBK-1823.
- For  $\hat{a}$  versus “ $a$ ” analysis, decision threshold level (DTL) is needed. This level is based on achieving desired flaw detectability size but also on limiting POF as calculated using noise data and flaw response data.
- Noise data is taken and DTL versus POF curve is determined.
- Once, the DTL level is determined to provide acceptable POF, it is used to complete the POD analysis and PO curve equation is determined.



## Approach (continued)



- Similarly, for hit-miss analysis, if the NDE technique does not use a DTL, then data on false calls is obtained and analyzed per MIL-HDBK-1823 using Clopper-Pearson method.
- Here, we may reduce the POF by increasing size of a relevant indication which also then results in larger  $a_{90/95}$ .
- If the POF value is acceptable and  $a_{90/95}$  is acceptable, then we can proceed to design the point estimate set.
- Here we need to recognize that the point estimate flaw size is larger than true  $a_{90}$  by certain margin denoted by  $\Delta$  (delta).

$$a_{requirement} \geq a_{90/95PE} = a_{90} + \Delta \quad (2)$$



# Case 1: Flaw-set with Uniformly Distributed Sizes with Average Size of $a_{90}$

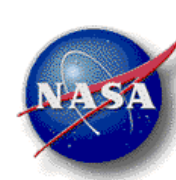


- Objective in Case 1 is to study effect of flaw range on PPD.
- The flaw-set has uniformly distributed flaws with median and average flaw size equal to  $a_{90}$ .
- We would use a POD model in the study. Standard deviation of a uniformly spaced flaw-set is 30.4% of the range.
- It is assumed that signal response  $\hat{a}$  relates to flaw size “ $a$ ” as follows.

$$\hat{a} = \beta_1 a + \beta_0 + \delta \quad (3)$$

- Although a linear relationship is chosen, other relationships as given in MIL-HDBK-1823 also apply.
- First, a symmetrical POD function curve based on error function (erf) is chosen. This is given by cumulative density distribution of a probability density function which is chosen to be a normal distribution.
- This meets the key assumption that POD increases with flaw size. Probability density function (PDF) in the form of normal distribution is given by,

$$f(a) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(a-\mu)^2}{2\sigma^2}}. \quad (4)$$



- POD function is given by cumulative density distribution (CDF) of the normal distribution PDF. It is given by,

$$g(a, \mu, \sigma) = \frac{1}{2} \left[ 1 + \operatorname{erf} \left( \frac{a - \mu}{\sigma \sqrt{2}} \right) \right], \quad (5)$$

- Where,  $\mu$  is mean and  $\sigma$  is standard deviation of the PDF and CDF functions. 90% POD is given by following expression,  $0.9 = g(1.2815, 0, 1)$ . Following CDF expression from Matlab is used.

$$g(a) = 0.5 [1 + \operatorname{erf}(C_1 a - C_2)] \quad (6)$$

$$C_1 = 1/\sigma\sqrt{2}, \text{ and} \quad (7)$$

$$C_2 = \mu/\sigma\sqrt{2}. \quad (8)$$

- We choose  $C_1 = 30.8$ ;  $C_2 = 7.8$  to generate the POD function.



# Assumed POD Function



- We calculate the following flaw sizes in arbitrary units.
- At  $g = 0.1$ ,  $a_{10} = 0.2239$ .
- At  $g = 0.9$ ,  $a_{90} = 0.283$ .
- Standard deviation can be calculated as,

$$\sigma = (a_{90} - a_{10}) / (2 \times 1.285) \sim 0.023 \quad (9)$$

- From Eq. 7, we can calculate the standard deviation in POD model as,

$$\sigma = \sqrt{2}C_1. \quad (10)$$

- From Eq. 8, mean used is calculated as,

$$\mu = C_2\sigma\sqrt{2}. \quad (11)$$

- **Here, we construct a set of 29 flaw sizes that are uniformly spaced with median and mean at  $a_{90}$ . Range  $R$  is given by,**

$$R = a_{max} - a_{min}, \quad (12)$$

- where,  $a_{max}$  and  $a_{min}$  = are the maximum and minimum flaw sizes in the set of 29 flaws.



**Case 1 condition: a set of 29 flaw sizes that are uniformly spaced with median and mean at  $a_{90}$**

- Range is given as percentage of standard deviation by following equation,

$$R_{\%} = 100R/\sigma. \quad (13)$$

- We can also express the standard deviation of flaw sizes as percentage of standard deviation by following equation.

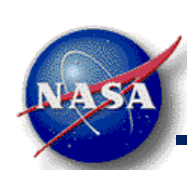
$$\sigma_{flaw\%} \cong 30R/\sigma. \quad (14)$$

- This allows us to investigate effect of range as percentage of standard deviation of POD function and therefore conclusions would be generically applicable, if POD functions are similar. We can simply multiply the range by 30% and get the corresponding standard deviation in the flaw sizes. Range and  $a_{max}$  and  $a_{min}$  relate to the median size as follows,

$$a_{min}, a_{max} = a_{median} \pm R/2. \quad (15)$$

- Spacing between flaws and the range are related by,

$$S = R/28. \quad (16)$$



# Average Probability of 29/29 Demonstration



**Case 1 Condition: a set of 29 flaw sizes that are uniformly spaced with median and mean at  $a_{90}$**

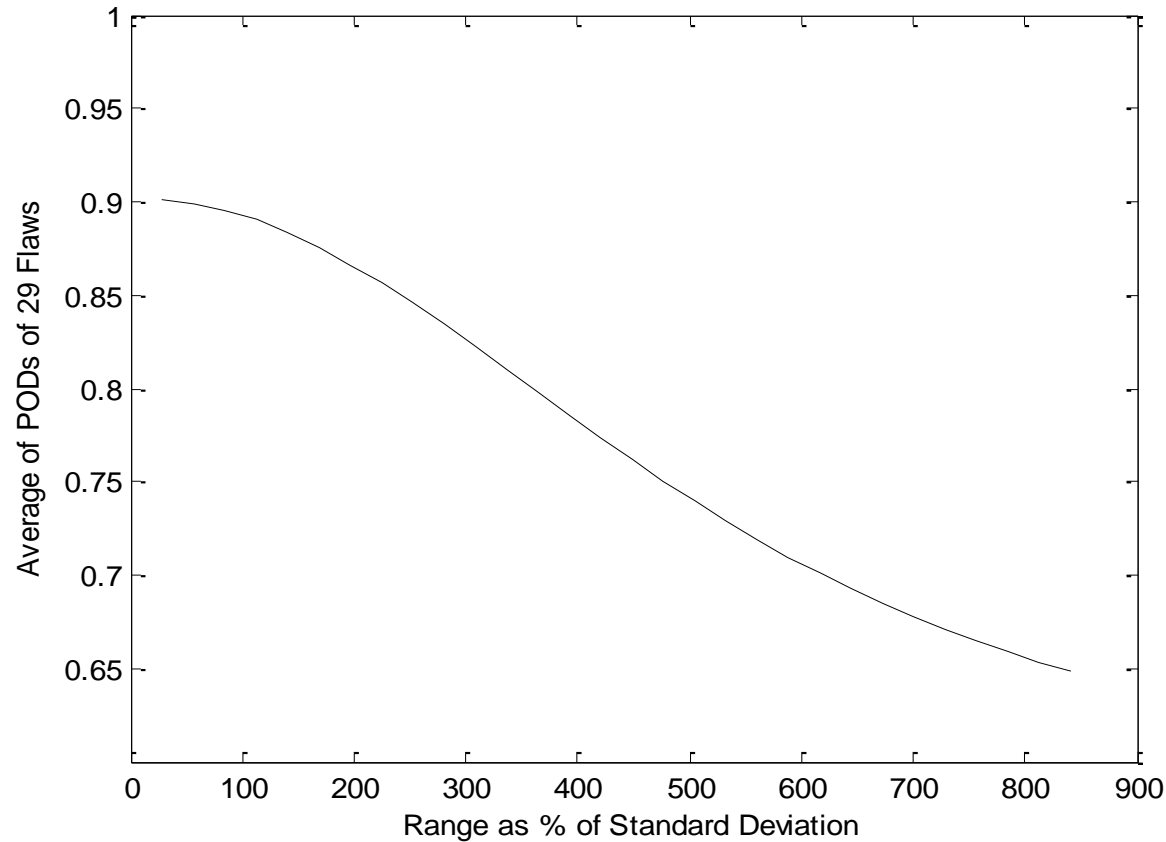


Fig. 1: Average probability of 29/29 demonstration

- Next we calculate average of PODs of the 29 flaws.
- Plot of average probabilities is shown in Fig. 1.
- At a range equal to standard deviation, the average probability 0.89.
- The average POD in above chart is either 0.9 or lower.



**Case 1 Condition: a set of 29 flaw sizes that are uniformly spaced with median and mean at  $a_{90}$**

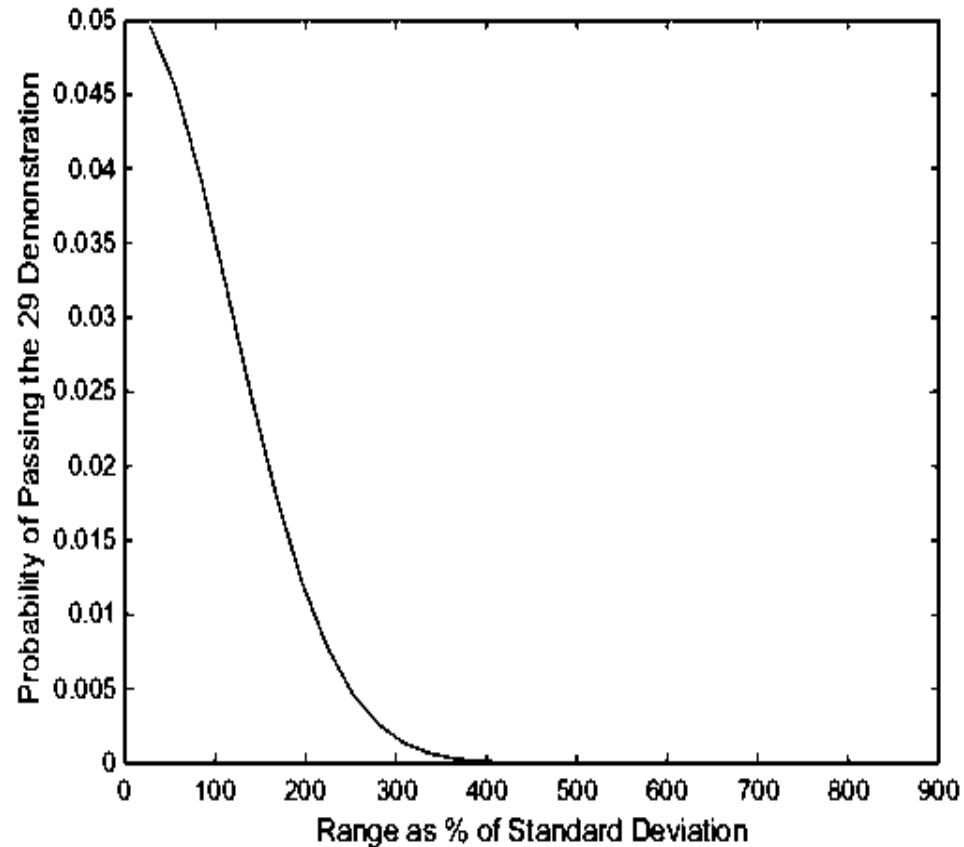


Fig. 2: Probability of passing the demonstration

- We also calculate product of PODs of 29 flaws which is same as the probability of passing (PPD) the 29/29 demonstration.
- The POD needs to be paired with its corresponding PPD for proper interpretation. Probability of passing the demonstration is given in Fig. 2.
- The above chart indicates that as range increases, it becomes harder to pass the 29/29 demonstration or PPD gets lower.
- Since, goal is to pass the demonstration, it is wise to have shortest range in flaw sizes.
- At a range equal to one standard deviation, PPD is 0.035 with confidence of 95%.
- This is also equivalent to PPD of 0.5 with confidence of 0.5.
- This PPD is too low for establishing successful demonstration protocol.
- We would only use 95% confidence in rest of the paper. At range equal to two standard deviations, PPD decreases from 0.05 to 0.012.
- Therefore, although there is no requirement on range of the flaws, if average flaw size is close to true  $a_{90}$  then flaw range would be very small too.
- Here we compute PPD versus PODs for single size flaw 29/29 demonstration by using following equation,

$$PPD_{SS} = 1 - \alpha \quad (18)$$

# NASA Probability of Passing Demonstration for Flaw Sizes in a Range and for Single Size



Case 1 Condition: a set of 29 flaw sizes that are uniformly spaced with median and mean at  $a_{90}$

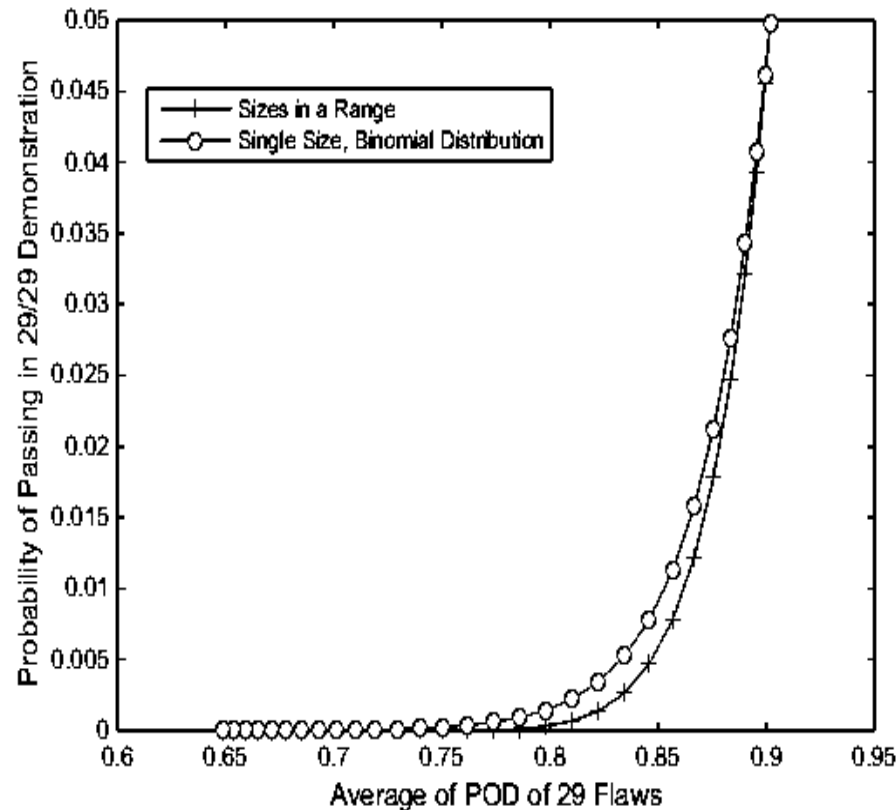


Fig. 3: Probability of passing 29/29 demonstration for flaw sizes in a range and for single flaw size

- Here, binomial distribution is used to compute confidence for the same average POD values.
- PPD versus POD for single size flaws is compared with the same for flaws in a range.
- For lower values of average POD, the range of flaw sizes is larger. See also Fig. 1.
- Fig. 3 indicates that for same average flaw size, PPD with flaws in a range is less than or equal to PPD for single size flaws.
- Therefore, average flaw size of a set of 29 flaws can be used as a more conservative  $a_{90,95}$  estimate compared to single size  $a_{90,95}$  estimate, provided POD increases with flaw size.
- If we increase the median flaw size to be greater than the true  $a_{90}$ , then it should be easier to pass the demonstration at the cost of increased  $a_{90,95}$  from the demonstration.
- This point is discussed in Case 2.

**Case 2 condition: Probability of passing demonstration equals 5%**

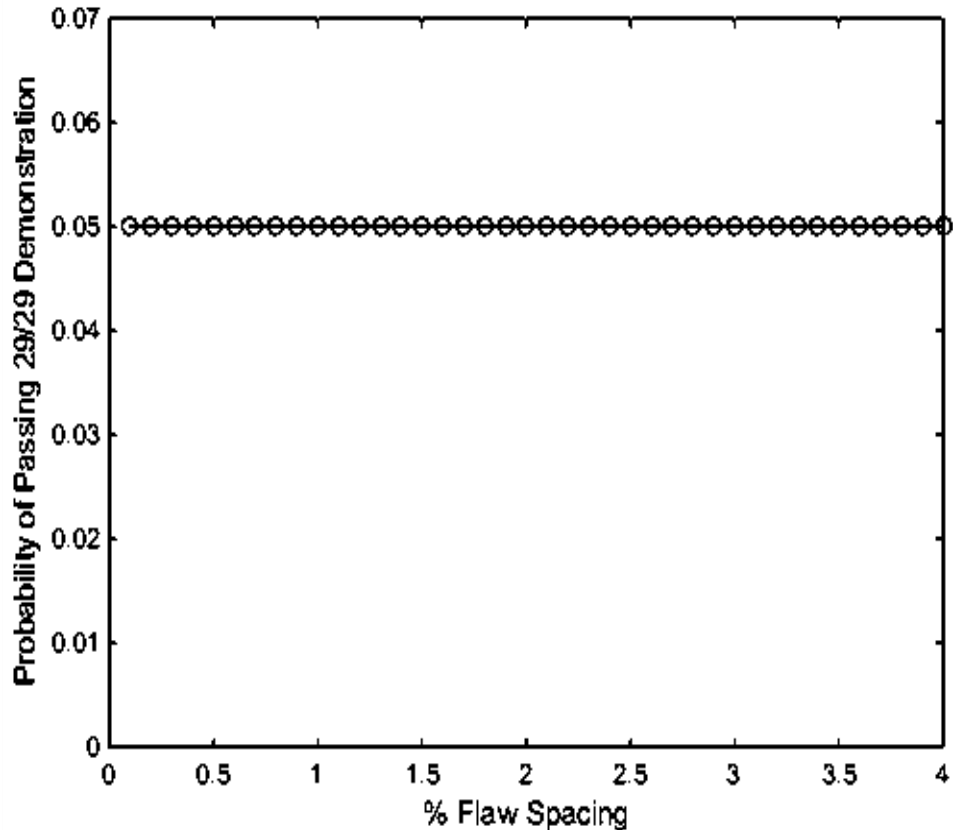


Fig. 4: Calculated PPD versus % spacing when optimal delta was calculated.

- In Case 2, our goal is to study relationship between the flaw size range of a uniformly distributed 29 flaw-set and the median (or average) flaw size in relation to  $a_{90}$  based on assumed POD curve to obtain a PPD of 0.05.
- We would continue with the same POD curve as in Case 1. Here, we define delta as difference between median size of the flaw-set and true or model input  $a_{90}$  as follows,

$$\Delta = a_{median} - a_{90} \tag{19}$$

- We express delta as percentage of standard deviation by,

$$\Delta_{\%} = 100 (a_{median} - a_{90}) / \sigma \tag{20}$$

- We choose flaw spacing or range as % of standard deviation of the POD curve and construct a uniformly distributed set of 29 flaws about a median flaw size providing probability of passing the demonstration  $\approx 0.05$ .
- Fig. 4 shows a number of % spacing values for which corresponding % delta values providing PPD of 0.05 were determined. See Fig. 4.



## Case 2: Median Flaw Size $\geq a_{90}$



### Case 2 condition: Probability of passing demonstration equals 5%

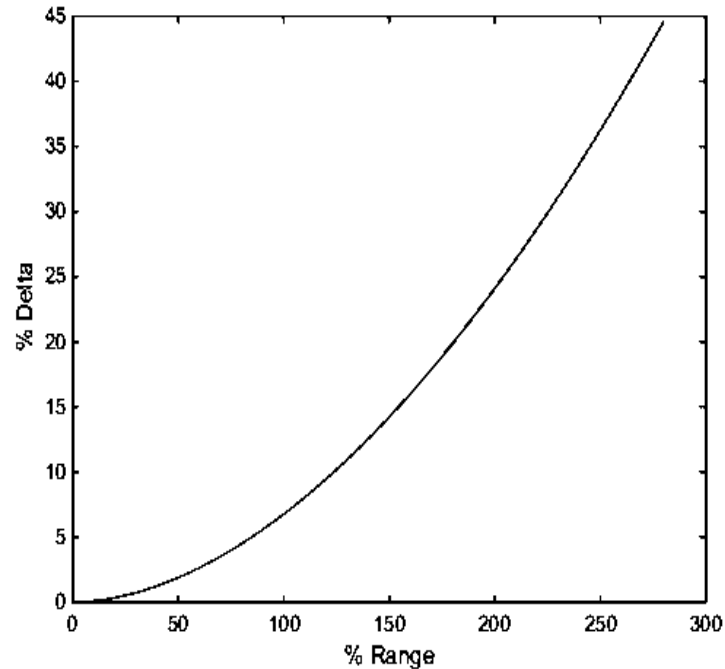


Fig. 5: % delta versus % range providing PPD of 0.05.

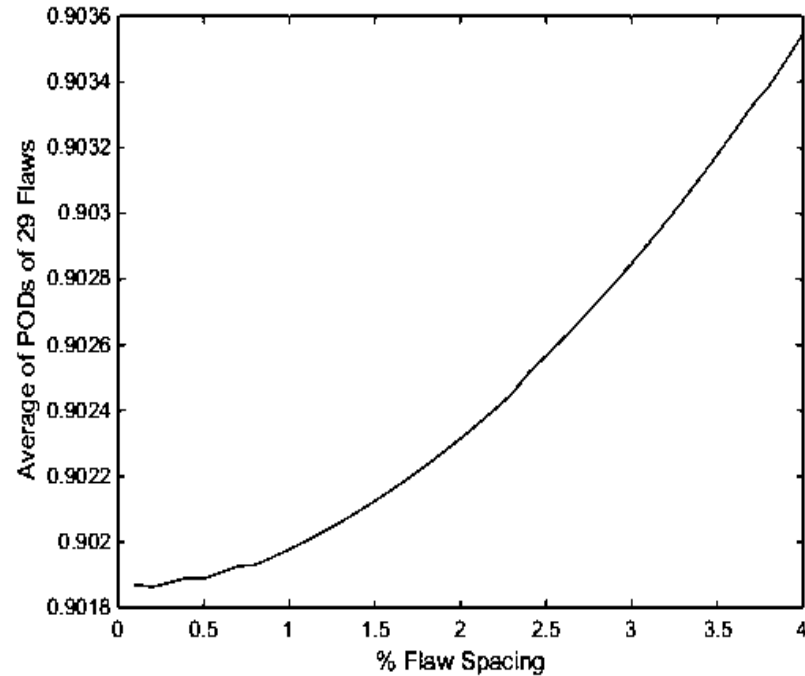


Fig. 6: Average POD versus % flaw spacing providing PPD of 0.05

- The % delta versus % range providing PPD of 0.05 are provided in Fig. 5.
- As expected, the average POD would increase with flaw spacing when PPD is held constant.
- In order to satisfy a condition that PPD of uniformly spaced 29 flaws is same as PPD of single sized 29 flaws with POD of 0.9, % delta and % range get related. % delta increases as % range increases. In this study, % delta and % spacing are approximately related by following equation.



# Case 2: Median Flaw Size $\geq a_{90}$



## Case 2 condition: Probability of passing demonstration equals 50%

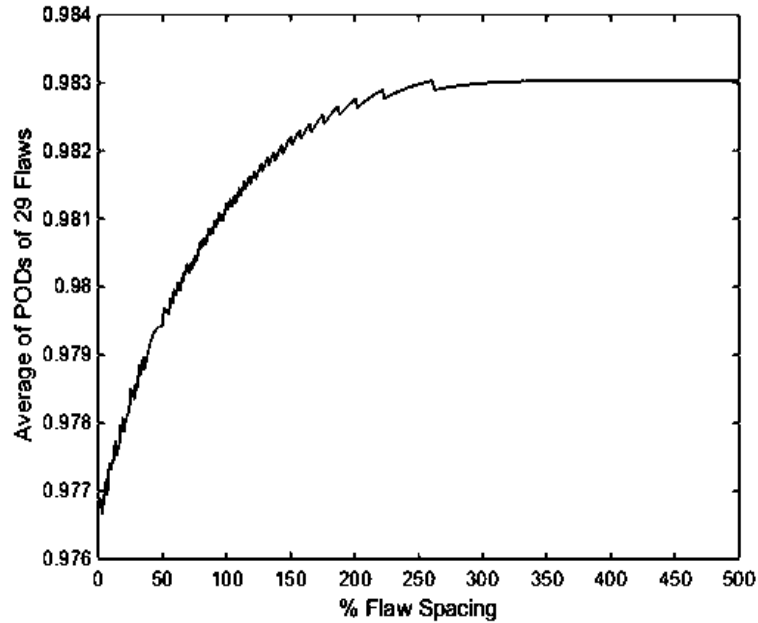


Fig. 8: Average POD versus % flaw spacing for 0.5 PPD

This curve also has some small error due to discrete values used in iterative calculations.

- The curve starts at 0.97637 POD for single size flaws.
- Average POD seems to level out at 0.983 for flaw spacing of ~250% for PPD = 0.5. Fig. 9 shows % delta versus % flaw spacing.

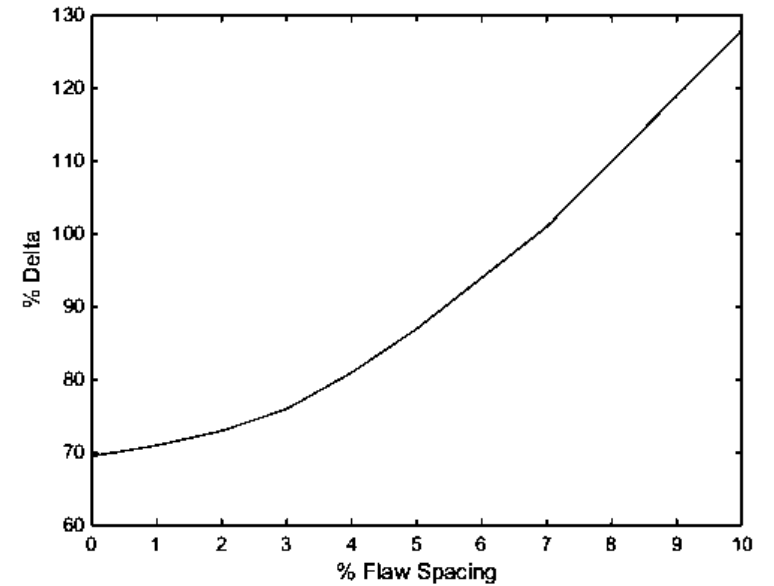
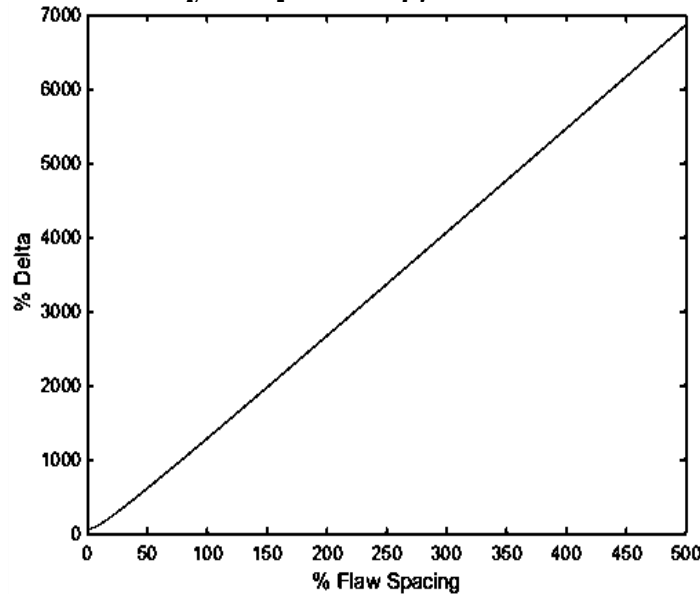


Fig. 9: Average POD versus % flaw spacing and delta versus range

- We determine  $a_{90}$  for the same decision threshold level using MIL-HDBK-1823. Fig. 9 indicates that, in order to get PPD value of 50%, a % delta of 69.5% (~70%) is necessary for single size flaw-set.
- In general,

$$\Delta_{50\%} = a_{90/95PE} - a_{90} \geq 0.695\sigma. \quad (24)$$

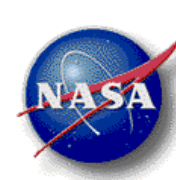
- For flaw spacing of 7% of POD model standard deviation and % delta with 0.5 PPD is about standard deviation.

$$\Delta_{50\%} = a_{90/95PE} - a_{90} \cong \sigma. \quad (25)$$

## Case 2 condition: Probability of passing demonstration equals 50%

- For flaw spacing of 7%, corresponding standard deviation of flaw-set is,  $0.7 \times 28 \times 0.304 = 59.6\%$ .
- This indicates that the point estimate demonstration that is reasonably easy to pass is likely to be very conservative.
- It also implies that the decision threshold level (DTL for  $\hat{a}$  versus “ $a$ ” testing) signals or the relevant flaw size (hit-miss testing) which establishes  $a_{90}$  is lower than average signal from flaw size  $a_{90/95}$  by about 1 standard deviation in signal. Signal response delta is given by,

$$\hat{a}(\Delta_{50\%}) = \hat{a}(a_{90/95PE}) - \hat{a}(a_{90}) \cong \hat{a}(\sigma). \quad (26)$$



# Signal to Noise Ratio



- DTL signal-to-noise ratio should be high (i.e.  $\geq 2$ ) to limit POF.
- Here, signal margin for  $a_{90/95PE}$  flaw size above DTL is same as standard deviation for noise times  $\Delta_{50\%}$ .
- Noise  $\hat{a}_{noise}$  can be measured as signal in areas where no flaws are present. mh1823 calculates percentile noise (e.g. cumulative 95% or  $\hat{a}_{noise95\%}$ ) and gives a curve for POF as a function of percentile noise. Therefore, the DTL-to-noise ratio can be calculated as,

$$SNR_{DTL\%} = \hat{a}_{DTL} / \hat{a}_{noise99\%}. \quad (27)$$

- If we do not have actual noise measurements, we can still use standard deviation of the POD curve as a measure of noise.

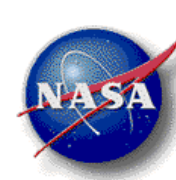
$$SNR_{DTL\sigma} = \hat{a}_{DTL} / \hat{a}(\sigma). \quad (28)$$

- Even this signal-to-noise ratio should be high ( $> 2$ ). If both these SNR are high it provides a low POF.
- In hit-miss analysis, signal response may not be measured but here the ratio with relevant flaw size may be of use. Relevant flaw size is due to signal or indication from the smallest flaw that needs to be reported. Therefore, it is a kind of DTL in flaw size units. Following flaw size ratio (FSR) would affect POF. Higher ratio would imply lower POF. Possibly, a minimum ratio above certain value (e.g. 1) would be desirable.

$$FSR_{rel,\sigma} = a_{rel} / \sigma \quad (29)$$

- Following flaw size ratio (FSR) would affect POD. Higher ratio would imply higher POD. Possibly, a minimum ratio above certain value (e.g. 1) would be desirable.

$$FSR_{90/95PE} = (a_{90/95PE} - a_{rel}) / \sigma \quad (30)$$



# Steps for Optimizing Point Estimate POD



- Step 1. Choose PPD value for the demonstration, e.g. 50%.
- Step 2. Choose decision threshold or relevant flaw size.
- Step 3. Collect hit-miss data or  $\hat{a}$  versus “ $a$ ” data for the flaw detection application.
- Step 4. Determine POD curve using MIL-HDBK-1823. Determine  $a_{90}$  ( $a_{90/50}$ ), mean ( $a_{50/50}$ ) and standard deviation in flaw units.
- Step 5. Choose a delta above 70% and determine % flaw spacing.
- Step 6. Construct flaw-set as follows.

$$a_{median} = a_{90} + \% \Delta_{50\%} \sigma / 100, \quad (31)$$

$$S = \% S_{50\%} \sigma / 100, \text{ and} \quad (32)$$

$$a_{min}, a_{max} = a_{median} \pm 14S. \quad (33)$$

- One could calculate standard deviation of the flaw-set from  $S$  and use mean as median to design the flaw-set for non-uniformly spaced flaws. This would give an approximate answer.
- Once flaws are chosen, then the PPD can be directly calculated by multiplying the PODs for 29 flaws and adjustment to the flaw-set made to provide acceptable PPD, POF and average POD.
- This leads us to Case 3, where we would compare several  $a_{90/95}$  of a simulated data with simulation input  $a_{90}$  value.





# Case 3: Comparison of Various $a_{90/95}$ Estimates from Simulated Data



Simulated data with programmed  $a_{90} = 0.2984$

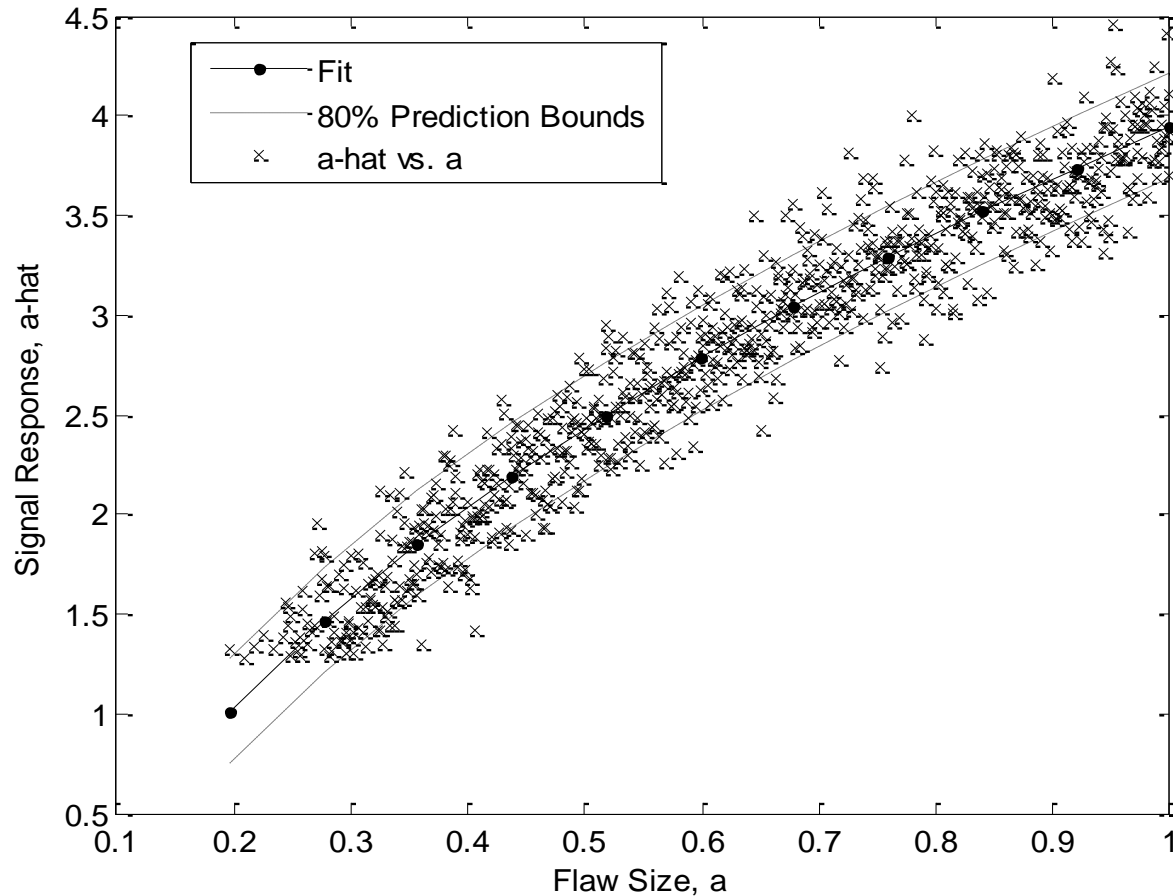


Fig. 10: Simulated signal response versus flaw size data

Table 1: Comparison of various  $a_{90/95}$  estimates from simulated data

Method	Quantity	Value (Arbitrary units)	% Difference with Model $a_{90}$	Rank
Theoretical (based on simulation model)	$a_{90}$	0.2984	Not applicable	
mh1823 $\hat{a}$ versus "a"	$a_{90/50}$	0.2907	-2.6	
	$a_{90/95}$	0.2946	-1.3	1
mh1823-hit/miss	$a_{90/50}$	0.2977	-0.2	
	$a_{90/95}$	0.3119	4.5	3
Matlab curve fit $\hat{a}$ versus "a"	$a_{90/50}$	0.2895	-3.0	
	$a_{90/95}$	0.2989	0.2	2
Point estimate based on average of 29 contiguous flaws	$a_{90/95PE}$	0.323	8.2	4



# Noise Analysis for Probability of False Calls

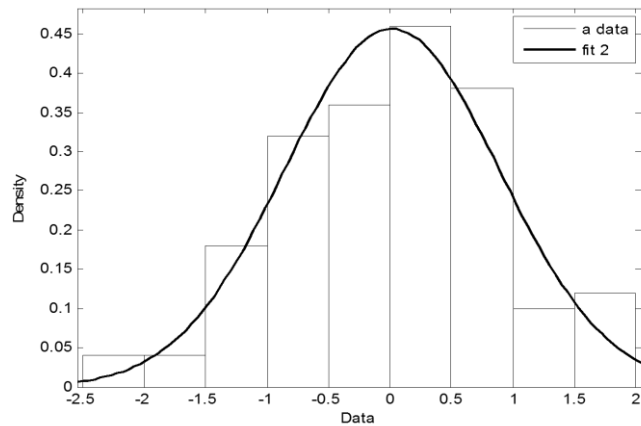


Fig. 11: Distribution function fit to noise data

Fit: Maximum Likelihood Estimation (MLE),  
 Distribution: Normal,  
 Log likelihood: -143.012,  
 Domain:  $-\infty < y < \infty$ ,  
 Mean,  $\mu$ : -0.110793,  
 Variance: 1.0329,  
 Sigma: 1.01631.

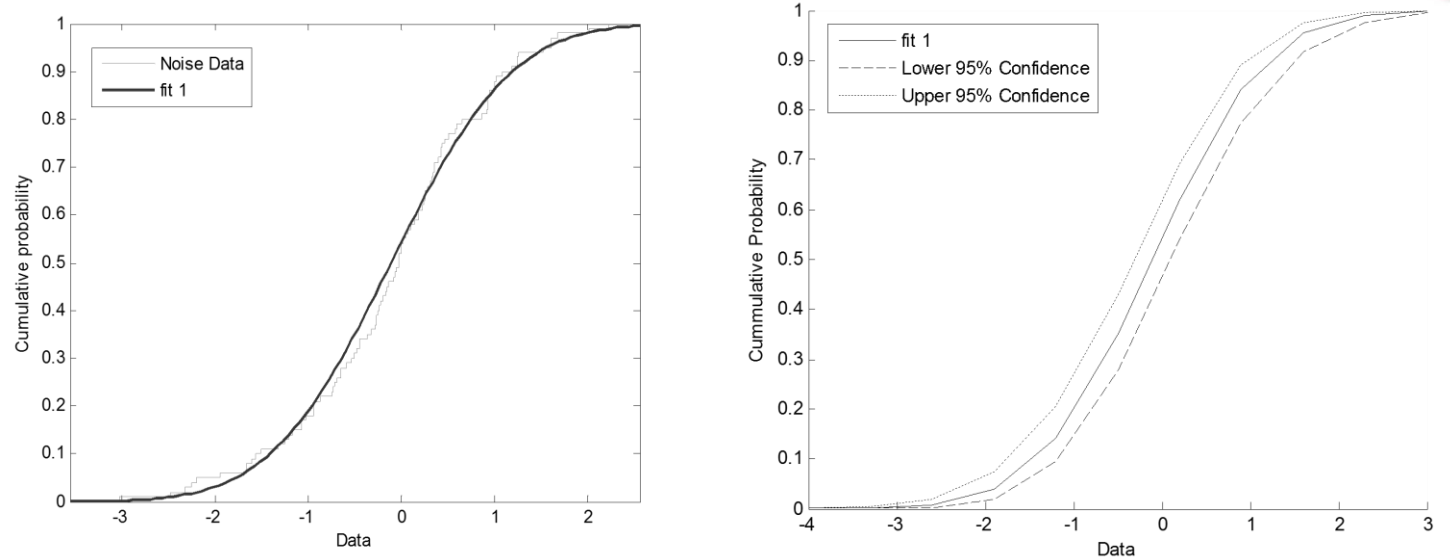
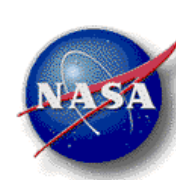


Fig. 12: a) Cumulative noise data and cumulative probability density function, b) Cumulative probability density function or  $POD_{noise}$  and confidence bounds.

- Probability of false calls (POF) is given by  $POD_{noise}$ , where noise level is equal to the decision threshold.

$$POF = 1 - POD_{noise} \quad (34)$$

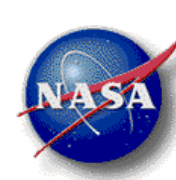
- Refer to MIL-HDBK-1823 for in depth analysis of noise and estimation of probability of false calls.



# Conclusions



- Case 1 indicates that for same average flaw size, PPD with flaws in a range is less than or equal to PPD for single size flaws.
  - Therefore, average flaw size of a set of 29 flaws can be used as a more conservative  $a_{90,95}$  estimate compared to single size  $a_{90,95}$  estimate, provided POD increases with flaw size.
- Case 2 indicates that in order to get PPD of 50%, a % delta of 69.5% is necessary for single size 29 flaw set.
- For flaw spacing of 7% of POD model standard deviation and PPD of 50%, % delta is equal to 100% of standard deviation.
- For flaw spacing of 7% of POD model standard deviation, the standard deviation of the flaw-set is 59.6% of POD model standard deviation.
- In Case 3, POD analysis on simulated data indicates that, point estimate  $a_{90,95PE}$  flaw size is the most conservative among the  $a_{90,95}$  from four POD analysis methods discussed.
  - PPD was calculated to be 53%.
  - Delta is calculated to be 107%.
  - Flaw spacing is calculated 4.35% using the same standard deviation.
  - Flaw-set standard deviation is 37% of the POD model standard deviation.
  - The DTL SNR were calculated to be much greater than 2.



## Conclusions (continued)



- In Case 3, Simulated data analysis results agree with Case 2 predictions to some extent.
- Compared to  $\hat{a}$  versus “ $a$ ” method, point estimate method has a lower DTL to qualify the same  $a_{90/95}$ .
- Lowering the DTL would detect smaller flaws reliably but also may cause higher false call rate.
- The POD input data model and mh1823  $\hat{a}$  versus “ $a$ ” model differ and therefore the Case 3 results do not come very close to the Case 2 predictions but they are of same order.
- The paper indicates that for PPD of 50%, point estimate  $a_{90/95PE}$  is very conservative possibly by as much as the standard deviation of the POD model with standard deviation for 29 flaw-set equal to 40-60% of standard deviation of the POD model.
- In order to design point estimate 29 flaw-set, the paper recommends using mh1823 POD analysis first.
- Results of the mh1823 POD analysis e.g. POD curve, noise percentile, POF and POD versus DTL analysis shall be used to design the optimal point estimate flaw-set that provides smallest possible  $a_{90/95PE}$  with acceptable PPD and POF.
- Procedure for optimizing the point estimate demonstration set is provided.