Designation: ~~E3023 – 15~~ E3023 – 21

Standard Practice for
Probability of Detection Analysis for *â* Versus *a* Data[1]

This standard is issued under the fixed designation E3023; the number immediately following the designation indicates the year of original adoption or, in the case of revision, the year of last revision. A number in parentheses indicates the year of last reapproval. A superscript epsilon (ε) indicates an editorial change since the last revision or reapproval.

1. Scope

1.1 This practice defines the procedure for performing a statistical analysis on Nondestructive Testing (NDT) *â* versus *a* data to determine the demonstrated probability of detection (POD) for a specific set of examination parameters. Topics covered include the standard *â* versus *a* regression methodology, POD curve formulation, validation techniques, and correct interpretation of results.

1.2 *Units*—The values stated in inch-pound units are to be regarded as standard. The values given in parentheses are mathematical conversions to SI units that are provided for information only and are not considered standard.

1.3 *This standard does not purport to address all of the safety concerns, if any, associated with its use. It is the responsibility of the user of this standard to establish appropriate ~~safety~~ safety, health, and ~~health~~environmental practices and determine the applicability of regulatory limitations prior to use.*

1.4 *This international standard was developed in accordance with internationally recognized principles on standardization established in the Decision on Principles for the Development of International Standards, Guides and Recommendations issued by the World Trade Organization Technical Barriers to Trade (TBT) Committee.*

2. Referenced Documents

2.1 *ASTM Standards:*[2]

E178 Practice for Dealing With Outlying Observations
E456 Terminology Relating to Quality and Statistics
E1316 Terminology for Nondestructive Examinations
E1325 Terminology Relating to Design of Experiments
E2586 Practice for Calculating and Using Basic Statistics
E2782 Guide for Measurement Systems Analysis (MSA)
E2862 Practice for Probability of Detection Analysis for Hit/Miss Data
E3080 Practice for Regression Analysis with a Single Predictor Variable

2.2 *Department of Defense Document:*[3]

MIL-HDBK-1823A Nondestructive Evaluation System Reliability Assessment

3. Terminology

3.1 *Definitions of Terms Specific to This Standard:*

---

[1] This ~~test method~~ practice is under the jurisdiction of ASTM Committee E07 on Nondestructive Testing and is the direct responsibility of Subcommittee E07.10 on Specialized NDT Methods.

Current edition approved ~~June 15, 2015~~Feb. 15, 2021. Published ~~August 2015~~March 2021. Originally approved in 2015. Last previous edition approved in 2015 as E3023 – 15. DOI: ~~10.1520/E3023-15.~~10.1520/E3023-21.

[2] For referenced ASTM standards, visit the ASTM website, www.astm.org, or contact ASTM Customer Service at service@astm.org. For *Annual Book of ASTM Standards* volume information, refer to the standard's Document Summary page on the ASTM website.

[3] Available from Standardization Documents Order Desk, DODSSP, Bldg. 4, Section D, 700 Robbins Ave., Philadelphia, PA 19111-5098, http://dodssp.daps.dla.mil.

3.1.1 *analyst, n*—the person responsible for performing a POD analysis on $\hat{a}$ versus a data resulting from a POD examination.

3.1.2 *decision threshold,* $\hat{a}_{dec}$, *n*—the value of $\hat{a}$ above which the signal is interpreted as a find and below which the signal is interpreted as a miss.

3.1.2.1 *Discussion—*
A decision threshold is required to create a POD curve. The decision threshold is always greater than or equal to the noise threshold and is the value of $\hat{a}$ that corresponds with the flaw size that can be detected with ~~50%~~50 % POD.

3.1.3 *demonstrated probability of detection, n*—the calculated POD value resulting from the statistical analysis of the $\hat{a}$ versus a data.

3.1.4 *false call, n*—~~—~~the perceived detection of a discontinuity that is identified as a find during a POD examination when no discontinuity actually exists at the inspection site.

3.1.4.1 *Discussion—*
A synonym for "false call" is "false positive."

3.1.5 *noise, n*—signal response containing no useful target characterization information.

3.1.6 *noise threshold,* $\hat{a}_{noise}$, *n*—the value of $\hat{a}$ below which the signal is indistinguishable from noise.

3.1.6.1 *Discussion—*
The noise threshold is always less than or equal to the decision threshold. The noise threshold is used to determine left censored data.

3.1.7 *probability of ~~detection,~~ detection (POD), n*—the fraction of nominal discontinuity sizes expected to be found given their existence.

3.1.8 *regression analysis, n*—a statistical procedure used to characterize the association between two or more numerical variables for prediction of the response variable from the predictor variable. **E456, E3080**

3.1.8.1 *Discussion—*
This practice focuses on (but is not limited to) regression analysis with a single predictor variable. Appendix X1 in this standard includes details on this topic as applied to Probability of Detection. See also Practice E3080 for an overview of linear regression with a single predictor variable.

3.1.9 *saturation threshold,* $\hat{a}_{sat}$, *n*—the value of $\hat{a}$ associated with the maximum output of the system or the largest value of $\hat{a}$ that the system can record.

3.1.9.1 *Discussion—*
The saturation threshold is used to determine right censored data.

3.2 *Symbols:*

3.2.1 *a*—discontinuity size.

3.2.2 $\hat{a}$—the measured signal response for a given discontinuity size, *a*.

3.2.2.1 *Discussion—*
The measured signal response is assumed to be continuous in nature. Units depend on the NDT inspection system and can be, for example, scale divisions, number of contiguous illuminated pixels, or millivolts.

3.2.3 a_p—the discontinuity size that can be detected with probability *p*.

3.2.3.1 *Discussion—*
Each discontinuity size has an independent probability of being detected and corresponding probability of being missed. For example, being able to detect a specific discontinuity size with probability *p* does not guarantee that a larger size discontinuity will be found.

3.2.4 $a_{p/c}$—the discontinuity size that can be detected with probability *p* with a statistical confidence level of *c*.

3.2.4.1 *Discussion—*
According to the formula in MIL-HDBK-1823A, $a_{p/c}$ is a one-sided upper confidence bound on a_p.$a_{p/c}$ represents how large the true a_p could be given the statistical uncertainty associated with limited sample data. Hence $a_{p/c} > a_p$. Note that POD is equal to *p* for both $a_{p/c}$ and a_p.a_p is based solely on the observed relationship between the *â* and *a* data and represents a snapshot in time, whereas $a_{p/c}$ accounts for the uncertainty associated with limited sample data.

4. Summary of Practice

4.1 ~~This practice~~ In general, the POD examination process is comprised of a specimen set design, study design, examination administration, statistical analysis of examination data, documentation of analysis results, and specimen set maintenance. This practice is focused only on and describes, step-by-step, the process for analyzing nondestructive testing *â* versus *a* data resulting from a POD examination, including minimum requirements for validating the resulting POD ~~curve.~~curve, and documenting the results.

4.2 This practice also includes definitions and discussions for results of interest ~~(e.g.,~~ (for example, $a_{90/95}$) to provide for correct interpretation of results.

4.3 Definitions of statistical terminology used in the body of this practice can be found in Annex A1.

4.4 A more general discussion of the POD analysis process can be found in Appendix X1.

4.5 A mathematical overview of the underlying model commonly used with *â* versus *a* data resulting from a POD examination can be found in Practice E3080.

5. Significance and Use

5.1 The POD analysis method described herein is based on well-known and well-established statistical methods. It shall be used to quantify the demonstrated POD for a specific set of examination parameters and known range of discontinuity sizes under the following conditions.

5.1.1 The initial response from a nondestructive evaluation inspection system is measurable and can be classified as a continuous variable.

5.1.2 Discontinuity size is the predictor variable and can be accurately quantified.

5.1.3 The relationship between discontinuity size (*a*) and measured signal response (*â*) exists and is best described by a linear regression model with an error structure that is normally distributed with mean zero and constant variance, σ^2. (Note that ~~"linear" refers to~~ in linear regression, "linear" means linear with respect to the model coefficients. ~~For example,~~ Though a quadratic model $\hat{y}=\beta_0+\beta_1 \cdot x+\beta_2 \cdot x^2$ ~~is a linear model.)~~does not have a linear shape when plotted, for example, it is classified as a linear model in regression analysis since it is linear with respect to the model coefficients.)

5.2 This practice does not limit the use of a linear regression model with more than one predictor variable or other statistical models if justified as more appropriate for the *â* versus *a* data.

5.3 This practice is not appropriate for data resulting from a POD examination on nondestructive evaluation systems that generate an initial response that is binary in nature (for example, hit/miss). Practice E2862 is appropriate for systems that generate a hit/miss-type response (for example, fluorescent penetrant).

5.4 Prior to performing the analysis, it is assumed that the discontinuity of interest is clearly defined; the number and distribution of induced discontinuity sizes in the POD specimen set is known and well documented; the POD examination administration procedure (including data collection method) is well designed, well defined, under control, and ~~unbiased;~~ unbiased (see X1.2.2 for more detail); the initial inspection system response is measurable and continuous in nature; the inspection system is calibrated; and the measurement error has been evaluated and deemed acceptable. The analysis results are only valid if the *â* versus *a* data are accurate and precise and the linear model adequately represents the *â* versus *a* data.

5.5 The POD analysis method described herein is consistent with the analysis method for *â* versus *a* data described in

MIL-HDBK-1823A and is included in several widely utilized POD software packages to perform a POD analysis on $\hat{a}$ versus a data. It is also found in statistical software packages that have linear regression analysis capability. This practice requires that the analyst has access to either POD software or other software with linear regression analysis capability.

6. Procedure

6.1 The POD analysis objective shall be clearly defined by the responsible engineer or by the customer.

6.2 The analyst shall obtain the $\hat{a}$ versus a data resulting from the POD examination, which shall include at a minimum the documented known induced discontinuity sizes, the associated measured signal response, and any false calls.

6.3 The analyst shall also obtain specific information about the POD examination, which shall include at a minimum the specimen standard geometry ~~(e.g.,~~ (for example, flat panels), specimen standard material ~~(e.g.,~~ (for example, Nickel), examination date, number of inspectors, type of inspection method ~~(e.g.,~~ (for example, Eddy Current Inspection), pertinent information about the instrument and instructions for use ~~(e.g.,~~ (for example, settings, probe type, scan path), and pertinent comments from the inspector(s) and test administrator.

6.3.1 In general, the results of an experiment apply to the conditions under which the experiment was conducted. Hence, the POD analysis results apply to the conditions under which the POD examination was conducted.

6.4 Prior to performing the analysis, the analyst shall conduct a preliminary review of the POD examination procedure to identify any issues with the administration of the examination. The analyst shall identify and attempt to resolve any issues prior to conducting the POD analysis. Identified issues and their resolution shall be documented in the report. Examples of examination administration issues and possible resolutions are outlined in the following subsections.

6.4.1 If problems or interruptions occurred during the POD examination that may bias the results, the POD examination should be re-administered.

6.4.2 If the examination procedure was poorly designed ~~and/or~~or executed, or both, the validity of the resulting data is questionable. In this case, the examination procedure design and execution should be reevaluated. For design guidelines, see MIL-HDBK-1823A.

6.5 Prior to performing the analysis, the analyst shall conduct a preliminary review of the $\hat{a}$ versus a data to identify any data issues. The analyst shall identify and attempt to resolve any issues prior to conducting the POD analysis. Identified issues and their resolution shall be documented in the report. Examples of data issues and possible resolutions are outlined in the following subsections.

6.5.1 Any apparent outlying observations shall be reviewed for correctness. If a typo is identified, the typo shall be corrected prior to performing the analysis. If the value is correct, it shall be retained in the analysis and its influence on the $\hat{a}$ versus a model shall be evaluated during the model diagnostic assessment. The analyst should also reference Practice E178.

6.5.2 POD cannot be modeled as a continuous function of discontinuity size if all the measured signal responses are below the noise threshold or above the saturation threshold. If this occurs, the adequacy of the nondestructive testing system should be evaluated.

6.6 Only $\hat{a}$ versus a data for induced discontinuities shall be used in the development of the linear regression model. False call data shall not be included in the development of the linear model when using standard linear regression analysis methods.

6.7 The analyst in conjunction with the responsible engineer shall determine the value of the noise threshold, $\hat{a}_{noise}$, and saturation threshold, $\hat{a}_{sat}$, prior to performing the analysis.

6.7.1 The value of $\hat{a}_{noise}$ is determined by performing a noise analysis. A noise analysis is typically accomplished by assessing the distribution of measured signal responses from sites with no known discontinuity (false calls) ~~and/or~~or measured signal ~~responses~~ responses, or both, that are not influenced by the size of the discontinuities (noise). Details on performing a noise analysis can be found in MIL-HDBK-1823A.

6.8 The analyst shall select an appropriate linear regression model to establish the relationship between $\hat{a}$ and a. Selection of a linear model may be an iterative process as the significance of the predictor variable(s) and the appropriateness of the selected model are typically assessed after the model has been developed.

6.8.1 "Linear" refers to linear with respect to the model coefficients. For example, $\hat{y}_i = b_0 + b_1 \cdot (x^2)$ and $\hat{y}_i = b_0 + b_1 \cdot x_1 + b_2 \cdot \ln(x_2)$ are linear regression models. (For more detail, see definition in Annex A1 and discussion in Practice E3080 Annex A1.1.)

6.8.2 In general, only significant and uncorrelated predictor variables are included in a regression model. If more than one predictor variable is being considered for inclusion in the model, a preliminary graphical analysis of the response variable against each predictor variable may help identify which predictor variables appear to influence the response and the type of relationship (for example, direct, inverse, quadratic). In addition, a preliminary graphical analysis of all possible pairings of predictor variables shall be performed to verify independence of the predictor variables. When plotted against each other, there should be no apparent relationship between any two predictor variables.

6.8.3 The appropriateness of a selected model is determined by how well the model fits the observed data and how well the underlying regression analysis assumptions are met.

6.9 The analyst shall use software that has the appropriate linear regression analysis capabilities to perform a linear regression analysis on the $\hat{a}$ versus a data.

6.9.1 If censored data are present, the analyst shall do the following:

6.9.1.1 Include and identify the censored data in the analysis (according to the notation required by the software).

6.9.1.2 Use the method of maximum likelihood to estimate the model coefficients.

6.9.1.3 Verify that convergence was achieved. If convergence is not achieved, the resulting $\hat{a}$ versus a model shall not be used to develop a POD curve.

6.9.1.4 Check the number of iterations it took to converge, provided that information on convergence and the number of iterations it took to converge is included in the analysis software output. If more than ~~twenty~~20 iterations were needed to reach convergence, the model may not be reliable.

6.9.1.5 Include a statement in the report indicating that convergence was achieved and the number of iterations needed to achieve convergence.

6.9.2 If no censored data are present, the method of maximum likelihood or the method of least squares shall be used.

6.10 If included in the analysis software output, the analyst shall assess the significance of the predictor variables in the model. Only significant predictor variables should be included in the model. (See X1.2 for more detail.)

6.11 Once the $\hat{a}$ versus a model is estimated, the analyst shall use, at a minimum, the model diagnostic methods listed below to assess the underlying linear regression analysis assumptions. The methods listed below shall be performed using only non-censored data. If available, other formal diagnostic methods (noted in ~~Appendix X1~~X1.2) should be used to assess the linear regression analysis assumptions.

6.11.1 There are three main underlying assumptions in a linear regression analysis: *(1)* residuals are normally distributed with mean 0 and constant variance, σ^2, *(2)* the residuals are independent, and *(3)* the relationship is in fact linear. The residual is calculated as $e_i = y_i - \hat{y}_i$ and represents the difference between the observed result, y_i, and the predicted value, $\hat{y}_i$, for the ith case. In general, the results of a linear regression analysis are not valid unless these assumptions hold. At a minimum, the following analyses of the residuals shall be performed to verify the assumptions.

6.11.1.1 A histogram of the residuals shall be constructed to assess the normality assumption and centering of the residuals. A histogram of the residuals should be roughly bell-shaped and symmetric around zero. In general, bell shape and symmetry around zero are more important than strict normality since traditional estimation procedures are typically only sensitive to large departures from normality (particularly with respect to skewness).

6.11.1.2 The constant variance and linearity assumptions shall be verified by plotting the residuals (*y*-axis) against the predicted values (*x*-axis). If the residuals fall in a horizontal band centered around zero, with no systematic preference for being positive or negative, then the assumption of constant variance and a linear relationship holds. (See Fig. X1.2 in Appendix X1.) In general, meeting the constant variance assumption is more important than meeting the normality assumption.

6.12 The analyst shall use at a minimum the methods listed below to assess the goodness-of-fit, influential points, and multicollinarity among predictor variables. If available, more formal methods (noted in Appendix X1) should be used.

6.12.1 A plot of predicted values versus actual values shall be used to assess goodness-of-fit. The plotted points should fall roughly on the $y = x$ line. Plotted points deviating from the $y = x$ line in a systematic way may be an indication of poor fit.

6.12.2 The analyst shall assess the influence of data that appears to be outlying on the established $\hat{a}$ versus a model. The histogram of the residuals and plot of the residuals versus predicted values can help identify outlying values. The influence of a suspected outlying value shall at a minimum be evaluated by removing the outlying value from the data and re-running the analysis to assess its influence on the $\hat{a}$ versus a model. A data point is said to be influential (or have high leverage) if its exclusion from the analysis has a relatively large effect on the $\hat{a}$ versus a model. Both analysis results (with and without the outlying data) shall be included in the report along with a discussion of the impact to the resulting POD curve and confidence bound (if applicable).

6.12.3 If the model includes more than one predictor variable, a graphical analysis shall be performed to verify independence of the predictor variables. (This step may be done during model selection as described in Appendix X1.)

6.13 The responsible engineer shall determine the value of $\hat{a}_{dec}$ that is most appropriate with respect to end use of the POD analysis results. A value for the decision threshold is required to create a POD curve. The value must be greater than or equal to the value of the noise threshold. That is, $\hat{a}_{dec} \geq \hat{a}_{noise}$.

6.14 The analyst shall use the decision threshold to determine a POD value for each discontinuity size given the established relationship between $\hat{a}$ and a, the formula for which can be found in Appendix X1. The resulting POD values shall be plotted against discontinuity size to produce a POD Curve.

6.14.1 POD curves tend to be *s*-shaped when a simple linear regression model is selected.

6.14.2 If more than one predictor variable is included in the model, POD is a response surface rather than a single curve.

6.14.3 The analyst shall determine the most appropriate way to plot the results.

6.15 If a c% level of confidence is specified by the responsible engineer or the customer, the analyst shall put a c% lower confidence bound on the POD curve by calculating a c% lower confidence bound on the $\hat{a}$ versus a model fit. Methods for constructing a confidence bound around a regression fit can be found in MIL-HDBK-1823A as well as statistics text books on linear regression.[4]

6.15.1 If, for example, the objective of the analysis is to determine the discontinuity size that can be detected with ~~90%~~90 % probability and ~~95%~~95 % confidence, denoted $a_{90/95}$, then the analyst shall put a ~~95%~~95 % lower confidence bound on the POD curve by calculating a ~~95%~~95 % lower confidence bound on the $\hat{a}$ versus a model fit. The formula for the ~~95%~~95 % lower confidence bound on the POD curve, which is based on the ~~95%~~95 % lower confidence bound around the regression fit, can be found in Appendix X1.

6.16 The analyst shall analyze any false call data and shall report the false call rate.

6.16.1 The responsible engineer or the customer shall clearly define what constitutes a false call.

6.16.2 A distributional analysis of false call or noise data, or both, is typically performed to assess the false call rate, a discussion of which can be found in MIL-HDBK-1823A.

[4] Neter, J, Kutner, M, Nachtsheim, C, Wasserman, W. *Applied Linear Statistical Models*, The McGraw-Hill Companies.

6.17 Acceptable false call rates shall be determined by the responsible engineer or by the customer.

7. Report

7.1 At a minimum the following information about the POD analysis shall be included in the report.

7.1.1 The specimen standard geometry ~~(e.g.,~~ (for example, flat panels).

7.1.2 The specimen standard material ~~(e.g.,~~ (for example, nickel).

7.1.3 Examination date.

7.1.4 Number of inspectors.

7.1.5 Type of inspection ~~(e.g.,~~ (for example, Eddy Current).

7.1.6 Pertinent information about the instrument and instructions for use ~~(e.g.,~~ (for example, settings, probe type, scan path).

7.1.7 Any comments from the inspector(s) or test administrator.

7.1.8 The documented known induced discontinuity sizes.

7.1.9 The associated measured signal responses, including information about censored data.

7.1.10 Any false calls.

7.1.11 The linear regression model describing the relationship between the observed $\hat{a}$ versus a data and confidence bound (if applicable).

7.1.12 A statement indicating that convergence was achieved and the number of iterations to convergence, if maximum likelihood estimation was used.

7.1.13 A statement about the model diagnostic methods used and conclusions.

7.1.14 The estimate of the error around the regression fit (calculated as the square root of the mean square error, which is typically included in the software output).

7.1.15 Summary of the noise analysis and rationale for selection of the decision threshold.

7.1.16 A plot of the resulting POD curve and confidence bound (if applicable).

7.1.17 Specific results of interest as required by the analysis objective ~~(e.g.,~~ (for example, $a_{90/95}$).

7.1.18 Any deviations from the POD examination procedure or standard POD analysis.

7.1.18.1 If the POD examination was re-administered, the original results and rationale for re-administration shall be documented in the report.

7.1.18.2 If a discontinuity is removed from the analysis, the specific discontinuity and rationale for removal shall be documented in the final report.

7.1.18.3 If the impact of outlying data was assessed, the results shall be included in the report along with an explanation.

7.1.19 Summary of false call analysis, including a definition of what constitutes a false call, the false call rate, and the method used to estimate the false call rate.

7.1.20 Name of analyst and company responsible for the POD calculation.