NOT MEASUREMENT SENSITIVE

MIL-HDBK-1823

30 April 1999

DEPARTMENT OF DEFENSE

HANDBOOK

NONDESTRUCTIVE EVALUATION SYSTEM RELIABILITY ASSESSMENT

THIS HANDBOOK IS FOR GUIDANCE ONLY

DO NOT CITE THIS HANDBOOK AS A REQUIREMENT

AMSC N/A AREA NDTI

FOREWORD

1. This handbook is approved for use by all Departments and Agencies of the Department of Defense (DoD).

2. This handbook is for guidance only. This handbook cannot be cited as a requirement. If it is, the contractor does not have to comply.

3. Beneficial comments (recommendations, additions, deletions) and any pertinent data which may be of use in improving this document should be addressed to: ASC/ENSI, 2530 Loop Road West, Bldg 560, Wright-Patterson AFB OH 45433-7101, by using the Standardization Document Improvement Proposal (DD Form 1426) appearing at the end of this document or by letter.

CONTENTS

CONTENTS

CONTENTS

1. SCOPE

1.1 Scope.

This handbook applies to all agencies within the DoD and industry involving methods for testing and evaluation procedures for assessing Non-Destructive Evaluation (NDE) system capability. This handbook is for guidance only. This handbook cannot be cited as a requirement. If it is, the contractor does not have to comply.

1.2 Limitations.

This handbook provides uniform guidance requirements for establishing NDE procedures used to inspect new or inservice hardware for which a measure of NDE reliability is required. They are, specifically, Eddy Current (EC), Fluorescent Penetrant (PT), Ultrasonic (UT), and Magnetic Particle (MT) Testing. This document may be used for other NDE procedures if they are similar in output to those listed herein, such as Radiographic testing, Holographic testing, Shearographic testing, etc.

1.3 Classification.

NDE systems are classified into either of two categories: those which produce only qualitative information as to the presence or absence of a flaw, i.e., hit/miss data, and systems which also provide some quantitative measure of the size of the indicated flaw, i.e., â vs. a data.

2. APPLICABLE DOCUMENTS

2.1 General.

The documents listed below are not necessarily all of the documents referenced herein, but are the ones that are needed in order to fully understand the information provided by this handbook.

2.2 Government documents.

2.2.1 Specifications, standards, and handbooks.

The following specifications, standards and handbooks form a part of this document to the extent specified herein. Unless otherwise specified, the issues of these documents are those listed in the latest issue of the Department of Defense Index of Specifications and Standards (DoDISS) and supplement thereto.

STANDARDS

DEPARTMENT OF DEFENSE

MIL-STD-410	Nondestructive Testing Personnel Qualification and Certification (Cancelled) see NAS-410
MIL-STD-1783	Engine Structural Integrity Program
JSSG-87221	Aircraft Structures, General Specification for

(Unless otherwise indicated, copies of the above specifications, standards, and handbooks are available from the DoDSSP, Bldg 4D, 700 Robbins Ave., Philadelphia PA 19111-5094.)

2.3 Non-Government publications.
The following document(s) form a part of this document to the extent specified herein. Unless otherwise specified, the issues of the documents which are DoD adopted are those listed in the latest issue of the DoDISS, and supplement thereto.

ANSI/ASNT CP-189	ANST Standard for Qualification and Certification of Nondestructive Testing Personnel
	Box, Hunter, and Hunter, Statistics for Experimenters (Wiley, 1978)
SNT-TC-1A	"Personnel Qualification and Certification in Non-Destructive Testing", American Society for Nondestructive Testing (ASNT), (1983)
UDR-TR-88-12	Berens, Hovey, Donahue, and Craport, "User's Manual for Probability of Detection," University of Dayton Research Institute, (January 1988)

(Non-Government standards and other publications are normally available from organizations that prepare or distribute documents. These documents also may be available in or through libraries or other informational services.)

2.4 Order of precedence.
In the event of a conflict between the text of this document and the references cited herein, the text of this document takes precedence. Nothing in this document, however, supersedes applicable laws and regulations unless a specific exemption has been obtained.

3. DEFINITIONS

a, flaw size	Actual physical dimension of a flaw; can be its depth, surface length, or diameter of a circular, or radius of semi-circular or corner flaw having the same cross-sectional area.
â, a-hat	Measured response of the NDE system, to a flaw of flaw size, a. Units depend on inspection apparatus, and can be scale divisions, counts, number of contiguous illuminated pixels, or millivolts.
a_{50}	Flaw size at 50% POD
$\hat{a}_{dec}$,decision threshold	Value of â above which the signal is interpreted as a hit, and below which the signal is interpreted as a miss. It is the â value associated with 50% POD. Decision threshold is always greater than or equal to inspection threshold.
$\hat{a}_{sat}$, saturation	Value of â as large, or larger than, the maximum output of the system or the largest value of â that the system can record.

$\hat{a}_{th}$, signal, or inspection threshold	Value of $\hat{a}$ below which the signal is indistinguishable from the noise or the smallest value of $\hat{a}$ that the system records. Inspection threshold is always less than or equal to decision threshold.
β_0, β_1	Intercept and slope of the linear relationship between Log $\hat{a}$ and Log a
$\hat{\beta}_0, \hat{\beta}_1$ and $\hat{\delta}$	Maximum likelihood estimators of parameters β_0, β_1, δ
censored data	Signal response either smaller than $\hat{a}_{th}$, and therefore indistinguishable from the noise (left censored), or greater than $\hat{a}_{sat}$, (right censored), and therefore a saturated response
crack	A subset of flaws
$\hat{d}$	A calculated flaw depth estimated from its signal response
δ	Standard error of residuals of regression of Log $\hat{a}$ on Log a
ET	Eddy current testing
factor	A variable whose effect on POD(a) is to be evaluated
false call	An NDE system response interpreted as having detected a flaw but associated with no known flaw at the inspection location.
flaw	An imperfection or discontinuity that may be detectable by nondestructive testing inspection and is not necessarily rejectable
hit	An NDE system result interpreted as having detected a flaw
inspector	The person who actually applies the NDE technique, interprets the results, and determines the acceptance of the material per the applicable specifications. The inspector must be certified to the same level required for production inspectors, per MIL-STD-410 or SNT-TC-1A, for the NDE technique being applied.
maximum likelihood estimation	A standard statistical method used to estimate numerical values for model parameters, $\beta_0, \beta_1, \delta, \mu$ and σ
miss	An NDE system response interpreted as not having detected a flaw
MT	Magnetic particle testing
NDE	Nondestructive evaluation, which encompasses both the inspection itself and the subsequent statistical and engineering analyses of the inspection data
NDE system	A collection that can include hardware, software, materials, and procedures intended for the application of a specific NDE test

	method. NDE systems can range from fully manually operated to fully automated.
noise	Signal response containing no useful flaw characterization information
POD	Probability of detection
POD(a), probability of detection	The fraction of flaws of nominal flaw size, a, which are expected to be detected (found)
PT	Fluorescent penetrant testing
residual	The difference between an observed signal response and the response predicted from the model
system operator	The person in charge of an automated or semi-automated system, and who is responsible for the mechanical, electrical, computer, and other systems being maintained in proper operating condition. The system operator should be certified to the same level required for production inspectors, per MIL-STD-410 or SNT-TC-1A, for the NDE technique being applied.
test monitor	The person assigned to monitor the system reliability testing per this document, and to assure that all requirements of this standard are being met.
UT	Ultrasonic testing

4. GENERAL REQUIREMENTS

4.1 General.

This section addresses the general requirements for assessing the capability of an NDE system in terms of the probability of detection (POD) as a function of flaw size, a . These general requirements are applicable to all NDE systems of this handbook and addresses responsibilities for planning, conducting, analyzing, and reporting NDE reliability evaluations. Specific requirements that pertain to eddy current test (ET), fluorescent penetrant test (PT), ultrasonic test (UT), and magnetic particle test (MT) inspection systems are contained in Appendices A through D.

4.2 System definition and control.

Evaluation of the NDE system in terms of the limits of operational parameters and range of application and demonstrate that the system is in control. In addition to the physical attributes of the NDE system, this may include planned statistical assessments of those components responsible for system variability.

4.3 Demonstration design.

To ensure that the assessment of the NDE system is complete, suitable documentation may be developed which specifies the experimental design for the inspections; the method of obtaining and maintaining the structural specimens to be inspected; the procedures for performing the inspections; and the process for ensuring the inspection system is under control. The topics that may be addressed in each of these areas include the following.

4.3.1 Experimental design.

The prime objective of an NDE reliability demonstration is to determine the POD versus flaw size relationship which defines the capability of an NDE system under representative application conditions. Variation in NDE system response (and, hence, uncertainty in detectability) is caused by both the physical attributes of a flaw and the NDE process variables or parameters. The uncertainty caused by differences between flaws is accounted for by using representative specimens with flaws of known size in the demonstration inspections (see 4.3.2). The uncertainty caused by the NDE process is accounted for by a test matrix of different inspections to be performed on the complete set of specimens. If the experiment is properly designed and executed, a secondary objective of identifying those factors which significantly influence POD for the system can also be met.

a. The experimental design defines the conditions related to the NDE process parameters under which the demonstration inspections will be performed. In particular, the experimental design comprises:

1. The identification of the process variables which may influence flaw detectability but cannot be precisely controlled in the real inspection environment;

2. The specification of a matrix of inspection conditions which fairly represents the real inspection environment by accounting for the influencing variables in a manner which permits valid analyses;

3. The order for performing the individual inspections of the test matrix. (The number of flawed and unflawed inspection sites in the experiment could also be considered as part of the experimental design, and this topic is addressed in 4.3.2.1.)

b. Although general guidelines for these areas are presented in the following paragraphs, it is recommended that a qualified statistician participate in the preparation of the experimental design.

4.3.1.1 Test variables.

It is assumed that the inspection process has been defined and is under control for the demonstration testing. Even so, there will be factors which cannot be completely controlled or can only be controlled within reasonable operational limits. To evaluate the inspection system in the application environment, these factors may be identified so that they can be fairly represented in the demonstration tests. For example, in a manual inspection, it would not be acceptable to use only the known best inspector in the demonstration tests. Rather, the entire population of inspectors may be represented, as is discussed in 4.3.1.2.

a. The contractor may generate a list of process variables which can be expected to influence the efficacy of the NDE system. This list may provide the basis for generating the evaluation test matrix. To assure a thorough evaluation, it is recommended that the initial matrix include as many variables as possible. If early in the test program it is demonstrated that a particular variable is not significant, it may be eliminated from further consideration, thus resulting in a revised, smaller test matrix. To be eliminated, it may be shown that the variable has no significant effect on POD using the analysis methods as specified in Appendices G and H. The government reserves the right to expand or reduce the list of variables to be included in the test matrix.

b. As a minimum, the following types of variables should be considered in generating the list of test variables:

1. Part preprocessing: This variable type includes factors such as part cleaning, preparation, contour, and surface condition. It could also include such things as the application of the penetrant for fluorescent penetrant readers. Early in the definition of the system acceptance test plan, a decision may be made as to how far upstream the requirements should extend. For a penetrant reading system, it may be determined not to consider the penetrant application as a variable and every effort should be made to hold that as a constant for all systems being compared. If, however, a new system is being evaluated specifically because it may be less sensitive to pre-processing variables, these variables should be included in the test plan. The range of the variables to be considered in this case should be those allowed by the procedures used at the application site.

2. Inspector: In many applications the human conducting the inspection is the most significant variable in the process. Conversely, some inspection systems have been demonstrated to be very inspector-independent. The test plan should include the inspection results obtained by several operators selected at random from among the population eligible to conduct the inspections. Eligibility may be defined in terms of some particular certification, training, or physical ability.

3. Inspection materials: Particular chemicals, concentrations, particle sizes, and other material-dependent variables may be used in a given inspection. For example, PT inspections will use penetrants, emulsifiers and developers, each of which may have a significant impact on inspection capability. System evaluation may be conducted considering the range of materials expected to be used in production. If different penetrants, for example, are used, the penetrant should be considered as a variable in defining the test matrix. If the operating procedures for the system preclude the use of alternate penetrants, others need not be included, but this restriction clearly limits the generality of the system assessment.

4. Sensor: If the sensor used in the inspection system is replaceable, or if different sensors are used for different applications of the system such as is the case for eddy current or ultrasonic inspections, sensors should also be a variable in the test matrix. The sensors used in the demonstration tests may be selected at random from a production lot. Sensor designs typical of each planned for use with the system should be included in the test plan, with several of each being evaluated.

5. Inspection setup (Calibration): Electronic inspection processes in particular require instrumentation adjustments to assure the same sensitivity inspection independent of time or place. To evaluate the potential variation introduced to the inspection process by this calibration operation, the test matrix should include calibration repetitions, allowing random variations that are consistent with the process instructions. If more than one calibration standard is available (e.g., production sets), the effect of the variation between standards should also be considered as a test variable by repeating the specimen inspection after calibrating on each of the available standards.

6. Inspection process: The inspection process specifies controls on such inspection parameters as dwell time, current direction, scan rates, and scan path index. The system test matrix should include evaluation of these parameters. If an allowable range is specified, the test plan should evaluate the inspection at the extreme of this range. If the parameter is automatically to be held constant, repetitions of the basic inspection may be sufficient evaluation of this variable.

4.3.1.2 Test matrix.

The contractor should generate a test matrix to be used in the reliability demonstration. The test matrix is a list of planned process test conditions which collectively define one or more experiments for assessing NDE system capability. A process test condition is defined as a set of specific values for each of the process variables deemed significant (see Appendix E). The complete set of test specimens should be inspected at each test condition of the test matrix. The complete matrix can comprise more than one experiment to allow for preliminary evaluation of variables which may only marginally influence inspection response of the system. To the extent possible, the individual inspections of a single experiment should be performed in a random order to minimize the effects of all uncontrolled factors which may influence the inspection results.

a. The inspection test conditions are to be representative of those that will be present at the time of a future inspection. Therefore, to eliminate potential bias, the values assigned to each test variable in a test condition may be selected at random from the population of possible values for that variable. For example, if a future inspection is to be performed by any of a given population of inspectors and three inspectors are to be included in the experiment, then the three inspectors should be chosen at random from the population. Similarly, if two different probes of identical design are to be used in the experiment, they should be selected at random from the population of probes. Note, that if the population of probes (or inspectors) includes those not yet available, it may be assumed that the available probes (or inspectors) are representative of those that may be obtained in the future.

b. The analysis methods for combining multiple inspections in the calculation of a single POD(a) function with confidence limits requires that the levels of all of the variables be balanced. This is most easily achieved when the test matrix comprises a full factorial experiment in which all combinations of all levels of the variables are in the test matrix. It is readily apparent that factorial experiments can rapidly lead to very large test matrices. There are other methods of designing balanced experiments in the statistical literature which do not require all combinations of the levels of the variables (see Appendix E, and Box, Hunter, and Hunter (1978)). These can and should be employed when necessary.

c. In general, a final test matrix is a compromise between the number of variables that can be included, the number of levels (values) for each of the variables, and the available time and money. To ensure that all desired objectives of the demonstration can be met, it is imperative that all trade-offs be evaluated before inspections begin.

d. It should also be noted that experiments to evaluate the effects of inspection process parameters on POD can be designed and analyzed using the methods of appendices E, G, and H. Such experiments should be performed prior to the capability demonstration as a planned approach to optimizing the process.

4.3.2 Test specimens.

The test specimens may reflect the structural types that the NDE process will see in application with respect to geometry, material, part processing, surface condition, and, to the extent possible, flaw characteristics. Since a single NDE process may be used on several structural types, multiple specimen sets may be required in a reliability assessment. The contractor should determine the characteristics of the test specimens required for the demonstration and recommend the required number of flawed and unflawed specimens. All test specimens available to the contractor should be evaluated to determine if existing test sets meet the requirements of the reliability demonstration. The contractor should insure that the specimens should not become familiar to the inspectors or inspection system. Specimens which have

become familiar to the inspectors or the inspection system will bias the resulting POD(a) curves and so will be considered as unsuitable for reliability demonstration. When necessary, new specimen sets should be designed and fabricated to meet the requirements. A plan for maintaining and re-validating the specimens should be established. All of these results should be documented in the Demonstration Design Document. The following subparagraphs present minimum considerations in obtaining and maintaining the demonstration test sets. Further guidelines for fabricating, documenting, and maintaining test specimens are presented in Appendix F.

4.3.2.1 Flaw sizes and number of flawed and unflawed inspection sites.

The statistical precision of the estimated POD(a) function depends on the number of inspection sites with flaws, the size of the flaws at the inspection sites, and the basic nature of the inspection result (hit/miss or magnitude of signal response). Unflawed inspection sites are necessary in the specimen set to insure integrity and to estimate the rate of false indications. The following recommendations are made regarding these topics:

a. The flaw sizes should be uniformly distributed on a log scale covering the expected range of increase of the POD(a) function. Cracks which are so large that they are always found (or saturate the recording device) or so small that they are always missed (or yield a signal which is obscured by the system noise) provide only limited information concerning the POD(a) function. Since the region of increase of the POD(a) function is initially unknown, only engineering judgment can be made regarding this range of increase. It should be noted that there is a tendency to include too many “large” flaws in NDE reliability demonstrations.

b. To provide reasonable precision in the estimates of the POD(a) function, experience suggests that the specimen test set contain at least 60 flawed sites if the system provides only hit/miss results and at least 40 flawed sites if the system provides a quantitative response, â, to a flaw.

c. To allow for an estimate of the false call rate, it is recommended that the specimen set contain at least three times as many unflawed inspection sites as flawed sites. An unflawed inspection site need not necessarily be a separate specimen. If a specimen presents several locations which might contain flaws, each location may be considered an inspection site. To be considered as such the sites may be independent, that is, knowledge of the presence or absence of a flaw at a particular site may have no influence on the inspection outcome at another site. It is advisable to have at least 10 - 20 unflawed specimens for PT testing.

4.3.2.2 Physical characteristics of the test specimens.

The final geometry of the specimen should represent to the NDE method to be used, the same degree of difficulty as the critical areas of the components to be inspected. Specimens may represent the shapes of the actual hardware for inspections where problem manipulation or inspection media (such as magnetic field, sound waves, and line of sight) are geometry dependent. Bolt holes, flat surfaces, fillets, radii, and scallops are some typical shapes that influence inspections. Residual stress may influence the inspection due to configuration. Another geometric consideration for all inspection techniques is flaw location, for example, corner flaws versus surface cracks. Flaw location on specimens may be oriented and positioned to represent actual parts.

a. The initial geometry of the specimen should allow the insertion of flaws of the required shape and size in the specified locations. The specimen should be designed such that the required flaws can be inserted, and then the final geometry can be obtained by machining or other forming methods that will also retain the flaws of the necessary size, shape, and

orientation and be within 0.002 inches of the intended locations. Specimens should be manufactured to tolerances typical of the component they represent.

b. For UT, ET, PT, and MT methods, the contractor should select alloys, material forms, and raw material processing that represent the physical properties (of the components to be inspected) significant to the NDE method being evaluated. For example, if an actual part is made of INCO 718, forged to near finished shape, the specimen should be made of INCO 718 and fabricated by the same processes. In addition, for ultrasonic inspection, the internal noise and attenuation should be as defined by the statement of work for the components to be inspected. For magnetic particle inspection, the magnetic properties should be comparable to the components to be inspected.

c. The processing (forged, cast, or extruded) of the raw material and the heat treat are critical to insure that the specimen simulates the same metallurgical properties as the actual part. Surface condition of the final product and specimen will influence all inspection signal to noise ratios. Some examples are as follows: Grain size can have a large influence on signal to noise ratio for ET and UT, and magnetic field for MT. Also, processing can develop mechanical properties which can influence PT results. Material strength can influence the amount of smear metal which can obscure defects from penetrant inspection and residual compressive stress may influence PT or UT. Residual stresses can also be influenced by flaw propagation (flaws grow to relieve the stress field in which they reside) and final machining. Final machining of the specimen should be consistent with final machining of the part. The surface finish of the specimen and actual part should be consistent so that the common surface finish between specimen and part provide similar signal responses. For example, if the part is turned on a lathe, the specimen should be turned on a lathe whenever possible. If the surface texture of the part and specimen are not similar, for instance “record groove” finish on the part due to lathe turning and ground finish on the specimen from grinding, the false call rate may be higher on the parts due to the macro finish of record groove even though the micro surface finishes are similar.

4.3.2.3 Specimen maintenance.

The contractor should derive a plan for protecting the specimens from mechanical damage and contamination that would alter the response of the NDE process for which they are used. This plan would require as a minimum that the specimens be:

1. Individually packaged in protective enclosures when not in use;
2. Carefully handled when in use;
3. Cleaned immediately and returned to the protective enclosure after each use;
4. Re-validated at intervals specified by the contracting agency when the specimens are intended for periodic usage.

4.3.2.3.1 Specimen flaw response measurement.

Specimen flaw responses will be measured periodically by (the contractor, as monitored by the appropriate procuring activity) using the same test technique and procedure used in the original specimen verification (see Appendix F). The flaw response may fall within the range of the responses measured in the original verification process. If it does not, the results may be examined to consider if they are acceptable, if the specimen has been unacceptably compromised, or if the specimen needs to be re-characterized and verified.

4.3.2.3.2 Multiple specimen sets.

When multiple specimen sets are required for periodic use, the contractor should initially select one set as a master set. The remaining sets should be demonstrated to have a response within a specified tolerance of the master set. Periodic re-verification against the master set can then be performed.

4.3.2.4 Hardware specimens.

Note that in many cases when a development system is first being evaluated, the specific part geometries and surface conditions may not be known, or if known, representative flawed specimens may not be available. This reemphasizes the necessity for the inspection of actual hardware as a part of the qualification program. Again, these may not reflect exactly the conditions to be seen in the specific application of the system, but they will be significantly more realistic than just the laboratory flawed specimens. The parts should have defects in them to provide signals for the inspection. For ET and MT systems, EDM notches may be sufficient for evaluating scan plan coverage but are inadequate to assess system response to actual fatigue flaws. For UT, drilled holes may be preferable; for PT, fluorescent markings may be the best available, though they may be too bright to verify system capabilities. An ideal test would use actual service flawed hardware, if a representative selection of such parts can be collected.

4.3.3 Test procedures.

The contractor should develop and report a detailed plan for executing the demonstration tests at the application facility. The procedures to be used in the demonstration may follow the procedures and work instructions planned for the production inspection of parts. This includes all fixed process parameters, data analysis algorithms (for automated systems), accept/reject criteria and other items covered by the System Configuration Control Document. The inspections should be performed by production inspectors, as designated by the experimental design. A test monitor should be designated who should assure that all requirements of this handbook are being met both prior to initiation and during the performance of the tests.

a. Every inspection technology depends on certain conditions being met that the operator may not be able to verify as a part of the daily inspection setup. Examples of this may include the scan speed or index of mechanical manipulators, the drive frequencies of eddy current or ultrasonic instruments, or the purity of chemicals or solutions being used. Prior to the NDE system evaluation, it is important that significant variables such as these be calibrated. It is suggested that this be done using NIST-traceable standards and procedures. Note that any nonconformance not corrected will likely degrade the NDE system performance. Periodic recalibration of the NDE system after acceptance should be conducted in accordance with local procedures.

b. In addition too specific requirements of the NDE process (Section 5 and Appendices A,B, C, or D), the following should be considered in the development of the test procedure plan:

1. System software controlling any data collection, reduction, and processing should be that planned for use in production implementation. Any differences between the test and reality could negate the ability of the POD curve to be applied to the actual testing situation.

2. Appropriate fixturing of specimens can make the inspection procedure similar to actual parts; that is, the demonstration fixturing and the actual component would ideally have the same inspection system arrangement of probe, orientation, manipulation, and scan plan.

3. Signal evaluation and decision levels used during the testing should be those planned for use in production. In many cases it may not be known in advance what thresholds can be

practically implemented in production, in such a situation the detection capabilities should be established as a function of these process parameters.

4. Scanning motions for the demonstration tests should be similar to those planned for production. This similarity should extend to the manipulator axes used, feeds and speeds, alignment routines (such as eddy current bolthole probe centering), and scanning procedures. This may not be strictly possible for the inspection of some of the low cycle fatigue (LCF) specimens, but every effort to achieve similarity should be made.

5. Accurate data acquisition, recording, and documentation are also important. The data should be recorded in the form which is compatible with the disposition of the part. For example, an eddy current inspection may record the data as voltage output of signal â or a signal-processed calculated "depth" $\hat{d}$. If the part is to be rejected by $\hat{d}$, (which is not a recommended practice) but the demonstration data were recorded and analyzed in â, the reject standard separating good from bad parts would necessarily be in terms of â. Therefore, the reject level for actual parts would be unknown, because â cannot be easily converted to $\hat{d}$, which is based on some signal processing algorithm rather than the mandatory break open data for specific geometries and stress fields. The test would then have to be repeated and the appropriate data, $\hat{d}$, in this example, collected and then reanalyzed in the appropriate metric, $\hat{d}$. Proper planning prior to data collection will avoid such difficulties and provide meaningful results the first time.

4.3.4 Demonstration process control.

The contractor will develop a plan for insuring that the NDE process is in a state of control at the start of the demonstration and remains in a state of control throughout the demonstration period, regardless of length of time. The plan will include routine quality, instrumentation, and calibration checks, and should also incorporate inspection responses to real structure or specimens. The process control plan should be the basis for process control during extended periods of production inspections using the system (see 4.2).

4.4 Demonstration tests.

The sets of inspections as defined in the Demonstration Design Document should be carried out at the production inspection facility under normal operational conditions. The test monitor should be available during all testing. Inspectors should inspect all specimens in accordance with the Demonstration Design Document, the matrix of test variables, the applicable NDE process specifications, and any work instructions deemed necessary for the inspection of the test specimens for the reliability test program. The inspection procedures should conform to the test procedures used for production components, modified only as necessary to accommodate the test specimen configuration. A log should be kept of the inspections, showing the order in which the inspections were performed, the inspector who performed the inspection, the specification identification and serial number, and the date and time the inspection was performed.

4.4.1 Inspection reports.

The inspector should prepare a report (or collect required data from automated reporting systems) on each inspection performed. The reports should be delivered to the test monitor and should contain, as a minimum, the inspector identification (possibly coded), specimen identifications including any serial numbers, inspection date and time, and the results of the inspections including the NDE responses and locations of any indicated defects. The data collection may be compatible with the reporting requirements of 4.5.

4.4.2 Failure during the performance of the demonstration test program.

In the event of failure in one or more of the systems during the performance of the demonstration test program, the contractor should remedy the cause of the failure. The periodic evaluation (see 4.3.4) for assuring that the process is under control should be performed to assure that no problems have arisen due to the failure. The particular matrix element being evaluated at the time of the failure should be completely reevaluated.

4.4.3 Preliminary tests.

With the agreement of the contracting agency, preliminary tests of the system may be carried out at the contractor's facility. Tests at the contractor's facility, however, should be directed toward preliminary acceptance and the results should not be used to modify hit/miss decision criteria.

4.5 Data analysis.

The purpose of the NDE demonstration is to produce quantitative descriptions of inspection system performance, POD(a) curves, and statistics for comparing NDE systems based on these curves and statistics.

a. Inspections can be grouped into two categories: those for which only the inspection outcome is known, hit or miss, and those providing additional information as to apparent flaw size, â vs. a.

b. The analysis of these data to produce POD(a) curves should be accomplished using a standard IBM PC computer program to be supplied by the Air Force. The latest version of the program and User's Manual for Probability of Detection Software System (POD/SS) can be obtained from ASC/ENFP, Wright-Patterson AFB, OH 45433.

4.5.1 Missing data.

All of the inspections called for by the test matrix should be performed. If the design of the experiment is a factorial (all possible combinations of the factors being varied) and some of the inspections are not performed, the POD analysis program cannot be directly used. The assistance of a professional statistician is recommended to assist in the evaluation of such data. If the experiment is designed to evaluate only the variability associated with different flaws and one other factor, the POD analysis program will provide valid answers even if some of the inspections are not performed.

a. Note that the program distinguishes between a missing inspection (i.e., no inspection result was obtained) and a missed flaw (i.e., the inspection was performed but the flaw was not detected). See program users manual for details.

b. A description of the statistical methods employed to generate these curves for both types of NDE data, the procedures for estimating their confidence limits, and analysis techniques for comparing POD curves is provided in Appendices G-H.

c. The design of the NDE demonstration (4.2 and Appendix E) provides the foundation for the entire system evaluation. No amount of clever analysis can overcome a poorly designed experiment.

4.6 Presentation of results.

The contractor should submit a permanent record of data and a summary test report for each NDE reliability experiment. To facilitate potential inclusion into a database, the data should be partitioned into four areas:

1. The description of the NDE system,
2. The experimental design,
3. The individual test results, and
4. The summary test results.

Each experiment should be assigned a unique identification. The identification should comprise codes which identify the NDE method, the NDE system, the inspecting organization, the type of specimens, and an experiment number. The identification numbers will be assigned by the Air Force. The experiment identification code is the tie between the four data types. Data included in one of the categories need not be repeated in another but, for ease of access, general information should be repeated on the various submittal forms. The data to be submitted for the permanent record should be from all four categories and should comprise data sheets, tables, and plots as described below.

4.6.1 Category I - NDE system.

The System Configuration Control Document may be sufficiently detailed to account for all factors which have a major influence on the accept/reject decision. The purpose in recording this information is to specifically identify the system that was evaluated. If the results are to be extrapolated to different, but similar, systems, it should be possible to identify and evaluate the sources of potential differences between the systems. The minimum information required in the description of each NDE method is listed in the data sheets in the specific requirements of Section 5 and Appendices A through D.

4.6.2 Category II - Experimental design.

The experimental design identifies the specimen set to be used in the demonstration; the test matrix of the levels of the factors of the controlled variables and the number of replications of test conditions; and the order in which the steps of the test matrix are to be run. Note that the specimen set determines the number of flaws in the experiment while the number and levels of the controlled factors determine the number of inspections of each flaw. All specimens would be subjected to the inspections that are specified by the combinations of the levels of the controlled factors of the Demonstration Design Document.

a. Sample data report sheets are included in Appendix J, and discussed as an example here. Assume that the assessment of an eddy current system was to include the effects of two operators, two probes, and two replications. An example data sheet for reporting this data is presented in the list of the test combinations of figure 21. The same information is contained in the table of test conditions of figure 22. This latter format is unwieldy if the experiment contains more than four factors or more than three levels of the factors. However, the table format more clearly shows the levels of all of the factors being evaluated and could assist in the analysis of the data.

b. A unique test identification is assigned to each combination of levels of the factors (each line of the test matrix) to facilitate reporting individual test results. The test identification in the examples correlate exactly with the levels of the experimental factors. This degree of

identification refinement is not necessary but if consistently used aids in the interpretation of data from different experiments.

4.6.3 Category III - Individual test results.

The data collected during the actual inspections are not necessarily the data to be recorded in the permanent individual test result of the experiment. However, the original data should be preserved by the organization conducting the experiment to resolve problems which may arise. In general, inspection result data sheets should be obtained from the original data recordings and should summarize the findings of all inspections of each flaw. Figure 23, is the data sheet for the permanent record of the individual test results of an inspection experiment. Figure 23 also arranges the data in a convenient format for input to the analysis programs. A magnetic disk containing the inspection result input files in IBM P/C-compatible format should be submitted with the summary of experimental results.

4.6.4 Category IV - Summary results.

Summary results are obtained from the analysis of the individual test results for a particular experiment. These may include POD(a) function parameters, plots of POD(a) functions, plots of log â versus log a, verification of assumptions of the analysis, and an analysis of the significance of test variables (if called for by the objectives of the experiment) as specified by the CDRL. All of this information should become part of the permanent record of each NDE experiment

a. The PC software analysis program should automatically output the required summary statistics for a given analysis. When requested, the program should also generate files for plotting POD(a) vs. a, the lower confidence bound on POD(a) versus a, the observed detection probabilities for each flaw vs. a, and log â vs. log a. Figures 24 and 25 are examples of summary output from â vs. a and hit/miss analyses, respectively. In both of these examples, the analysis provided complete sets of parameter estimates. If the likelihood equations cannot be maximized for a particular data set, the program so indicates. In either type of analysis, if the probability of detection is not significantly related to flaw size, the lower confidence bound on the POD(a) function will not be monotonically increasing. In this case, the program does not output an estimate of a lower confidence bound on POD(a) and writes a message that the model does not adequately fit the data. Tests of the assumptions of the analysis should be made on the basis of the log â vs. log a data (for â vs. a data) and from the superposition of the POD(a) function on the observed detection probabilities (for hit / miss data). Other analysis procedures are discussed in Appendices G and H. All departures and potential discrepancies from the standard analysis should be specifically identified and reported.

b. Figures 26 and 27 are the POD(a) functions and 95 percent confidence limits for the example analyses of figures 24 and 25, respectively. These figures indicate the information that may be included on all plots of POD(a) functions when used to illustrate the capability of an inspection system for each of the basic types of inspection data. Figure 28 presents the log â vs. log a data for the analysis of figure 24. These plots may be generated for all sets of â vs. a data. Any deviations from assumptions (e.g., restricting the set of test flaws to a range of linear log â vs. log a) may be corrected prior to analysis or specifically noted on all characterizations of the capability of the system. In the hit/miss type of data, the estimated POD(a) function should be compared to the detection probabilities for each flaw in the specimen set as in figure 29.

4.6.5 Summary report.

The results of each capability experiment should be documented in a summary report as specified by the CDRL. This report should interpret the results of the experiment and conclude whether or not the system met specifications. If the system failed to meet the specification, the cause and reason for the failure should be identified. Future actions regarding qualification of the system should be presented. As a minimum, this report should contain the following information:

a. The NDE system description data sheet;

b. A description of the factors being included in the experimental design and the levels of each factor;

c. The output summary sheets from the analysis;

d. Plots of log â vs. log a, if applicable;

e. Plot of the properly annotated POD(a) function and its lower 95 percent confidence bound;

f. Plot of the POD(a) function superimposed on the observed detection-probabilities for hit/miss data;

g. A statement concerning the validity of the assumptions of the analyses linear relation between log â and log a and approximately equal scatter of the residuals;

h. Identification of significance of test factors and interpretation in terms of capability characterization; and

i. A statement of conclusions and recommendations for further actions.

4.6.5.1 Summary report documentation.

More than one experiment can be documented in the same report but the information from each experiment may be contiguous. Comparisons of data from different experiments and extensive summaries across comparable experiments are recommended whenever possible.

4.7 Retesting.

If the system does not meet the capability and reliability requirements of the contract, the contractor should conduct a review of the possible causes for the failure. This may include some of the multi-factor statistical analysis described in Appendix E as well as function tests on the various subsystems. A plan, which includes a discussion of the possible causes for the failure, should be generated which describes how the system will be modified and what additional testing will be performed. This new plan should be, in effect, a second Demonstration Design Document (see 4.3), except that it should also include the discussion of the possible reasons for the failure and resolutions.

4.8 Process control plan.

After the system has been demonstrated as reliable by satisfying the requirements as specified in the Data Item Document (DID), the contractor should provide a written plan for assuring that the process is under control. This plan should include a periodic evaluation of the processes involved including all mechanical, electrical, calibration, and computing systems. Control charts or other proper permanent records should be required as an integral part of the plan.

5. DETAILED REQUIREMENTS

5.1 General.

The detailed requirements for determining the test and evaluation NDE procedures are contained in Appendices A through D. The contractor should establish the basic process parameters prior to conducting the reliability demonstration. Once the demonstration has been completed, the process parameters used in the demonstration should not be changed without another demonstration program which shows the effect of changing the parameter. The reliability of the system, the overall POD curve, and the lower bound will be determined as a result of some sort of statistical experimental design. A factorial design is preferred. A discussion of a factorial design and the sampling approach is given in Appendix E.

6. NOTES

(This section contains information of a general or explanatory nature that may be helpful.)

6.1 Intended use.

This handbook is intended to provide procedures for assessing NDE inspection capability that will permit quantitative comparison of one system with another with respect to known specimen standards.

6.2 Trade-offs between ideal and practical demonstrations.

Ideally, the test designed according to this document should include all variables of concern in the test matrix. The conditions found in real part inspections should be matched exactly. In reality, these constraints cannot always be made. For example, the number of different geometries in a complete engine, and the requirement that each be tested as suggested by the ideal test design, may drive testing costs and times to the point where it is impractical to do such a test. This same situation could involve test parameters, probes, and mechanical parameters. The number of parameters that could possibly be tested is immense.

6.2.1 Solution.

The solution to this problem is to allow the terms reasonable and representative to govern any concessions made to reality. The term reasonable argues for a balanced definition of the test, one which does not force the ideal too much. Important variables should be tested, while unimportant variables may not have to be tested. It implies avoidance of extremes in testing, and application of logical considerations in compromise. The term representative also argues for limiting the number of variables tested, but in a manner which gives reasonable representation of the real inspections. This philosophy of testing recognizes that not all variables will be tested, and accepts that some areas of inspection will be better than the test and some will be worse. By being reasonable and representative, a good quality test can be designed which will satisfy cost and time constraints. As mentioned elsewhere, the final test design may be submitted to the customer for approval, and becomes part of the design document.

6.3 Other topics.

The following notes are included as examples of ongoing work related to NDE system evaluation. The work has not progressed sufficiently to include these topics as standards, yet they are important and should be considered as part of any technical update of this document.

6.3.1 False call analysis.

When an inspection stimulus is applied to detail, the interpretation of the response determines whether or not a crack is judged to be present. Presumably, the inspection system is designed to produce a clear, unambiguous response to all cracks whose sizes exceed a specified value. If noise (from whatever source) is present in the signal response, false indications (false calls) can result if a noise response from a noncracked detail is interpreted as being caused by a crack. Although false indications are undesirable for economic reasons, they cannot be entirely eliminated since there is a trade off between the rate of false indications and the ability to detect very small cracks.

6.3.2 Rates of false indications.

Rates of false indications are currently quantified by a count of the number of indications that are given at locations for which no known crack is present. There have been data sets for which the false call rate was so high that very small “detected” cracks were more likely to be false indications at crack sites. These data produced POD(a) functions that did not adequately model the observed results. To incorporate the simultaneous estimation of the parameters of the POD(a) function and the false call rate, a modified analysis is being considered. This new model is based on the probability of obtaining an indication (rather than detection) at an inspection site.

a. Let POD(a) represent the probability of obtaining an indication in an inspection of a crack of size a. Let p represent the probability of a false indication for the inspection which depends on the inspection method, the inspector, the calibration, etc. Then

$$POI(a) = p + POD(a) - Prob [\text{ false call and detection }]$$

(Note that an inspection response signal could be such that both the response and the noise levels would be large enough to produce a crack indication). If the probability of a simultaneous detection and false indication are independent.

$$POI(a) = p + (1 - p)\, POD(a)$$

b. While this expression may be a reasonable model for the joint estimation of p and the parameters of the POD(a) function, the implementation of the model by maximum likelihood is not straightforward. Other approaches to estimating the parameters and placing confidence limits on the POD(a) function are being sought. At present a maximum false call rate of 5% is suggested to ensure proper POD(a) representation.

6.3.3 POD from multiple inspections.

Redundant inspection is the practice of performing multiple inspections on a single part. The philosophy behind multiple inspections is to increase the probability of detecting a flaw which may exist. If the POD fails to meet CDRL requirements, it may be possible to use redundant inspections to shift the POD curve and its lower bound.

a. Historically, calculations expressing the benefits of redundant fluorescent penetrant inspection have been made assuming complete independence between inspections. For example, if the probability of detecting (POD) a flaw of a certain size is 0.9, then the probability of a single miss (POM) is 0.1, the probability of two (independent) misses is 0.1(0.1) = 0.01, and so the POD for two inspections is 1 - 0.01 = 0.99, assuming independence.

b. Unfortunately, most inspections have been found to be not independent inspection-to-inspection. Events which cause this dependency include inspection of the same crack twice

(location, size, etc.), or the same inspector may investigate the crack twice, or the surface of the part, and the crack itself, may not be restored to its initial state between inspections.

c. In reality, quantifying the POD due to multiple inspections requires knowledge of this dependency. For double inspections, the calculation is:

POD(A or B) = POD(A) + POD(B) - POD(A and B)

where these POD equations are calculated as described in Appendix G, Modeling Probability of Detection, and where A and B refer to two inspectors.

d. Assuming that inspector A and inspector B equally share the responsibilities for flaw location, the difference between single and double inspections assuming inspection-to-inspection dependency can be expressed as:

Increase in POD = (POD for double inspection) - (POD for single inspection)

= {POD(A) + POD(B) - POD(A and B)} - {0.5 POD(A) + 0.5 POD(B)}

= 0.5 POD(A) + 0.5 POD(B) - POD(A and B)

This argument can be extended for multiple inspections greater than double inspections, or for a process parameter other than inspector, or for a system other than PT where redundant benefits may be needed.

e. For more details please see “Quantifying the Benefits of Redundant Fluorescent Penetrant *Review of Progress in Quantitative Nondestructive Evaluation*, Vol. 8B pp. 2221-2228.

6.3.4 Inspection of EDM-notched parts.

System Probabilities of Detection (PODs) established using the procedures of this handbook characterize the sensitivity of the system to the flaws in the specimens tested. The applicability of these PODs to the inspection of actual hardware is dependent upon the extent to which the specimens mirror the actual part conditions. That they are not perfect reflections is due to limitations in such factors as:

a. Full part geometry is not reproduced (e.g., dovetail slant, part radius curvature).

b. System manipulation routines are different (since not testing full parts).

c. Only typical geometries are represented, a full set of all features inspected is prohibitively expensive.

d. It may be difficult to initiate defects in the specimens that duplicate the positions, sizes, and shapes of flaws that are the targets of the part inspections.

6.3.4.1 Evaluation of applicability of PODs.

To make some estimate of how directly the established POD curves may be applied to the inspections of the parts it is appropriate to inspect actual hardware with artificial flaws machined in the critical locations. Note that the purpose of this test is not to modify the PODs already generated, but to evaluate their applicability to production inspections.

6.3.4.2 Example of eddy current inspection.

The rest of this discussion will use as an example eddy current inspection of EDM notched parts. The notches used for these tests may be sized to provide an â that can be referenced to the calibration, or to provide eddy current â values approximately equal to those of the crack sizes to be detected in the production inspections. The steps in establishing the size of this notch are as follows:

a. Determine the inspection goal (e.g., detection of a 0.010 inch crack in the part).

b. Determine from the POD testing the average â of this size crack in the specimen(e.g., 100 counts).

c. Machine several size notches in specimen blanks, to determine the size notch that yields an â of the same 100 counts level (interpolation on a log-log plot may be necessary).

6.3.4.2.1 Resolution of variances.

Notches may then be machined into the part features to be inspected. Significant variations of the notch â values from those expected may indicate that the POD curves established using the specimens may not be directly applicable to those part features being inspected. The causes of this, and some means of establishing representative PODs should be examined.

6.3.5 Ill-behaved data.

Because of an inadequate number of observations or an inappropriate range of flaw sizes, some inspection results contain little information, and taken by themselves, give nonsense POD(a) curves. One possible approach in this situation would be to simply declare the data unusable. This may ultimately prove to be the most prudent procedure.

a. However, there is some engineering information contained within these observations. A better idea might be to extract that information and evaluate it in light of prior knowledge about similar inspection processes. Then decide if more testing is required to augment (or replace) the data under consideration.

b. Bayesian statistics provides the framework for this analysis. The overall plan is to define the likelihood in terms of the observed data (as is currently done) and in terms of the expected. parameters values, based on prior experience. Parameter estimates can then be selected such that this new likelihood function achieves a maximum.

c. For this approach to be effective, the influence of the prior information should be small, when the data are well behaved, and only moderate otherwise. If the influence of the “prior” (as it is called) is too overwhelming, what little information contained within the data will be obscured and the entire exercise will be of no practical value. The prior, therefore, should provide stability to the data, without undue influence on the final outcome.

6.4 Subject term (key word) listing.

Nondestructive Evaluation (NDE)
Nondestructive Inspection (NDI)
Probability of detection (POD)
Reliability

APPENDIX A

EDDY CURRENT TEST SYSTEMS

A.1 SCOPE

A.1.1 Scope.

This appendix provides the detailed requirements and methods for testing evaluation procedures for assessing NDE system capability requirements for eddy current test compliance.

A.1.2 Limitations.

Eddy current test NDE procedures addressed in this appendix are those used to inspect gas turbine engine components.

A.1.3 Classification.

Eddy current test is classified using quantitative measurement, â vs. a data.

A.2 APPLICABLE DOCUMENTS

This section is not applicable to this appendix.

A.3 DETAILED REQUIREMENTS

A.3.1 Demonstration design

A.3.1.1 Test parameters.

The demonstration design for the capability and reliability of the eddy current system should include, but not be limited to, the following test variables. These requirements are in addition to those listed in Section 4.3.

a. Inspector Changes
b. Sensor Changes
c. Loading and Unloading of Specimens
d. Specimen Position
e. Calibration Repetition
f. Calibration Standard Variation, if applicable
g. Test Repetition

A.3.1.2 Fixed process parameters.

Fixed process parameters should include, but not be limited to, the following. These parameters will be required to mirror actual production inspection. Some of these parameters may be included in the matrix of test variables, if desired.

a. Drive frequency
b. Coil frequency and design
c. Probe body and/or holder design
d. Scanning technique
    1) Index amount
    2) Scanning speed
e. Digitization rate, if applicable
f. Digitization resolution, if applicable

g. Threshold levels

h. Filter values, low-pass and high-pass

i. Hardware and software configuration control number

A.3.2 Specimen fabrication and maintenance.

Specimens for the evaluation of eddy current inspection systems should have surface connected flaws, generated as described in 4.3.2. Following the initiation of the cracks and the grinding off of the EDM notches, the specimens should be further stress cycled to break the crack through any metal that may have been smeared over the cracks. At that time, the crack lengths should be measured. This is best done by loading the specimen to 60% of the load used to grow the cracks, and optically measuring the length using a 40 X magnifier. To characterize cracks further, a representative sample should be dyed or heat tinted and the cracks broken open, to confirm the surface length measurements and to establish the crack depths and shapes.

A.3.2.1 Crack area or crack depth.

Either crack area or crack depth, as agreed to by the Air Force, can be used to characterize the cracks. To make this more readily relatable to the detection requirements for a given application, this area can be expressed in terms of the radius of a sector of circular crack of that area. The sector is a quarter circle for corner cracks and a half circle for surface cracks. Actual crack aspect ratio (ratio of surface length to depth) is to be determined by break open procedures. The inspectors should be provided the orientation of potential cracks in the specimens, but should not know if a particular specimen is cracked, or if cracked, the specific location of those cracks.

A.3.2.2 Specimen maintenance.

The eddy current process would not itself degrade the specimens' condition, so no special precautions need to be taken for specimen maintenance beyond those listed in 4.3.2.3. An exception is the practice of touching the part with a metal probe during the part alignment, such as is sometimes used with a typical non-contact bolthole or scallop inspection. In this case, the test procedures may clearly prohibit this practice, to prevent damage to the cracked specimens.

A.3.3. Testing procedures

A.3.3.1 Test definition.

Procedures should be written prior to the test, clearly describing what tests are to be conducted, and the exact procedures for conducting them. They should be to the same level of detail as the day-to-day procedures to which production inspectors operate. In addition to those items outlined in Section 5.1.1, other items to be specified in this test definition are the following:

a. Part preprocessing requirements as appropriate. This is more of an issue for the inspection of actual production engine parts, preprocessing of the test specimens should be limited to cleaning only.

b. System inspector requirements. This will frequently refer to qualification/training requirements, but will also include the number of inspectors to be included in the test plan. At the start of the test matrix this may typically call for three inspectors to be involved in the system evaluations. This number is specified by the demonstration design.

c. Inspection materials are not a significant variable for eddy current inspections.

d. Depending upon the degree of system automation, sensors may be the most significant variable to be considered. The test plan should require the evaluation of the system using at least two samples of each distinct coil type used (such as end mount or side mount absolute coils, differential, reflection, printed circuit, etc.). The probe body needs to be a factor in this evaluation only to the extent necessary to allow inspection of the specific specimen designs.

e. Inspection setup (calibration) may be conducted using the same procedures planned for use in production. The signal responses may be set to the same values, with the same tolerances in both situations.

f. The production inspection process should be duplicated in the tests as much as possible. Thus the inspection feed rates, scan index rates, drive signal frequencies, filter settings and any signal processing may be the same. Because the cracked specimens may differ physically from the real parts to be inspected in production, the scanning motions for the specimens may necessarily differ from those used for the parts. Efforts should be made to minimize the differences, and recognized differences should be documented. For automated systems, software package version and revision numbers may be specified.

g. Inspection thresholds used in the test should be the same as those planned for production use. Inspection of the actual engine part specimens should help to establish how realistic those thresholds are for production inspections. Where the specific application of the system is known, typical production parts should be used to determine practical thresholds. It may be desirable to inspect the specimens at as low a threshold as possible, to establish the detection capabilities as a function of thresholds used. This will allow trade-offs between detection capability and production throughput to be made.

A.3.3.2 Test environment.

The environment in which the test is run should match the anticipated production environment as closely as possible and be conducted at the production site if possible. If the system is a new development, the initial tests may need to be conducted at the manufacturer's facility. To the extent possible, production conditions should be met. It is suggested that the manufacturer conduct a first evaluation prior to shipping the equipment and a second test one or two months after the system is installed on site.

A.3.4 Presentation of results.

Documentation of test results should include all raw data from the tests. If some of the data is classed as irrelevant and not included in the data reduction process, this should be noted, and an explanation given for why this decision was made (an indication was subsequently demonstrated to be due to a power surge, or to inadequate cleaning of the specimen, for example.) This provides the customer the option of accepting or not accepting that rationale.

A.3.4.1 Submission of data.

Data for the permanent record of eddy current NDE reliability experiments will be submitted in accordance with the requirements stated in Section 4.6. Figure 1 presents an example of the type of information required for description of eddy current inspection systems. Eddy current data should be in the â vs. a format and analyzed accordingly (see G.3.2).

Date:_____

Operator ID:_____

Part Number _____ Serial Number_____ Alloy_____

Engine _____ Part Number _____Surface Roughness_____

Attach Specification Sheet _____System Operating Ambient Temperature_____

State other Equipment Environmental constraints _____

Test Frequency_____Scan Speed _____Filtering_____

Horizontal Gain_____Vertical Gain_____Lift-Off-Technique_____

Coil Output Impedance _____

Probe

Contact _____ Noncontact _____

Differential _____ Absolute_____ Others_____

Pancake_____ Toroid Coil _____ Others_____

Coil Diameter _____ Shielding _____

Scanning Technique_____ Digitization_____

Calibration Level_____ Inspection Threshold_____

Attach a sketch of the inspection setup. Include part orientation with respect to flaw orientation and eddy current direction.

Describe technique for analyzing, rejecting, and recording a defect signal.

FIGURE 1. Eddy current data sheet.

APPENDIX B

FLUORESCENT PENETRANT TESTING SYSTEMS

B.1 SCOPE

B.1.1 Scope.

This appendix provides the detailed requirements and methods for testing evaluation procedures for assessing NDE system capability requirements for fluorescent penetrant testing systems.

B.1.2 Limitations.

Fluorescent penetrant test NDE procedures addressed in this appendix are those used to inspect gas turbine engine components.

B.1.3 Classification.

Fluorescent penetrant test is classified using quantitative measurement, â vs. a data.

B.2 APPLICABLE DOCUMENTS

This section is not applicable to this appendix.

B.3 DETAILED REQUIREMENTS

B.3.1 Demonstration design

B.3.1.1 Test parameters.

The demonstration design for the capability and reliability of the fluorescent penetrant system should include, but not be limited to, the following test variables. These requirements are in addition to those listed in 4.3.

a. Inspector Changes
b. Sensor Changes
c. Loading and Unloading of Specimens
d. Specimen Position
e. Calibration Repetition
f. Calibration Standard Variation, if applicable
g. Test Repetition

B.3.1.2 Fixed process parameters.

Fixed process parameters should include, but not be limited. to the following. Some of these parameters might be included in the matrix of test variables.

a. Penetrating fluid formulation
b. Penetrating fluid application method
c. Dwell times
d. Emulsifier formulation
e. Emulsifier/remover application method, concentration and contact time
f. Developer formulation
g. Developer application method

h. Drying time and temperature

i. Pre- and post-rinse temperature and time

j. Hardware and software configuration control number

B.3.2 Specimen fabrication and maintenance.

The specimens for evaluation of PT systems should contain Low Cycle Fatigue (LCF) surface connected cracks. The cracks should be generated and measured as described in 4.3.2. Because PT indications are more dependent on crack length than area, these cracks should be described by their surface length.

a. The specimens should have the cracks oriented and positioned randomly relative to the edges of the specimens, to minimize the tendency of a manual inspector to "learn the specimens." The inspectors should not know in advance if a particular specimen is cracked, or if it is, they should not know the location, orientation, or size of the crack.

b. Particularly for manual readers, it is important that a significant portion of the samples be crack-free, to help assess the false call rate that will be associated with a particular inspection capability.

c. Specimen maintenance is an issue for PT specimens, since inspection materials are being introduced into the cracks themselves. It is important that the specimens be thoroughly cleaned after each inspection. This cleaning should use an ultrasonic bath of heated acetone to assure that the penetrants are removed from the cracks.

d. Care may also be taken to assure that the chemicals in the inspection materials are not harmful to the specimens. The presence of such elements as sulfur is potentially harmful to some superalloys and may be avoided. All inspection materials and cleaning procedures should be carefully documented as a part of the test plan.

B.3.3 Testing procedures

B.3.3.1 Test definition.

Procedures should be written prior to the test, clearly describing what tests are to be conducted, and the exact procedures for conducting them. They should be to the same level of detail as the day-to-day procedures to which production inspectors operate. In addition to those items outlined in B.3.1, other items to be specified in this test definition are the following:

a. To assure specimen integrity, the specimens should be subject only to cleaning using chemicals that will not degrade the specimen surface or crack characteristics. An ultrasonic cleaning may be necessary to assure that all penetrant material has been removed from the cracks.

b. The definition of the system to be evaluated is critical at this point, to determine the controls being applied to the part processing. If the system being evaluated is a penetrant preprocessor (i.e., applies the penetrant, perhaps the emulsifier and developer) the test is to determine the effect of that system on the inspection results, so the system may be considered to include the reader. Similarly, if the test is to evaluate new penetrant chemicals, the system definition may also include the reader. If the component being evaluated is the reader (e.g., an automatic reader, as opposed to manual), the system may be defined more restrictively, and include only the reader. This assumes that it will be put in production without any changes to the existing preprocessing procedures. In this case, the evaluation should be conducted with no special controls applied to the pre-processing, and with production inspectors following their usual procedures. If it is intended to tighten control of production

preprocessing procedures, it will be necessary to consider the system being evaluated as including all of the pre-processing activities as well as the reader itself.

c. System inspector requirements should typically refer to certification and training requirements, but should also include the number of inspectors to be included in the test plans. Because of the larger scatter historically seen in PT, this is an important criterion. For automated PT readers, it may be practical to reduce the number of inspectors as detailed in 4.2.

d. Inspection materials used should be a significant factor in the evaluation of PT systems, and as such may be specified in the test plan. In many cases the materials (penetrants, emulsifiers, and developers) will be the subject of the evaluations. The chemicals used, their concentrations, and application will need to be detailed in the test procedure. The criteria used for the acceptance of the chemicals (e.g., viscosity, concentrations, etc.) may be those that are planned for production use.

e. The sensor in PT inspections should be considered to include the light source as well as the detector. The detector may be the person inspecting the specimens, or it may be a camera/computer arrangement. In any case, the sensor should be typical of that to be used in production inspections and should meet all of the calibration requirements specified for that equipment. In the case of the human inspector, that calibration may relate to the level of certification; for the light source, it may be intensity measured at some specified distance from the source; for the camera/computer system it may be tied into a software configuration control procedure and to filter types.

f. Inspection setup/calibration requirements should be the same as those used for production inspections, including the same tolerances and settings as may be appropriate for automated readers.

g. During the evaluation tests, the production inspection process should be duplicated as much as possible. Settings such as the time of penetrant application, dwell time, and rinse time all should follow production procedures. The methods of application (for example, dip, spray, or electrostatic spray) should match that planned for production. Scanning procedures should be described, including parameters such as distances of the light source and of the detector from the part/specimen. For the automated readers, the software version and revision numbers should be detailed. Because the cracked specimens are not the same as real hardware to be inspected in production, the scanning motions for specimens may not be the same as those for real components. Efforts should be made to minimize the differences, and recognized differences should be documented. Because the specimens will not provide the same line-of-sight or contour following difficulties as some of the actual production components, it is important that the evaluation plans include some real components with fluorescent markings.

h. Inspection thresholds used in the test should be the same as those planned for production use. With automated readers, this may be set in the signal processing software, and as long as the signal processing software is kept constant, the thresholds will be the same. For the manual reader, the scanning procedure in the test should reflect production procedures as closely as possible (e.g.. if an inspector would normally scan at a rate of 10 square inches per second without magnification, then during the tests he should not focus for prolonged periods on a 6 square inch specimen, or use a magnifier). If the manual reader sees fluorescent indications that he does not call out as cracks in the specimen, he should be prepared to explain why he did not call them out. This will be to minimize the effect of the inspector's learning the specimens.

B.3.3.2 Test environment.

The environment in which the test is run should match the anticipated production environment as closely as possible and be conducted at the production site if possible. If the system is a new development, the initial tests may need to be conducted at the manufacturer's facility. To the extent able, production conditions should be met. It is suggested that the manufacturer conduct a first evaluation prior to shipping the equipment and a second test one or two months after the system is installed on site.

B.3.4 Presentation of results.

Documentation of test results should include all raw data from the tests. If some of the data is classed as irrelevant and not included in the data reduction process, this should be noted, and an explanation given for why this decision was made. This provides the customer the option of accepting or rejecting that rationale.

B.3.4.1 Submission of data.

Data for the permanent record of fluorescent penetrant testing reliability experiments should be submitted in accordance with the requirements stated in 4.6. Figure 2 presents an example of the type of information required for description of penetrant testing systems. The PT inspection results are recorded in the hit/miss format for manual inspections, and should be in the â vs. a format for automated readers. The data are analyzed accordingly (see G.3.2 and G.3.3).

Date:_____

Operator ID:_____

Part Name _____ Part Number_____ Serial Number_____

Alloy _____ Engine _____

Penetrant System Model_____Manufacturer & Date _____

Attach Specification Sheet _____

Inspection Setup - Describe procedure including:

a. Precleaning method
b. Penetrant manufacturer & type. State contact angle.
c. Removal method - State water conditioning and sulfur and halogen content.
d. Drying temperature and time
e. Developer application and time. State manufacturer.
f. Inspection method
g. Post-cleaning method

Defect Evaluation - State technique for analyzing, rejecting, and recording a defect indication.

FIGURE 2. Liquid penetrant test data sheet.

APPENDIX C

ULTRASONIC TESTING SYSTEMS (UT)

C.1 SCOPE

C.1.1 Scope.

This appendix provides the detailed requirements and methods for testing evaluation procedures for assessing NDE system capability requirements for ultrasonic testing (UT) systems .

C.1.2 Limitations.

Ultrasonic test NDE procedures addressed in this appendix are those used to inspect gas turbine engine components.

C.1.2 Classification.

Ultrasonic test is classified using quantitative or qualitative measurement.

C.2 APPLICABLE DOCUMENTS

This section is not applicable to this appendix.

C.3 DETAILED REQUIREMENTS

C.3.1 Demonstration design

C.3.1.1 Test parameters.

The demonstration design for the capability and reliability study of the ultrasonic testing system should include, but not be limited to, the following test variables. These requirements are in addition to those listed in 4.3.

a. Inspector Changes
b. Sensor Changes
c. Loading and unloading of specimens
d. Calibration Repetition
e. Inspection Repetition

C.3.1.2 Fixed process parameters.

Fixed process parameters should mirror actual production inspections and should include, but not be limited to, the following. Some of these parameters might be included in the matrix of test variables.

a. Test frequency (instrument and transducer)
b. Pulser settings, damping, gain, frequency
c. Receiver settings, gain, frequency
d. Transducer size and type
e. Calibration standards (material, artificial defect size, metal travel)
f. Water path

g. Digitization rate and resolution, if applicable
h. Time Compensated Gain (TCG) setup
i. Gate parameters
j. Scanning Technique
    1) Scanning speed
    2) Index value
k. Incident angle of ultrasound
l. Threshold setting
m. Wave mode (shear, longitudinal, surface, Lamb, etc.)

C.3.2 Specimen fabrication and maintenance.

Ultrasonic inspection should use one or more of several inspection modes; including longitudinal, shear, or surface wave. These will require different test specimens, the specifics of which will depend upon the inspection requirements. Typically, the surface wave inspections may use the same specimens as are used for ET (A.3.2) with LCF surface connected cracks. The size characterizations of the specimens used for ET may also be used for UT surface wave. The use of surface wave UT assumes that the orientation of the cracks is known, so the specimens may have the orientation of the cracks defined (although the inspectors should not know if a particular specimen is cracked, or the location or sizes of the cracks).

C.3.2.1 Longitudinal and shear wave UT inspections.

Longitudinal and shear wave UT inspections would typically be evaluated using flat-bottom holes (FBH) at various depths from the entry surface of the specimen. The capability is then quoted in terms of the detectability of the various sizes of FBH at the different depths. Since the surface condition of the specimen can significantly affect this detectability, the specimen surface condition should simulate that of the parts to be inspected. If this surface condition is not known, the specimens should be made with a very good surface finish, and inspection of the typical production part specimens should be used to evaluate the expected noise. These holes should be drilled normal to the direction of sound propagation for the wave mode being evaluated. Hole sizes should be established by replication of the diameter and depth. Since material type and processing history critically affect the inspection capability, efforts should be made to assure that the material is typical of that anticipated for the production components.

C.3.2.2 Defects in diffusion bonded specimens.

Another specimen type that can be used contains internal defects in diffusion bonded specimens as described in F.3.2.3.1. These defects can be used to simulate mal-oriented defects, such as might arise from internal crack growth. Specimens should be made with the defects widely spaced, to avoid inspecting the entire specimen in an artificially severe evaluation mode. Placement of the defects near geometric discontinuities should be done only if that is specifically what is being evaluated. Care should be taken that the defects are not so close together that their UT signals interact. Flaws at greater depths require greater

separation than those closer to the surface. The proximity of the defects that is allowed is a function of the depth of the defect from the entry surface, as the deeper the defect, the greater the sound beam will spread before it reaches the defect.

C.3.2.3 Specimen maintenance.

Specimen maintenance should require no specific precautions, with the only exception being the need to assure that the couplant will not degrade the specimen material.

C.3.3. Testing procedures

C.3.3.1 Test definition.

Procedures should be written prior to the test, clearly describing what tests are to be conducted, and the exact procedures for conducting them. They should be to the same level of detail as the day-to-day procedures to which production inspectors operate. In addition to those items outlined in C.3.1, other items to be specified in this test definition are the following:

a. Part pre-processing requirements should be limited to cleaning the specimens and to the application of the couplant as appropriate.

b. System inspector requirements will frequently refer to qualification and training requirements, but will also include the number of inspectors to be included in the test plan. At the start of the test matrix, this may typically call for three inspectors to be involved in the system evaluations. This number may be reduced as detailed in 4.2.

c. Inspection materials (for example, couplant) are not significant variables.

d. The test plan should require the evaluation of the system using at least two samples of each distinct transducer planned for production use (including factors such as focal length and frequency). The probe body and the use of such things as reflectors need to be factors in this evaluation only to the extent necessary to allow inspection of the specific specimen designs.

e. Inspection setup/calibration may be conducted using the same procedures and calibration standards planned for use in production. The signal responses may be set to the same values, with the same tolerances in both situations. The production inspection process may be duplicated in the test as much as possible. Thus the inspection feed rates, scan index rates, drive signal frequencies, filter settings, water path distances, and any signal processing may be the same. Because the specimens are not the same as real components to be inspected in production, the scanning motions for the specimens may not be the same as those used for components. Efforts should be made to minimize the differences, and recognized differences should be documented.

f. Inspection thresholds used in the test should be the same as those planned for production use. Inspection of the actual fatigue cracked hardware described in 4.3.2.4 will help to establish how realistic those thresholds are for production inspections. Where the specific application of the system is known, typical production components should be used to determine practical thresholds. It may be desirable to inspect the specimens at as low a threshold as possible, to establish the detection capabilities as a function of thresholds used. This will allow trade-offs to be made between detection capability and production throughout.

C.3.3.2 Test environment.

The environment in which the test is run should match the anticipated production environment as closely as possible and be conducted at the production site if possible. If the system is a new development, the initial tests may need to be conducted at the manufacturer's facility. To the extent possible, production conditions should be met. It is suggested that the manufacturer conduct a first evaluation prior to shipping the equipment and a second test one or two months after the system is installed on site.

C.3.4. Presentation of results.

Documentation of test results should include all raw data from the tests. If some of the data is classed as irrelevant and not included in the data reduction process, this should be noted, and an explanation given for why this decision was made. This provides the customer the option of accepting or not accepting that rationale.

C.3.4.1 Submission of data.

Data for the permanent record of ultrasonic testing reliability experiments will be submitted in accordance of the requirements stated in 4.6. Figure 3 presents an example of the type of information required for description of ultrasonic testing systems. The UT inspection results should be recorded in the â vs. a format whenever possible. However, when the inspection mode does not quantify the flaw area (for example, shear wave detecting a corner of a crack) then the hit/miss format is necessary. The data are analyzed accordingly (see G.3.2 and G.3.3).

APPENDIX C

Date:_____
Operator ID:_____

Part Number_____ Serial Number_____ Alloy_____
Engine _____ Part Number_____ Surface Roughness_____

Equipment Model_____ Manufacturer & Date _____
Attach Specification Sheet System Operating Ambient Temperature_____
Other System Operating Environmental Constraints_____

Pulser -
Frequency _____ Voltage_____ Damping_____
Receiver _____ Rise Time_____ Pulse Width_____
Frequency _____ Gain_____ Filtering_____

Monitor Gate -
Delay _____ Width_____ Level_____
Time Compensate Gain
Attach Graph - Gain versus Time

Transducer -
Manufacturer_____ Date_____ Shelf Life_____
*Frequency _____ Piezo Electric Disk Material _____ Disk Diameter _____

This is the frequency of the finished transducer measured with a frequency analyzer.

Type -
Contact _____ Angled _____
Couplant _____ Couplant _____
_____ Wedge Material_____

Immersion -
Unfocused _____ Focus_____ Focus Distance _____
Operating Water Path _____
Mode of Operation -
Longitudinal _____ Transverse____ Surface_____
Scanning Technique _____ Digitization_____
Calibration Level _____ Inspection Threshold_____

Attach a sketch of the inspection setup. Include part orientation with respect to flaw orientation and ultrasonic beam direction.

FIGURE 3. Ultrasonic test data sheet.

APPENDIX D

MAGNETIC PARTICLE TESTING

D.1 SCOPE

D.1.1 Scope.

This appendix provides the detailed requirements and methods for testing evaluation procedures for assessing NDE system capability requirements for magnetic particle test (MT**)** systems.

D.1.2 Limitations.

Magnetic particle test NDE procedures addressed in this appendix are those used to inspect gas turbine engine components.

D.1.2 Classification.

Magnetic particle testing is classified using quantitative or qualitative measurement.

D. 2 APPLICABLE DOCUMENTS

This section is not applicable to this appendix.

D.3 DETAILED REQUIREMENTS

D.3.1 Demonstration design

D.3.1.1 Test parameters.

The demonstration design for the capability and reliability study of the magnetic particle inspection system should include, but not be limited to, the following test variables. These requirements are in addition to those listed in 4.2.

a. Inspector Changes
b. Sensor Changes
c. Loading and unloading of specimens
d. Calibration Repetition
e. Inspection Repetition

D.3.1.2 Fixed process parameters.

Fixed process parameters should include, but not be limited to, the following. Some of these parameters may be included in the matrix of test variables, if desired.

a. Magnetic suspension formulation and concentration
b. Magnetic current for a particular part number
c. Demagnetizing procedure
d. Method of magnetization (circular or longitudinal)
e. Method (e.g., fluorescent or visible)

D.3.2 Specimen fabrication and maintenance .

The specimens for evaluation of MT systems should contain LCF surface connected cracks. The cracks should be generated and measured as described in 4.3.2. Specimen geometry and material should represent production component.

D.3.2.1 Specimen treatment.

It is important that the specimens be treated carefully to prevent corrosion. They should be thoroughly cleaned after each use. Care may be taken to ensure that the chemicals in the inspection materials do not degrade the specimen material. The presence of some elements, such as sulfur, may be harmful to some alloys, and may be avoided. All inspection materials and cleaning procedures should be carefully documented as a part of the test plan.

D.3.3 Testing procedures

D.3.3.1 Test definition.

Procedures should be written prior to the test, clearly describing what tests are to be conducted, and the exact procedures for conducting them. They should be to the same level of detail as the day-to-day procedures to which production inspectors operate. In addition to those items outlined in D.3.1, other items to be specified in this test definition are the following:

a. To maintain specimen integrity, the specimens should be subject only to cleaning using chemicals that will not degrade the specimen surface or crack characteristics.

b. The definition of the system to be evaluated is critical to a determination of the controls to be applied to the part processing. If the system being evaluated is a preprocessor (i.e., applies the current and the particle material to the component) the test is to determine the effect of that system on the inspection results, so the system may be considered to include the reader. Similarly, if the test is to evaluate new particle materials, the system definition may include the reader. If the component being evaluated is the reader (e.g., an automated reader, as opposed to manual), the system definition may be defined more restrictively, and include only the reader. This assumes that it will be put into production without any changes to the existing pre-processing procedures. In this case, the evaluation should be conducted with no special controls applied to the preprocessing, and with production inspectors following their usual procedures. If it is intended to tighten control of production pre-processing procedures, it will be necessary to consider the system being evaluated as including all of the preprocessing activities as well as the reader itself.

c. Inspector requirements refer to certification and requirements, and will include the number of inspectors to be included in the test plans. Because of the scatter historically associated with what has historically been a very operator-dependent inspection, this is an important criterion. For automated readers, it may be practical to reduce the number of inspectors as detailed in 4.2.

d. Inspection materials used should be a significant factor in the evaluation of MT systems and as such may be specified in the test plan. In many cases the materials themselves will be the subject of the evaluations. The chemicals used, their concentrations, agitation, and their application will need to be detailed in the test procedure. The criteria used for the acceptance of these materials may be those that are planned for production use.

e. The sensor in MT inspections should be considered to include the light source as well as the detector. The detector may be the person inspecting the specimens, or it may be a camera/computer arrangement. In any case, the sensor should be typical of that to be used in production inspections, and should meet all of the calibration requirements specified for that equipment. In the case of the human inspector, that calibration may be related to the level of certification; for the lightsource, it may be intensity measured at some specified distance from

the source; for the camera/computer system it may be tied into a software configuration control procedure and filter types.

f. Inspection setup/calibration requirements may be the same as those used for production inspections, including the same tolerances and settings as may be appropriate for automated readers.

g. During the evaluation test, the production inspection process may be duplicated as much as possible. Settings such as the current, direction of current flow, particle application and agitations, etc., all should follow production procedures. The methods of application also may match that planned for production. Scanning procedures may be described, including parameters such as distance of the light source and of the detector from the part/specimen. For automated readers, the software version and revision numbers may be detailed. Because the cracked specimens are not the same as real components to be inspected in production, the scanning motions for the specimens may not be the same as those used for the components. Efforts should be made to minimize the differences, and recognized differences should be documented. Because the specimens will not provide the same line-of-sight or contour-following difficulties as some of the actual production components will, it is important that the evaluation plans include some real production components with artificial defects such as EDM notches.

h. Inspection thresholds used in the test should be the same as those planned for production use. With automated readers, this may be set in the signal processing-software, and as long as the signal processing software is kept constant. The thresholds will be the same. For the manual reader, the scanning procedure in the test should reflect production procedures as closely as possible (e.g., if an inspector would normally scan at a rate of 10 square inches per second without magnification, then during the tests he should not focus for prolonged periods on a 6 square inch specimen or use a magnifier). If the manual reader sees fluorescent indications that he does not call out as cracks in the specimen, he should be prepared to explain why he did not call them out. This will minimize the effect of the inspector's "learning the

D.3.3.2 Test environment.

The environment in which the test is run should match the anticipated production environment as closely as possible and be conducted at the production site if possible. If the system is a new development, the initial tests may need to be conducted at the manufacturer's facility. To the extent possible, production conditions should be met. It is suggested that the manufacturer conduct a first evaluation prior to shipping the equipment and a second test one or two months after the system is installed on site.

D.3.4 Presentation of results.

Documentation of test results should include all raw data from the tests. If some of the data is classed as irrelevant and not included in the data reduction process, this may be noted, and an explanation given for why this decision was made. This provides the customer the option of accepting or rejecting that rationale. The MT inspection results are recorded in the hit/miss format for manual inspections, and should be in the â vs. a format for automated readers. The data are analyzed accordingly (see G.3.2 and G.3.3).

APPENDIX E

TEST PROGRAM GUIDELINES

E.1 SCOPE

E.1.1 Scope.

This appendix presents the test program procedures of a NonDestructive Evaluation (NDE) demonstration. The purpose of an NDE demonstration is to produce a POD(a) curve, and lower bound, which accurately represent the capability of an inspection system. This is accomplished by recording the system responses which result from inspecting flaws of known sizes. The mathematical details of producing a POD(a) curve are discussed in Appendix G.

E.1.2 Limitations.

E.1.2 Classification.

Since the system response for ET, UT, PT, or MP is subject to variation in the input variables (eg: probe, inspector, penetrant type), it may be necessary to determine the impact of these variables on the system response. The plan for determining the best estimate of the overall POD(a) curve as well as the significance of the input variables is called an NDE experimental design.

E.2 APPLICABLE DOCUMENTS

This section is not applicable to this appendix.

E.3 EXPERIMENTS

E.3.1 Main effects and interactions.

Main effects are the changes in the NDE system response caused by the input variables acting individually. Main effects are additive. An interaction occurs between two variables if the effect of the two variables is not additive. If there is no interaction, then a pattern observed at a low level of a factor should result in the same pattern at the high level. Pictorially this is shown in figure 4, where inspector 2 produces a higher response than does inspector 1, regardless of which probe is used, and probe 1 is better than probe 2 regardless of inspector.

APPENDIX E

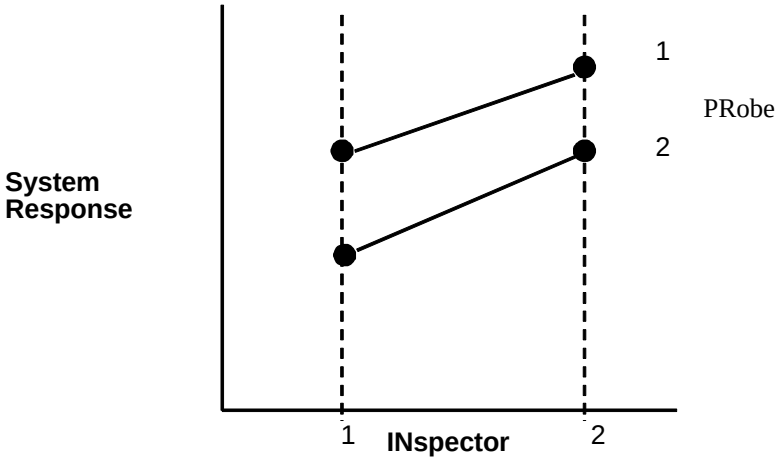


FIGURE 4. Parallel lines indicate No. 2 - factor interaction.

a. If there is interaction, then this pattern doesn't exist. This is illustrated in figure 5. Here inspector 1 using probe 2 produces a higher response, but the situation is reversed when the inspectors change probes. Notice that probe 1 is not uniformly better than probe 2.

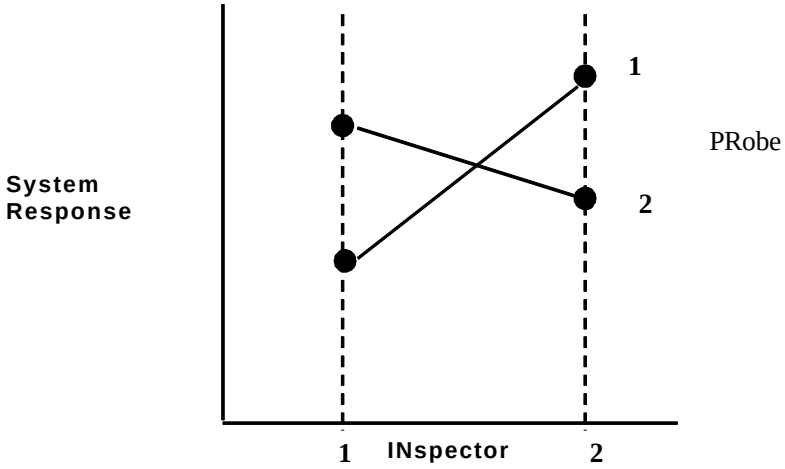


FIGURE 5. Interactions cause the lines to cross.

b. If an interaction is suspected, then the experiment should be designed so that the interaction effects can be separated from the main effects.

E.3.2 Experimental design.

a. Input variables can be divided into two groups: control factors and noise factors. The first group contains variables which are to be tested at different levels. (For ET, significant variables may be inspector, probe, and position; for PT, significant variables may include inspector, penetrant, or emulsifier processing times.) The second group contains those variables which either can be tested, but for some reason are deemed as less important to test, or can't be identified and therefore can't be tested (but can still cause variation in the system). Noise factors may be changes in surface preparation, or influence of laboratory humidity and temperature.

b. The output response can be expressed as:

$$y = f(x_1, \ldots x_p),\ x_{p+1},\ \ldots\ x_{p+r},\ x_{p+r+1}, \ldots)$$

where	$x_1, \ldots x_p$	are controlled in the test
	$x_{p+1}, \ \ldots$	are noise
	$x_{p+1}, \ \ldots\ x_{p+r}$	can be tested but are not
	$x_{p+r+1}, \ldots$	cannot be identified or tested

To quantify the POD(a) relationship for an eddy current system, a typical test program would proceed as follows. First, those knowledgeable of the specific inspection process would decide which variables are important in defining the response. If many variables are identified, a Pareto analysis may help determine which are the more important, and thus separate the significant few variables from the trivial many variables. Once the important variables are determined (say inspector, probe, and position of the specimen for ET), an NDE experiment is designed to determine their effect on the response. A factorial experiment, discussed in E.3.4, is recommended for most cases, although many designs exist and should be used as appropriate.

E.3.3 One-factor-at-a-time experiments.

A one-factor-at-a-time design, as the name implies, considers each factor in isolation. To test for a difference in probe under this plan, two probes would be selected and specimens tested using these probes while inspector and position are held constant. In the past, this has been a common method of experimentation. However, there are more efficient ways to gather the needed information (i.e., fewer tests are required using other methods). There are other problems with the one-factor-at-a-time method. Because the other variables are held unchanged, the observed NDE system responses are valid only for that specific setting of the other variables.

E.3.3.1 Interactive effects.

Therefore, interactive effects among input variables are undetectable. It is also more likely to confuse a correlation of input and response, with cause and effect, using this method of experimentation. Finally, the resulting POD(a) curves are less precise than they could otherwise be, because only one set of measurements is taken to estimate the influence of a specific variable.

APPENDIX E

E.3.4 Factorial experimentation.

A factorial NDE evaluation considers the influence of all factors simultaneously. A full factorial expe0riment is performed by choosing a number of levels for each of a number of factors (variables) and the experiment is conducted for each possible combination of the factors. If there are L1 levels for the first variable, L2 for the second, and Lk for the kth variable, then the experiment is called an L1 x L2 x ... x Lk factorial design. A 2 x 3 x 5 factorial design requires 2 x 3 x 5 = 30 runs. As an example, consider the 3 factors of the ET setup (PRobe, INspector, and POSition) each at 2 levels; this is a 2 x 2 x 2 = 8 run factorial experiment. Figure 6 is a plot of the three independent (input) variables for this example. A (+) indicates one level of either the probe (PR), inspector (IN), or position (POS) variable and a (-) indicates the second level. Notice that the cube represents the input factors only; the system response is not being plotted.

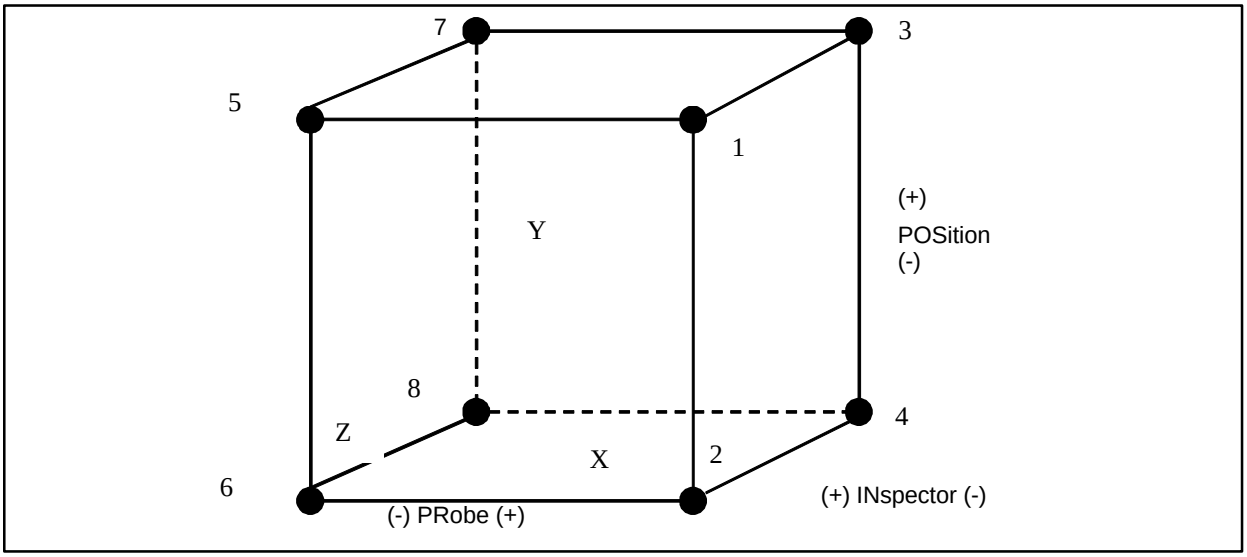


FIGURE 6. A cube representing a full (2x2x2) factorial experiment.

E.3.4.1 Test conditions.

The test conditions represented by this cube are provided in table I. In practice, run numbers are assigned to the tests in a random order. Randomization is required to minimize the effects of those factors which are sources of variation for the response and have not been controlled experimentally, i.e. the noise factors. Errors can result from attempts to save time, labor, or materials by choosing a particular nonrandom run sequence, so careful thought and planning are necessary prior to conducting the NDE system evaluation.

APPENDIX E

TABLE I. Full factorial test conditions for figure 6.

Test Condition	X PR	Y POS	Z IN	PR*POS	PR*IN	POS*IN
1	+	+	+	+	+	+
2	+	-	+	-	+	-
3	+	+	-	+	-	-
4	+	-	-	-	-	+
5	-	+	+	-	-	+
6	-	-	+	+	-	-
7	-	+	-	-	+	-
8	-	-	-	+	+	+

E.3.4.2 Level of a factor.

The number of levels of a factor to include in an experiment is based on several considerations. If the NDE system response is linear, then two levels are sufficient; nonlinear factors require three or more levels. The number of natural levels a variable possesses, or the amount of variation which is expected, can also influence the number of levels to test. Experience suggests that 2 or 3 levels are appropriate for testing variables in an ET, UT, PT, or MT system. (Other types of testing situations may require more than 3 levels or more than 3 variables; this will be discussed shortly.)

E.3.4.3 Factorial designs.

Factorial designs have three major benefits:

a. The design is more efficient, i.e., more information is gained for a given expenditure of labor, time, and material, than with other methods.

b. Comparisons across levels of a factor (e.g.. inspector or probe) are more precise since average values are used rather than single observations. That is, all observations contribute to all comparisons among all factors; no single test exists only to evaluate a single factor. Notice in table I that the average of test conditions 1, 2, 3, 4 compared to the average of test conditions 5, 6, 7, 8 is a comparison of probe 1 results to probe 2 results - each with a sample size of 4. A comparison of 1, 2, 5, 6 vs. 3, 4, 7, 8 can be used to check for a difference between inspectors. Specimen position effects are estimated by comparing 1, 3, 5, 7 vs. 2, 4, 6, 8.

c. Interactions can be estimated. For example, the average response from tests 1, 2, 7, 8 vs. the average resulting from 3, 4, 5, 6 provides an estimate of the magnitude of the interaction of probe and inspector.

E.3.5 Fractional factorial experimentation.

The number of tests required by a full factorial design increases rapidly as the number of factors is increased. Even with a 2 x 2 x 2 x 2 = 2^4 = 16 run factorial design, the labor, time, and material used to complete the design may be more than is available. It turns out,

however, that since the factorial design is efficient and estimates of variables effects are made more precisely than one-factor-at-a-time methods, the results can be achieved by performing only a fraction of the full factorial. However, since fewer NDE settings are evaluated, something is lost. The ability to discern the significance of the main effects (PR, IN, POS) from the effects of some of the interaction terms is traded for the reduced test matrix. For example, in a full factorial experiment, PR may be identified as having a significant effect on the NDE response. In a fractional experiment, the effect of PR may be confused with the effect of the IN*POS interaction, and therefore the significance may be attributed to the probe by itself or to an interaction of probe and position. If this problem occurs, further experimentation can be performed to investigate these interactive effects without having to design a completely new experiment. This is not true of the one-factor-at-a-time approach.

E.3.5.1 Examples.

The example in table II shows how the effects which are confused, or confounded, with one another can be determined by comparing the “signs” in each column; columns with all signs the same are confused. Here the effects of IN and the PR*POS interaction are confused, the effects of PR and the IN*POS interaction are confused, and the effects of POS and the IN*PR interaction are confused.

TABLE II. Fractional factorial test conditions for figure 7 (Columns with all signs the same are confounded).

Test Condition	X PR	Y POS	Z IN	PR*POS	PR*IN	POS*IN
1	+	+	+	+	+	+
4	+	-	-	-	-	+
6	-	-	+	+	-	-
7	-	+	-	-	+	-

a. Using this information, a fractional factorial can be designed by setting the factors of PR, IN, and POS at two levels each. This situation can be represented by the cube in figure 7.

APPENDIX E

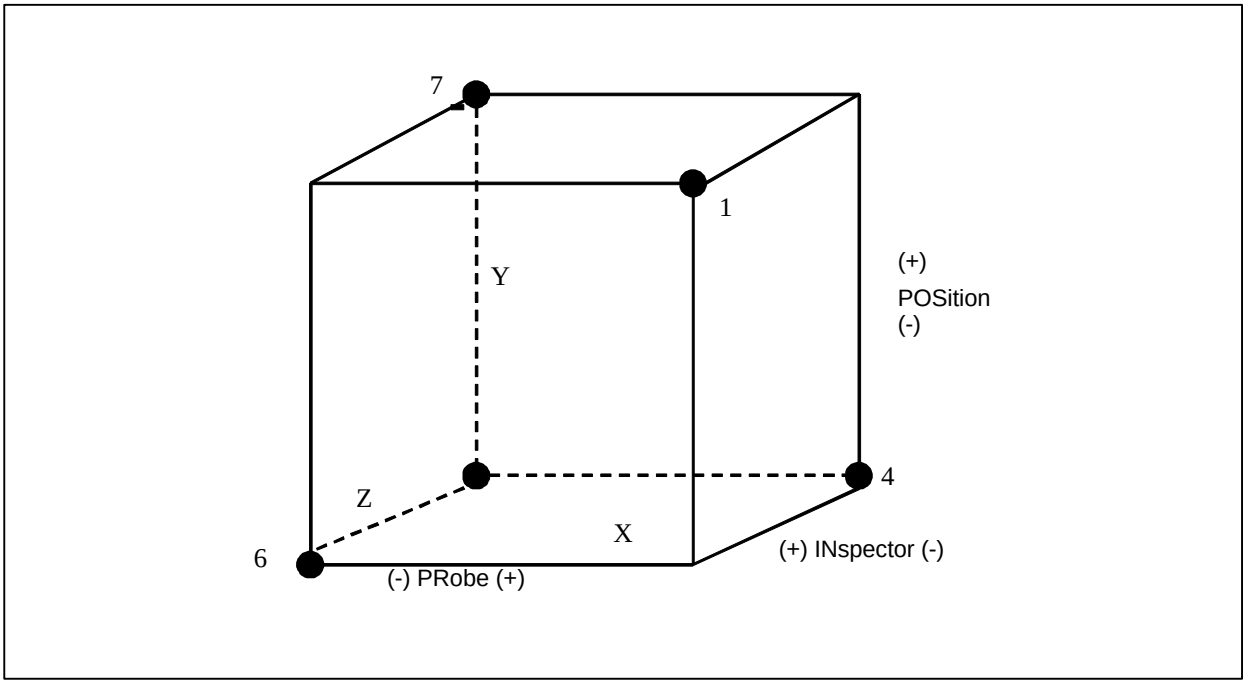


FIGURE 7. A cube representing a fractional factorial experiment.

b. Four tests under conditions 1, 4, 6, 7 of the full factorial matrix in table I would be made; these points are found in table II. The comparison between the probe levels would be made by comparing the average of the response from one level of probe (PR+) to the average response with the other level of probe (PR-). Notice that this same (fractional) data will also allow for a similar test between high and low levels of both inspector and position. Many commercially available software packages can perform these calculations. The analysis of NDE experiments is discussed in detail in Appendix H.

c. If the resulting difference in the response is significantly different from zero, then a change from one probe to another will have an influence on the NDE response. This would indicate that reducing the amount of variation in the POD(a) curve would require more consistent probes.

d. Some fractions of the full factorial experiment are better than others. A poorly designed fractional factorial experiment is illustrated in table III which shows a subset of the full factorial design shown in table I. Since the (+) and (-) signs are the same in the PR and IN columns, this test confuses the PR and IN variables with each other. Conclusions about PR would be the same as conclusions about IN since all levels are the same for each test condition. Due to the confused main effects of PR and IN , it is inconceivable that this test program would ever be run. To avoid this problem with confused variables, an experimenter may know before the test is conducted which variables and interactions are important or significant and design the test taking this into consideration.

APPENDIX E

TABLE III. An improper fractional factorial experiment confuses the main effects (Columns with all signs the same are confounded).

Test Condition	X PR	Y POS	Z IN	PR*POS	PR*IN	POS*IN
1	+	+	+	+	+	+
2	+	-	+	-	+	-
7	-	+	-	-	+	-
8	-	-	-	+	+	+

e. It may be necessary to extend the testing to more than three variables or more than three levels of the variables. A factorial or fractional factorial design, or one of several other classes of designs, can be created to test these situations. It is recommended that someone knowledgeable in statistical experimentation, most likely a professional statistician, assist in the NDE demonstration. Box, Hunter, and Hunter, *Statistics for Experimenters*, Wiley, 1978, provides an excellent discussion of the design and analysis of industrial experiments.

E.3.6 Experimentation by sampling.

An alternative NDE evaluation design may be purposely to confuse all effects of all variables with each other and with experimental error. That is, the output response can be expressed as:

$$Y = f(x_1, \ldots x_p, x_{p+1}, \ldots x_{p+r}, x_{p+r+1}, \ldots)$$

where $x_1, \ldots$ are noise

$x_1, \ldots x_{p+r}$ can be tested but are not

$x_{p+r+1}, \ldots$ cannot be identified or tested

a. To estimate the POD(a) relationship and the corresponding lower bound in a situation when the system has been demonstrated to be in statistical control, or for periodic reevaluation of NDE capability, a sampling approach may be appropriate. Here the overall system performance is to be quantified, as well as some measure of the variability which can be expected.

b. For example, consider a PT process with 20 inspectors, and a specified range of acceptable values for penetrant dwell time, emulsifier concentration, and emulsifier dwell time. Suppose also that the range for emulsifier concentration can be reasonably represented by its two end points, but the ranges of dwell times are large enough to require

a midpoint representation to augment the end point values. A full factorial evaluation would require 360 observations:

20 inspectors x 3 penetrant dwell times x 2 emulsifier concentrations x 3 emulsifier dwell times

c. To proceed with the sampling approach, a full factorial of these 360 observations would be tabulated. Next, a sample size, say 15 test runs, would be determined and a representative random sample of that size tested from the 360 possible observations. In this instance, randomly select 15 tests from the 360 possible. These tests would be performed in this randomly selected order. The resulting POD(a) would reflect error from all the combined influences. If a large variation were to be observed, as indicated by the POD(a) confidence limit, the source(s) would be indistinguishable from the noise. That is, there would be no way to associate a deviation with its cause.

APPENDIX F

FABRICATION, DOCUMENTATION & MAINTENANCE OF RELIABILITY ASSESSMENT SPECIMENS

F.1 SCOPE

F.1.1 Scope.

This appendix presents general guidance for manufacturing NDE reliability specimens for use when no existing specimen sets can provide an adequate evaluation of the NDE process under evaluation. Also included are general guidelines for maintaining the specimens between inspections.

F.1.2 Limitations.

F.1.3 Classification.

F.2 APPLICABLE DOCUMENTS

This section is not applicable to this appendix.

F.3 REQUIREMENTS

F.3.1 Design.

Specimen geometry should be similar to that of the parts being inspected. Holes should be typical of the sizes in typical engines. Specimens representative of particular part geometries should be used when that information is known and when there is reason to expect that the inspection will be geometry dependent. Specimen size should be such that inspection of the specimens is reasonably similar to the inspection of actual parts. Small specimens may require scanning motions completely divorced from those used in production. This should be avoided to the extent practical. Some system evaluation data may need to come from inspection of actual engine hardware. This is particularly true of systems dependent on line-of-sight inspection, such as for PT. The procuring agency will define a selection of preferably field cracked engine hardware for this system evaluation.

F.3.1.1 Machining tolerances.

Machining tolerances for the specimens should be similar to those for the engine hardware to be inspected. Specimens should be manufactured to cover the range of sizes allowed, e.g., if a typical hole has an allowable diameter range of 0.015 inch (including MRB and potential rework), the specimens used for inspection system evaluation should span at least that range. This may not be a significant concern for some features for particular inspection methods, for example, hole size tolerances may not be an issue for PT inspections.

F.3.1.2 Environmental conditioning.

Environmental conditioning, to represent such conditions as in-service oxidation, should be included in the specimen fabrication if they can be realistically simulated. This simulation should be demonstrated first on a small sample of specimens to verify its validity.

F.3.2 Fabrication

F.3.2.1 Processing of raw material.

To the extent that the specific applications of the NDE system are known, it may be possible to specify the raw material processing of the test specimens. Issues to be considered should include processing techniques e.g., (forging (isothermal, upset, flow patterns,... powder metal (mesh size, HIP)), casting, extruding, ...). Heat treatment of the specimens should reflect that seen by the parts, as should the machining processes (turning, grinding, broach, EDM, etc.). If the applications are not known precisely, specimens representative of production parts currently receiving similar inspections should be selected.

F.3.2.2 Establish machining parameters.

Machining parameters have to be established for each desired specimen geometry to simulate the component fabrication conditions. As an example, for a specimen with a crack located at the intersection of a cooling hole with a countersink as might be present in a turbine disk, the following details are presented. Figure 8 illustrates the component geometry. Figures 9 and 10 give the crack geometry relationship obtained from the destructive evaluation. Figure 11 shows how a given final crack can be plotted graphically for a given initial crack that has an 0.280 inch diameter hole drilled at a 25 degree angle to the surface with a 38 degree countersink. The machining of this specimen was accomplished on a Knight vertical milling machine. The specimen was held on an angled fixture which established the hole center line angle (25 degrees) and center line position (0.096 inches from the crack center). A drill guide was placed on top of the specimen and cobalt drills and reamers were used to generate the hole. Generation of the countersink machining parameters were done by trial and error with dummy holes until the proper depth and location was established, and then the countersink was machined in the specimen with the specimen held horizontal in the milling machine.

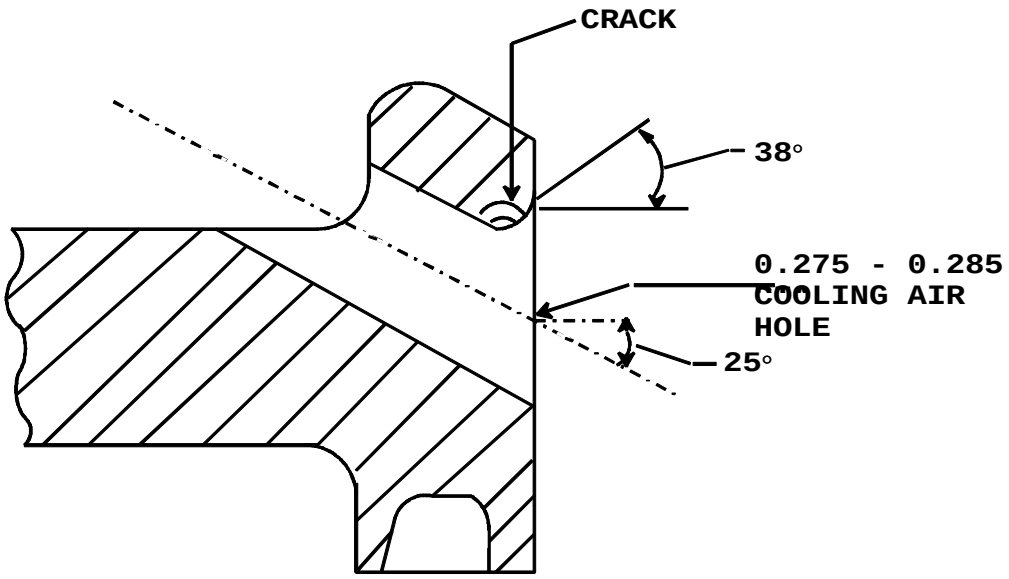


FIGURE 8. First turbine disk.

F.3.2.2.1 Machining parameters.

Because the final machining of the specimens has a direct effect on surface crack size, shape, and aspect ratio, and on internal defect location, it is important that the specimen blank be machined to the same tight tolerances as the final specimen will be. Since several thousandths (0.001 inch) of an inch of material will be subsequently machined off, the processing of the blank is critical only to the degree that the machining will produce cold working or some heat treatment to the depth of the finished specimen surface. For this reason, the machining parameters should specify such things as depth of cut, and these parameters should be held constant over the population of the specimens, and documented for future reference.

F.3.2.3 Defect insertion.

Simulated machining defects are inserted into the finish machined specimen. Surface cracks should be grown from EDM notches or tack welds. If the relation of specimen scanning and crack orientation is known, this should be accounted for in the crack generation. If this relation is not known, the crack orientation should be random, relative to the edges of the specimen. The machining of the EDM notch should be closely defined and documented to assure repeatable notches, in terms of the notch dimensions and also in the amount of recast layer and heat-affected zone. Cracks should be grown from these EDM notches by stress cycling the specimen at a stress sufficiently high to grow with no measurable plastic deformation. Cyclic lives (to the desired crack lengths) should be between approximately 10,000 and 50,000 cycles. Cyclic loads or strains should be well documented to assure consistent application over the specimen population. Depending upon specimen geometry, the cracks can be induced by a tensile load (applied uniformly over the cross-section of the specimen) or three-point or four-point bending. Environmental conditions under which service-induced cracking would be introduced will be simulated to the extent reasonable. This simulation should be tried first on a small sample of specimens to establish its realism.

F.3.2.3.1 Internal defects.

Internal defects can be generated by milling shallow (< 0.003 inch deep) holes into the face of a block to be diffusion bonded to a mating block. Because of the requirements of the diffusion bonding process, the mating surfaces may be very carefully machined. This will also facilitate the necessary flaw location and machining parameter documentation.

F.3.2.3.2 Flaw documentation.

Flaw documentation may include critical parameters, such as flaw depth, length, width, and bottom radius. For examples, see figures 12 through 15. All of the defects should be documented, including the position and orientation. For internal defects, size and shape of the defect should be recorded. For surface cracks, the size and shape of the starter notches should be kept, and also the stress cycling imposed to generate the cracks, including the loads and number of cycles.

F.3.2.4 Final machining.

Specimens will require final machining to remove misalignment of bonded surfaces, provide finished contour, and remove starter notches. Especially for the last function, it is critical that tight dimensional tolerances be maintained. The amount of material removed can have a significant effect on the final shape and size of the defect. A magnified visual inspection may be conducted to verify complete removal of the starter notch. Some of each population will need to be fractured for the specimen verification described in F.3.2.5.

F.3.2.4.1 Final machining procedures.

Final machining procedures for the specimens may be carefully followed and documented. The specimens used for system evaluation should be machined to the same parameters as the parts to be inspected. Where specific applications are not known, or where the specimens cannot be machined in this manner, specimens with surface conditions typical of the types of parts to be inspected should be used. Surface condition refers to such factors as finish and texture and to the presence or absence of machining or handling marks or damage.

F.3.2.5 Defect verification.

Both the aspect ratio and length of the fatigue cracks should be verified. Specimen dimensional information should be recorded. This data may concentrate on the characterization of the flaws as regards the position, orientation, and size. For surface connected cracks, measured lengths (and depths for hole specimens) should be recorded for all cracks. This measurement is best accomplished by magnified (~ 40 x) optical measurement with the specimen under ~ 60 % of the load used during the crack growth cycling. The aspect ratio should be verified by breaking open a sufficient number of specimens as defined in the CDRL prior to final machining. To break open a crack, cut to within 0.050 inches of each end of the crack with a saw or cut off wheel, then fracture the specimen with a single load application. Establish the crack contour to surface length relationship. Failure to meet the estimated aspect ratio within the limits specified by the Statement of Work (SOW) or failure to repeatedly reproduce an aspect ratio within the specified limits will require modification of the crack generation procedure until this requirement is met. Once the desired aspect ratio can be demonstrated, all fatigue crack lengths should be measured to within 0.002 inches in the final machined configuration.

F.3.2.5.1 Specimen flaw response.

Specimen flaw response should be documented for all specimens using a standard test technique that is available to WL/MLSA or other procuring agency that will be the specimen custodians. For systems for which the magnitude of signal response, â, will be used in determining the POD(a) relationship, the flaw response should be recorded at least six times to provide an estimate of test-to-test scatter. Specimen reverification will involve comparison of the results of periodic repetition of this test with these original results.

F.3.2.5.2 Imbedded defects.

The size and shape of the imbedded defects produced by diffusion bonding shall be verified by sectioning, as required by the CDRL or SOW. The size and shape of other types of imbedded defects shall be verified as specified by the contracting agency.

APPENDIX F

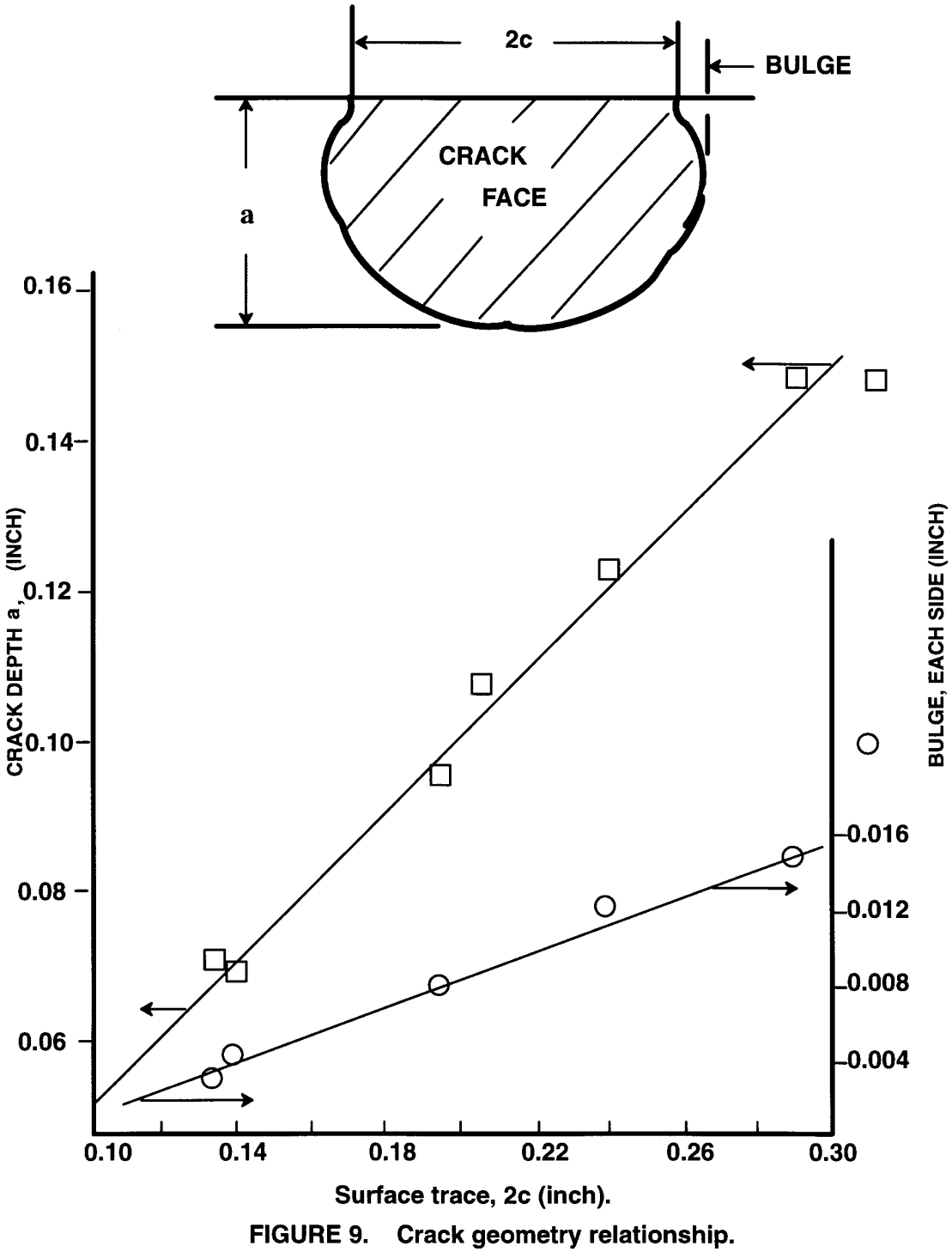


FIGURE 9. Crack geometry relationship.

APPENDIX F

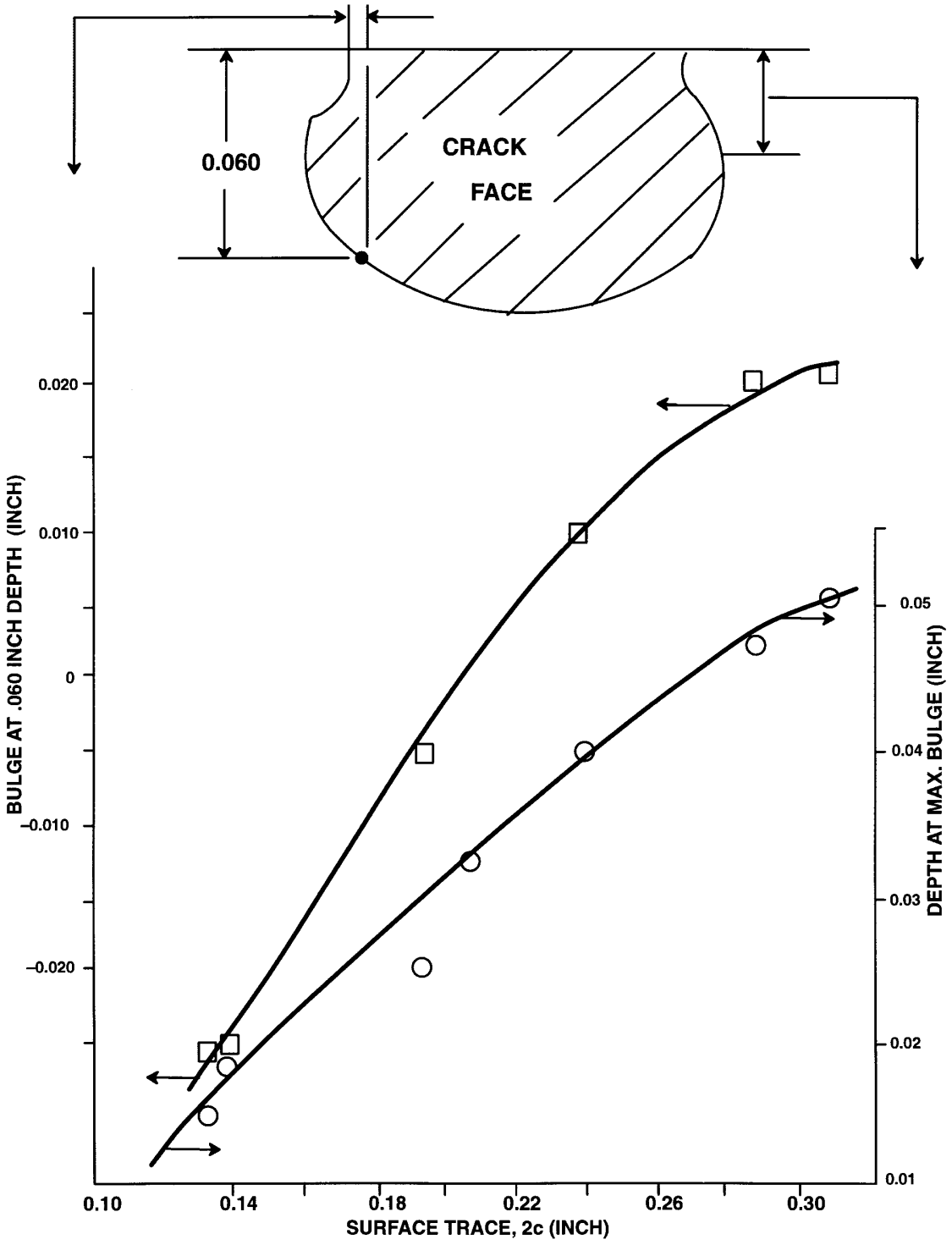
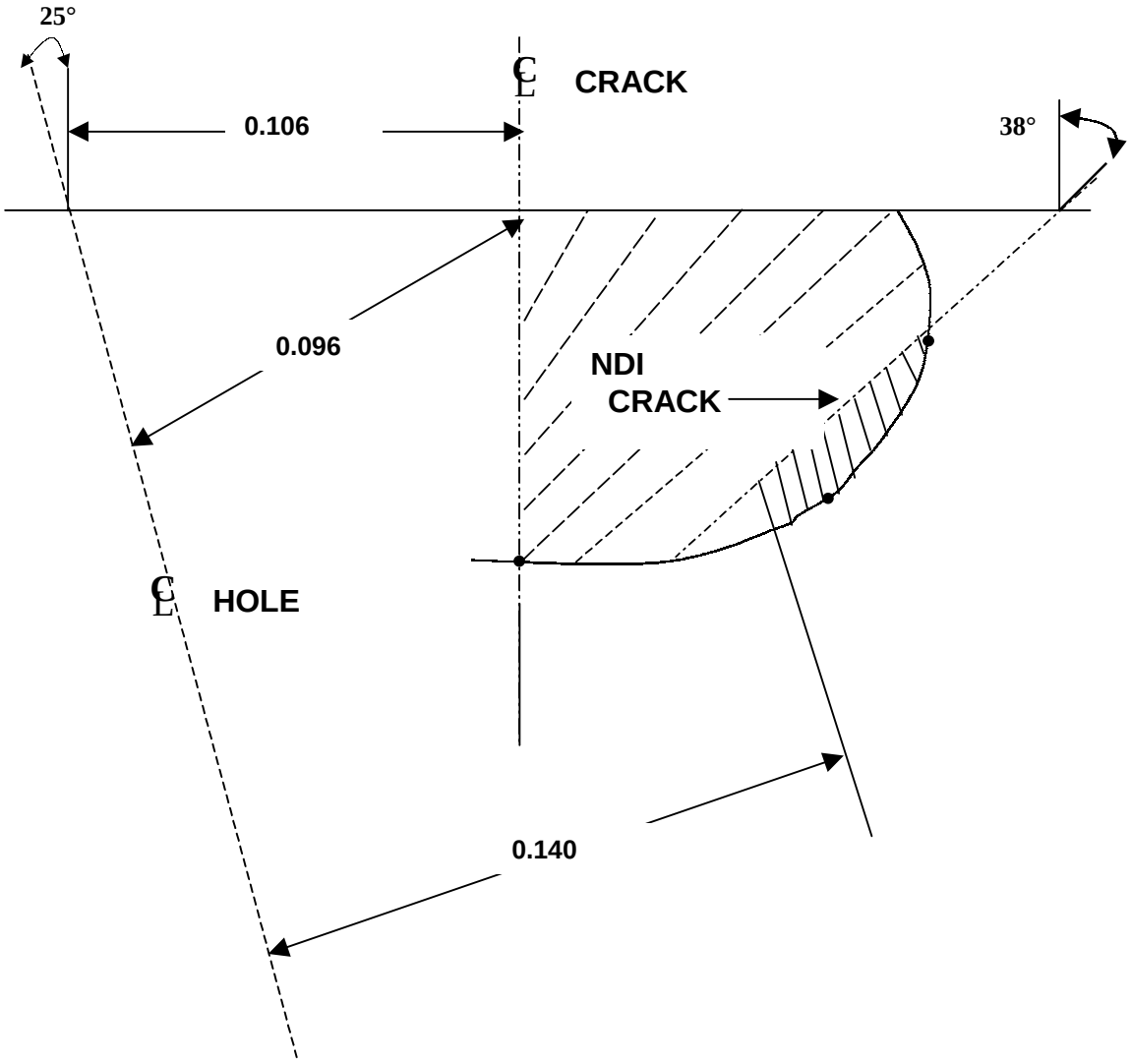


FIGURE 10. Crack geometry relationship at 0.060 depth.

APPENDIX F



2c	= 0.192 INCH
a	= 0.099 INCH
BULGE MAX	= 0.008 INCH AT 0.030 INCH DEPTH
Λ AT .060 INCH	= -0.003 INCH

FIGURE 11. Final crack manufacture.

APPENDIX F

Specimen FML #_____
Matrix/Test _____
Facility _____
FPI System _____

Operator _____
Date _____
Inventory No._____

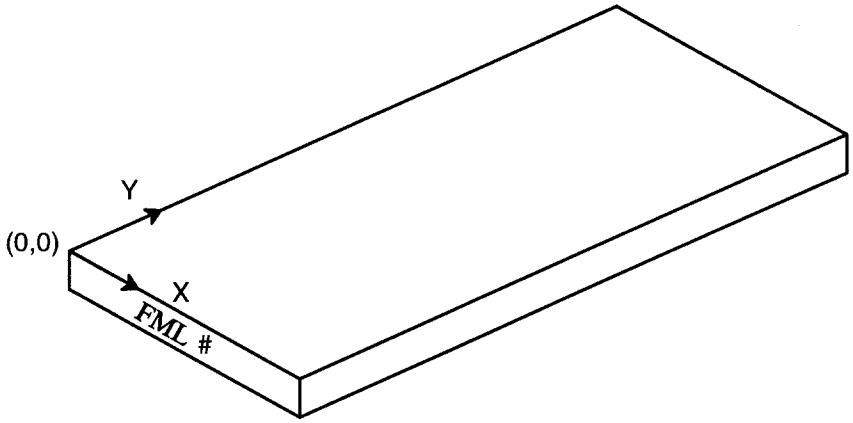
	LOCATION*			
S/N	UP	DOWN	X	Y

	LOCATION*			
S/N	UP	DOWN	X	Y

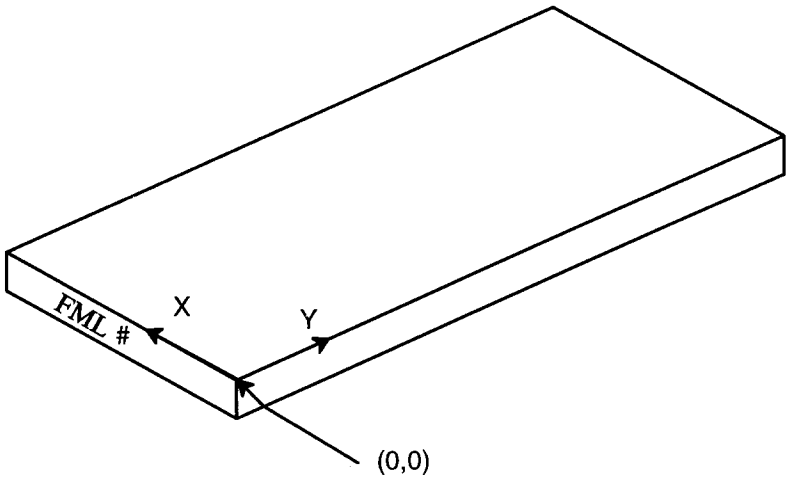
* See Figure 13 to reference flaw location with S/N.

FIGURE 12. F100–PW–229 ENSIP manufacturing inspection reliability test web/bore surface specimen FPI test data sheet.

APPENDIX F



X – Y REFERENCE WITH SPECIMEN IN UP POSITION



X – Y REFERENCE WITH SPECIMEN IN DOWN POSITION

FIGURE 13. Flaw location reference.

APPENDIX F

Specimen FML #_____
Matrix/Test _____
Facility _____
FPI System _____

Operator _____
Date _____
Inventory No._____

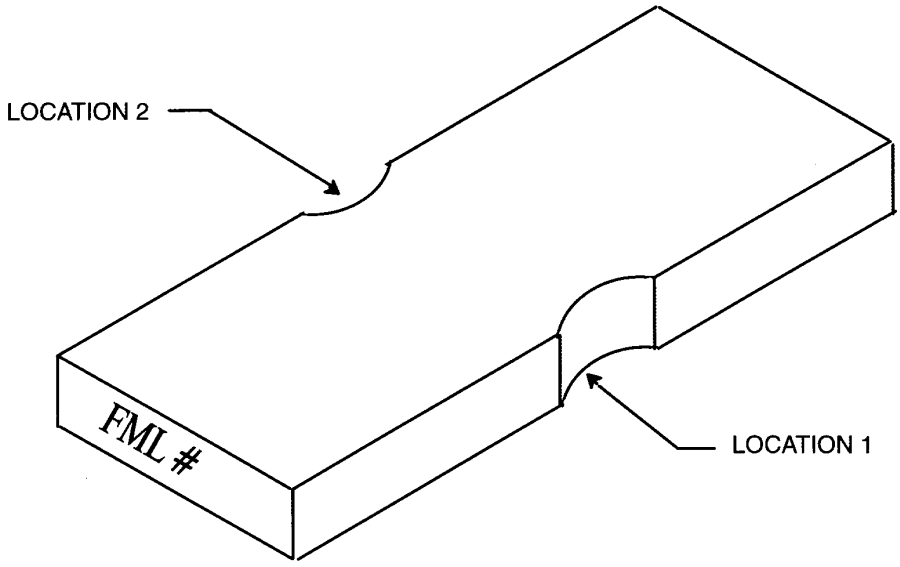
	LOCATION*	
S/N	1	2

	LOCATION*	
S/N	1	2

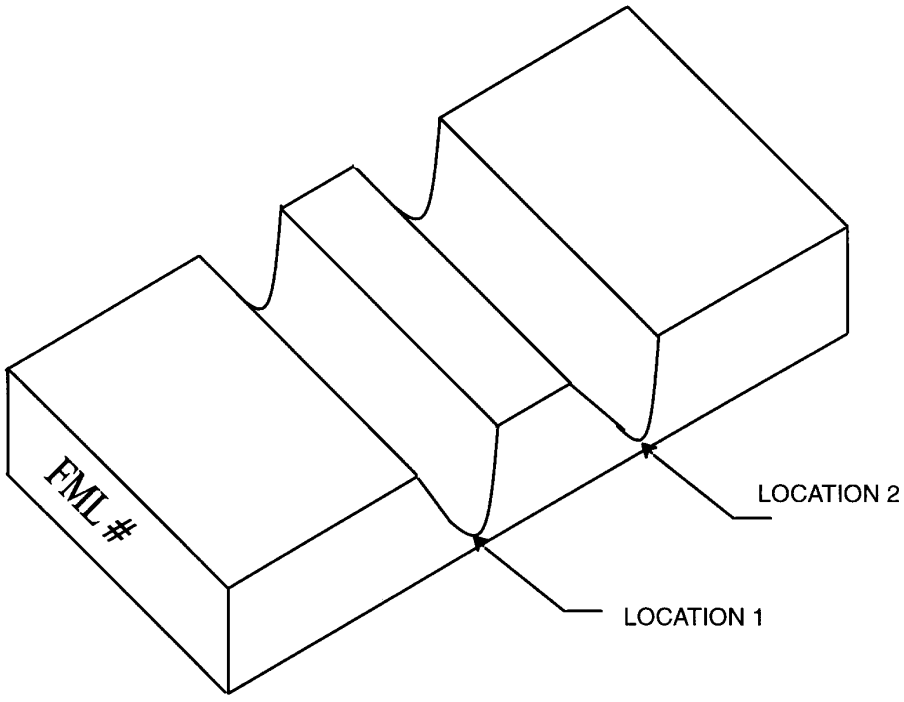
* See figure 15 to reference flaw location with S/N.

FIGURE 14. F100–PW–229 ENSIP manufacturing inspection reliability test, scallop and web/bore surface specimen FPI test data sheet.

APPENDIX F



LOCATION REFERENCE OF SCALLOP SPECIMEN



LOCATION REFERENCE OF WEB/BORE FILLET SPECIMEN

FIGURE 15. Flaw location reference.

APPENDIX F

F.3.3 Maintenance.

Specimens are to be maintained as described in 4.3.2. The goal of these requirements is to preserve specimen integrity for the purpose of inspection system evaluation.

F.3.3.1 Handling.

Specimens should be stored in carrying cases where they will not be subject to metal-to-metal contact. This is to prevent scratching the specimens or damaging the cracks in them accidentally. To assure truly back-to-back system evaluations, it is imperative that the specimens be the same from one test to the next.

F.3.3.2 Cleaning.

Because the inspection process may leave residual material in surface connected defects (e.g., penetrant from PT inspections) and that this material may affect later test results, it is imperative that each specimen be thoroughly cleaned after each use. When the inspection does not use a contaminating fluid (such as ET or UT) wiping the specimen with a soft, lint-free cloth may be sufficient. Use of acetone on the cloth may be useful. Where a penetrant is used, ultrasonic cleaning is necessary. Vapor degreasing may also be appropriate. All chemicals that contact the specimens should be checked to assure that they are not damaging to the specimen material.

F.3.3.2.1 Specimen integrity.

To maintain specimen integrity, the specimens should not be subject to any metal-removing process such as polishing, etching, or sanding.

F.3.3.3 Shipping.

Because the same specimens may be needed for several system demonstrations, and to lower the risk of damage to the specimens in transit, the cases containing the specimens should be handcarried from program to program, or shipped by Next Day Air Freight. Packaging must be sufficient to allow for the rough handling that can be expected.

F.3.3.4 Storage.

USAF specimens will be stored in an office-type environment at Wright-Patterson Air Force Base. WL/MLSA will be responsible for maintaining the inventory of the specimens. However, ASC / ENFP will be the point of contact for requesting use of the specimens for particular testing programs. Other Government agencies will be responsible for their own specimens.

F.3.3.5 Revalidation.

Specimen flaw responses will be measured at least annually or prior to use, by WL/MLSA or other procuring agency using the same test technique and procedure used in the original specimen verification (see F.3.2.5). The flaw response must fall within the range of the responses measured in the original verification process. If it does not, the results must be examined to determine if the specimen has been unacceptably compromised or is salvageable but needs to be recharacterized and verified.

APPENDIX G

MODELING PROBABILITY OF DETECTION

G.1 SCOPE

G.1.1 Scope.

This appendix discusses the mathematical and statistical procedures which have been implemented in the standard POD(A) software. This software is available through the United States Air Force, ASC/ENFP, Wright-Patterson AFB, Ohio, 45433.

G.1.2 Limitations.

G.1.3 Classification.

G.2 APPLICABLE DOCUMENTS

This section is not applicable to this appendix.

G.2.1 Non-Government publications.

The following documents form a part of this appendix to the extent specified.

Cheng and Isles, "Confidence Bands for Cumulative Distribution Functions of Continuous Random *Variables", Technometrics,* Vol. 25, No. 1, February, 1983

Cheng and Isles (1 988), "One-Sided Confidence Bands for Cumulative Distribution Functions", *Technometrics,* Vol. 30, No. 2, May, 1983

Cochran, W. G., "Errors in Measurement in Statistics", *Technometrics, Vol.* 1 0, No. 4, November, 1968

Cramer, H., Mathematical Models of Statistics, (Princeton University Press, 1946)

Kendall and Stewart, "Inference and Relationship", *The Advanced Theory of Statistics,* Vol. 2:, (Charles Griffin, London, 1961)

Lawless, Statistical Models and Methods for Lifetime Data (Wiley, 1982)

G.3 PROCEDURES

G.3.1 Background.

Early attempts to quantify probability of detection, POD, considered the number, n, of cracks detected, divided by the total number, N, of cracks inspected, to be a reasonable assessment of system inspection capability, POD = n/N. This resulted in a single number for the entire range of crack sizes. Since larger cracks are easier to find than smaller ones, cracks were often grouped according to size, and n/N calculated for each size range, as illustrated on figure 16. Grouping specimens this way improved the resolution in crack size, but the resolution in POD suffered because there were fewer specimens in each range. Any attempt to improve the resolution in POD by having more specimens in a given group would necessarily decrease the resolution in crack size. Several methods, such as moving averages and binomial distribution methods, were proposed to circumvent this problem but they required very large sample sizes and suffered from other analytical difficulties.

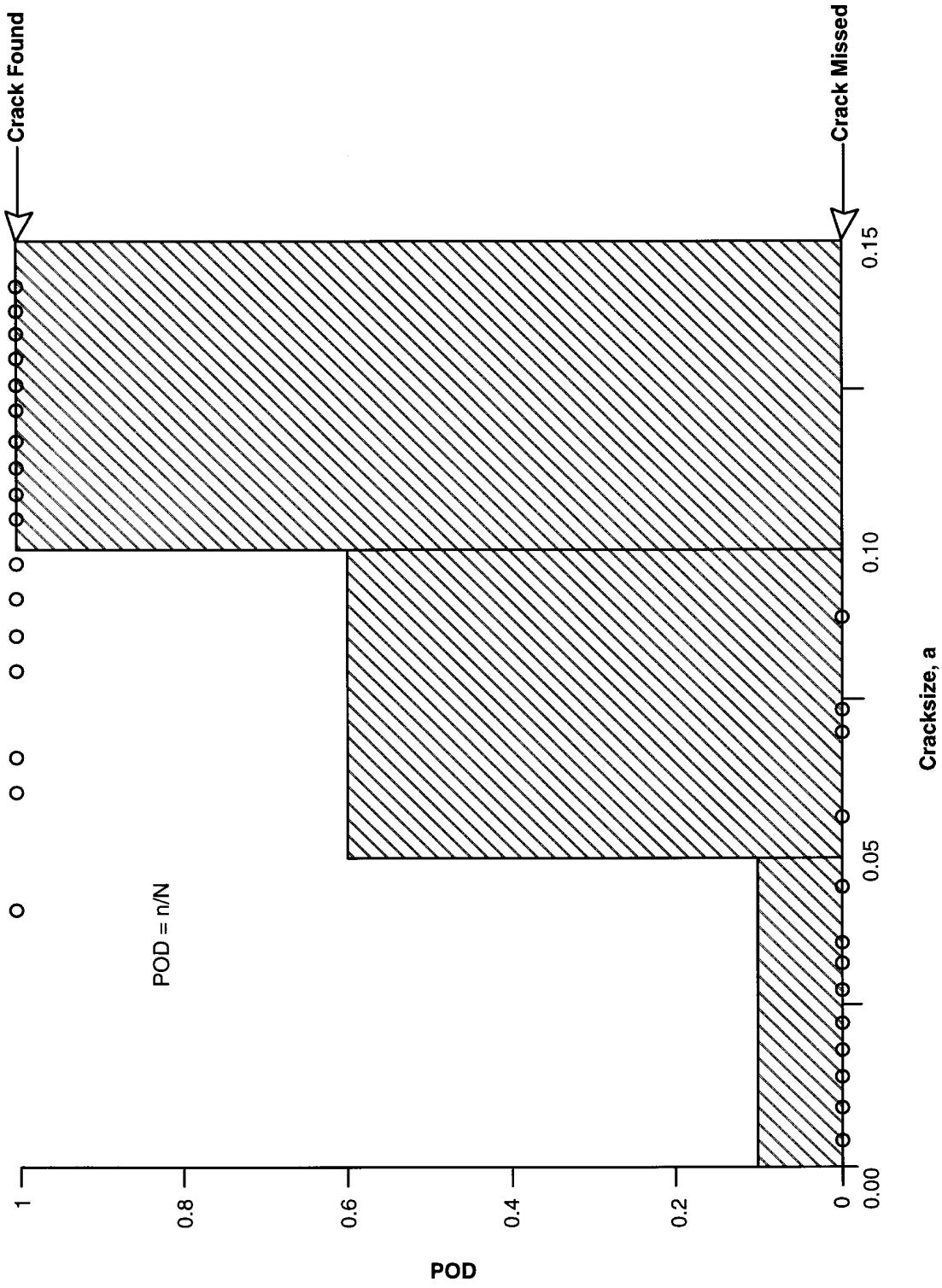


FIGURE 16. Resolution in POD vs. resolution in cracksize.

G.3.1.1 Methods.

The methods in this document are based on a POD(A) model, a mathematical description of the relationship between the size of a crack or defect, a, and its probability of detection, POD. The parameters of the model are estimated by choosing values which are most likely correct, given the results of the inspection being modeled.

G.3.2 Modeling probability of detection â vs. a.

The log normal formulation of the POD(A) model is a natural consequence of the observed behavior of â vs a data, and will be developed here in that context. The same log normal model will be seen to apply also to inspection data where no size information is available. The situation for pass/fail or hit/miss data will be discussed later.

a. Some NDE procedures provide a signal response that is correlated with crack size, If the crack is detected. The data presented as an example in table Iv are for eddy current testing, ET. The magnitude of the eddy current signal is quantitative by correlation with crack size.

TABLE IV. â vs. a Data

Bolthole Specimens, SemiAutomated Inspection

a	â	a	â	a	â
0.001	*	0.012	2.2	0.022	7.7
0.004	*	0.012	3.4	0.023	11.6
0.005	1.5	0.012	2.4	0.023	8.0
0.006	*	0.015	3.0	0.028	**
0.006	1.2	0.016	7.3	0.029	**
0.006	2.6	0.018	7.3	0.030	13.2
0.008	1.2	0.018	4.0	0.034	19.6
0.008	2.8	0.019	5.0	0.036	16.2
0.008	1.6	0.020	7.3	0.052	19.2
0.009	2.7	0.020	11.6	0.058	19.6

Notes:

1. a is crack size in inches
2. â is apparent size (see text)
3. *, ** censored observations
   * unknown, below $\hat{a}_{th}$ = 1.0
   * * unknown, above $\hat{a}_{sat}$ = 20.0

b. Fracture mechanics nomenclature defines crack depth as, a, and the NDE literature refers to crack size indication, or apparent crack size as â, the idea being that â is correlated with a. Consider the 30 specimens given in table IV, where every fatigue crack of size, a (measured in inches), has an associated apparent size, â (measured in scale divisions). The units of actual crack size are those usually associated with crack depth (e.g., mils, inches, mm, microns) although crack length or crack area is sometimes used as the correlative parameter. By contrast, the units of apparent crack size can be nearly anything, e.g.,millivolts, number of contiguous illuminated pixels, total signal counts, or percent of some maximum scale reading. In this discussion these units are major scale divisions representing signal output of the semiautomated system on which the measurements were made.

1. In any real inspection some, fatigue cracks may be too small to be detected by the inspection apparatus. The system output signal, â, is not zero, it is just indiscernible from the noise, i.e., less than $\hat{a}_{th}$. These misses have no associated â value and so are left-censored. Similarly, cracks which are sufficiently large can overwhelm the system, resulting in a saturated signal. Again, the apparent size, â, is unknown, other than that it exceeds some saturation level, $\hat{a}_{sat}$. These saturated observations are right-censored. Given the â vs a data, it is necessary to estimate the probability of detecting a crack of size a , POD(a). The POD(a) function is defined as

$$POD(a) = P(\hat{a} > \hat{a}_{dec}) \qquad [G\text{-}1]$$

where $\hat{a}_{dec}$ is a predetermined detection threshold. This threshold may be set near the system noise level for maximum crack detection sensitivity, or set somewhat above the noise level to improve the system discrimination.

$_{th}$ or a $\hat{a}_{dec}$ when there is no actual crack. This can result from noise introduced by the inspection itself (e.g., improper scan plan, surface irregularities, or probe lift-off) or from some real but innocuous discontinuity in electrical conductivity or magnetic permeability within the material, or from simply setting the pass/fail criterion too close to the material's noise threshold. (The difficulties in assessing these false calls are noted in Section 6.) In any case, a part found to have a questionable indication is subjected to further scrutiny, usually cellulose acetate replication and subsequent microscopic examination.

3. Signal responses which are either obscured by noise, or too large to be measured, are called censored observations. Censored observations are not the same as missing observations; the treatment of missing data is discussed in Section 4.5.

G.3.2.1 Developing the â vs. a model.

Referring to figure 17, it is seen that the logarithms of â and a can be linearly related. For purposes of this text, log(a) refers to the natural logarithm of a. The linear relationship between log â and log a , can be useful, so for the remainder of this discussion, let:

$x = \log a$ and $y = \log \hat{a}$.

The relationship between â and a can now be expressed as:

$$y = \beta_0 + \beta_1 x + \varepsilon$$

and in figure 17 the residual, E, is observed to be approximately normally distributed with zero mean and variance δ^2**.** Several dozen collections of similar data have been studied and the linear relationship with approximately normal residuals occurs quite frequently but not always. For some analyses, it has been necessary to restrict the range of crack size in the analyses to ensure these properties. The residuals of the ten inspections reported here are presented collectively in figure 18.

The POD(x), $P(Y > Y_{th})$, is illustrated as the shaded region under the normal density for log crack size, x in figure 17. As one moves along the x axis, the location (mean) of the normal density of log â values changes ($y= \beta_0 + \beta_1 x$**)** and thus the POD also changes.

Now under the above assumptions, $z = [y - (\beta_0 + \beta_1 x)]/\delta$ [G-2]

Has a standard normal distribution; i.e.,

$\phi(z) = \frac{1}{\sqrt{2\pi}} e^{-(z^2/2)}$, the standard normal pdf, and

$Q(z) = \int_z^\infty \phi(\xi) d\xi$, the standard normal survivor function

Then $POD(x) = P(y > y_{th}) = Q\left[\frac{y_{th} - (\beta_0 + \beta_1 x)}{\delta}\right]$ [G-3]

$POD(x) = 1 - Q\ \frac{x - (y_{th} - \beta_0)/\beta_1}{\delta/\beta_1}$

Hence the POD function is a cumulative normal distribution function with

parameters $\mu = \frac{y_{th} - \beta_0}{\beta_1}$, and $\sigma = \delta/\beta_1$

With these parameters,

$POD(a) = 1 - Q\frac{\log a - \mu}{\sigma}$ [G-4]

Notice that although POD(a) has the form of a cumulative distribution function, it does not represent the cumulative probability of occurrence of a crack of size, a. It represents the probability of detection of cracks of size, a.

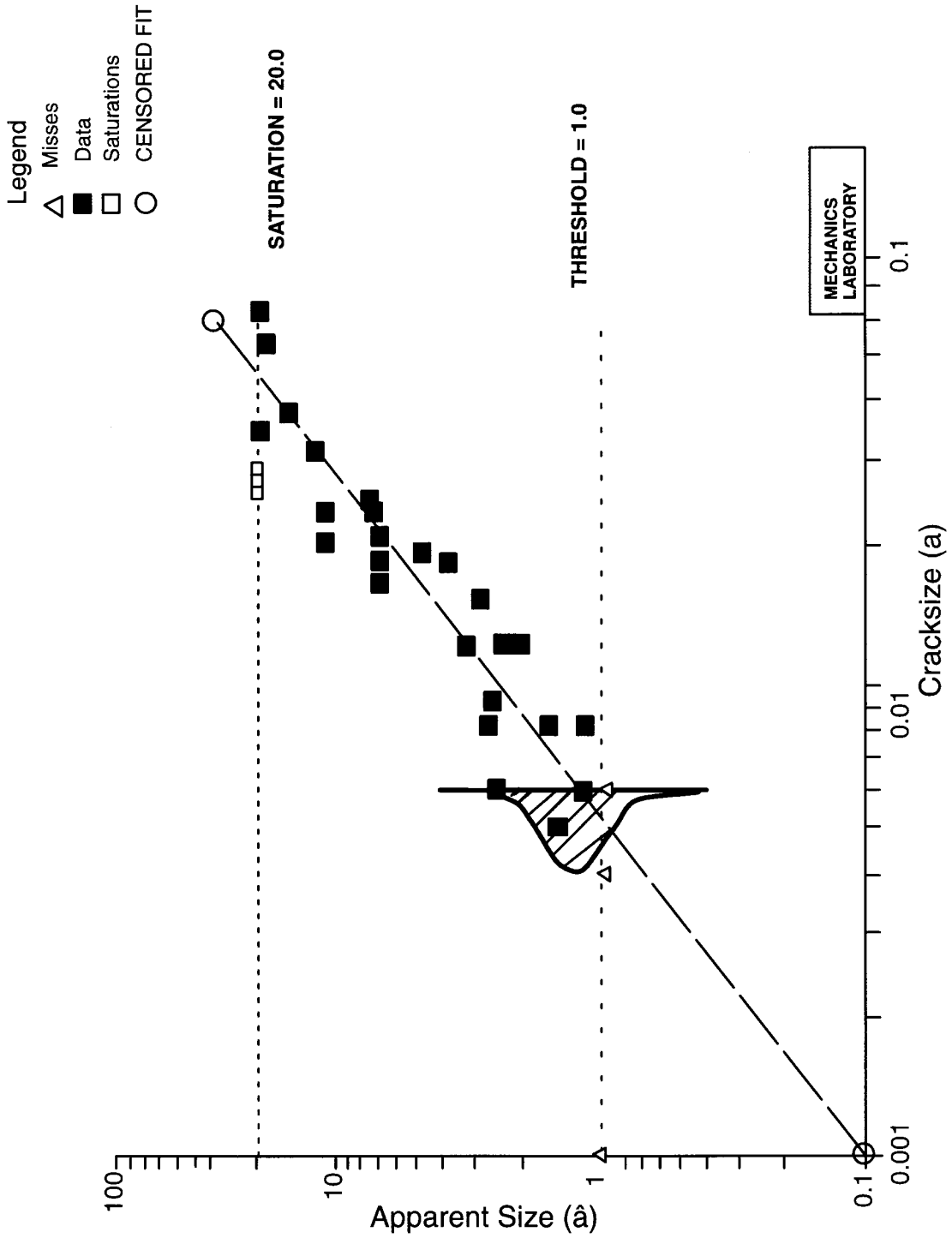


FIGURE 17. Large bolthole specimens. Shaded region is probability of detection.

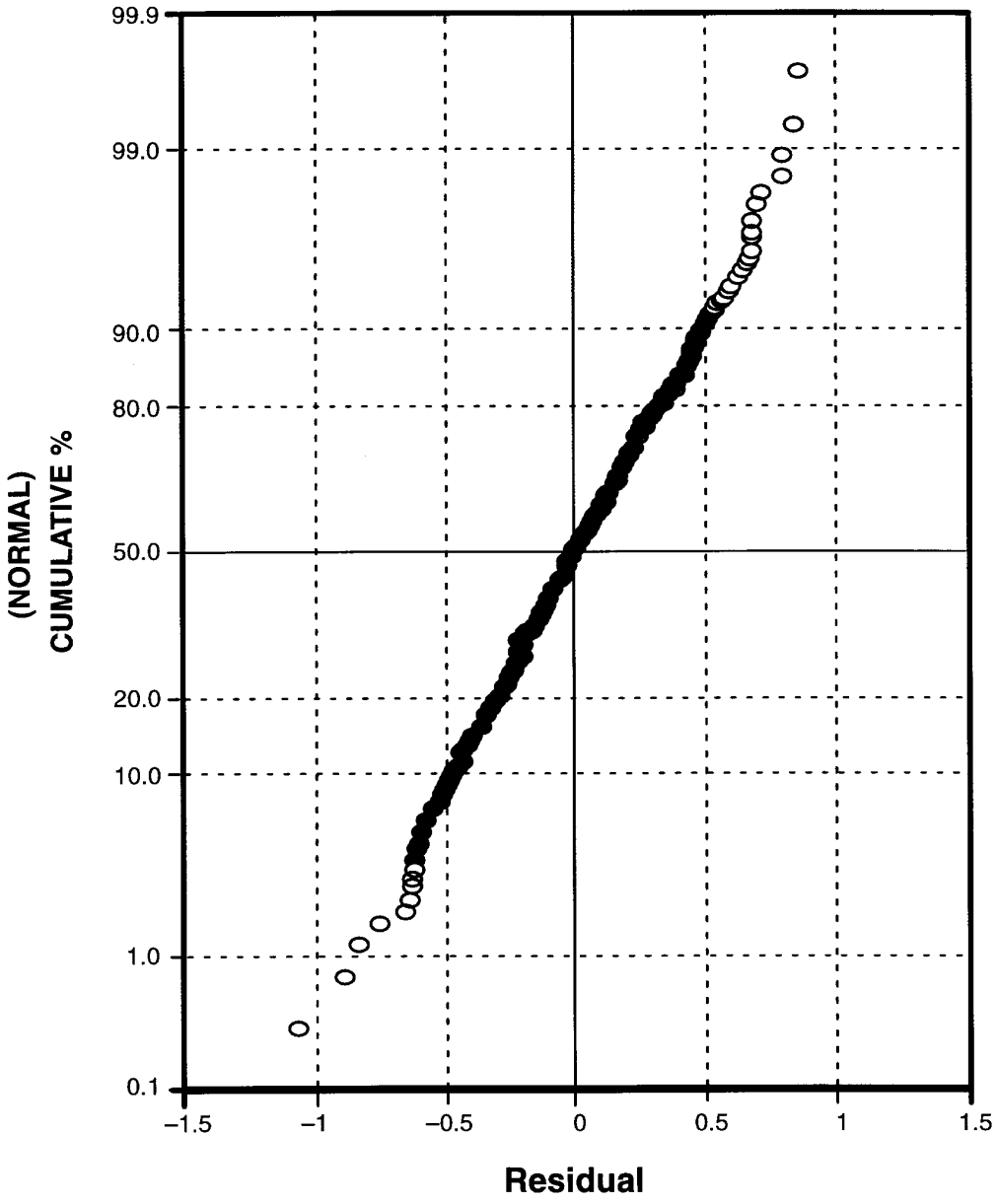


FIGURE 18. **Residuals of 10 inspections are approximately normally distributed.**

G.3.2.2 Effects of uncertainty in crack aspect ratio.
Equation G-4 expresses the probability of detection in terms of a crack size, a. In some experiments, the crack size in the test specimens might be known exactly. For example, in experiments for which the POD would be calculated in terms of a crack length measured on the surface or in experiments using diffusion bonded specimens with exactly defined subsurface voids, the true crack size would be known. In the general NDE reliability experiment, the crack size must be inferred from an assumed or observed crack aspect ratio based either on destructive tests of a few specimens or on experience with the method used to produce the test specimens. In this general case, the differences between the true and inferred crack sizes will have an effect on the POD(a) function. Given a set of specimens for which both the true and estimated sizes are known, the effect of using the estimated crack sizes in obtaining the POD(a) parameters can be quantified.

a. The following presents a method for assessing the magnitude of the effect of using an estimate of the crack size rather than the true (and generally unknown) value. (Cochran, 1968)

1. Define aspect ratio as c = crack length/crack depth. Assume the relation between the measurement of crack length, a_m and the true crack depth, a_t, is given by:

$$\text{Log } a_t = \log a_m - \log c + h$$

where η is normally distributed with zero mean and constant standard deviation σ_h , η accounts or the difference between the calculated crack depth assuming a constant crack aspect ratio and The true crack depth. η the initial analyses of this appendix, the random error term, η-7, was ignored, i.e., it was assumed that the aspect ratio exactly correlated crack length and depth.

2. Assuming that h has zero mean implies that the estimation of the true crack size is unbiased. Assuming that ηhas constant variance implies that the random error is proportional to the size of the crack. These assumptions were reasonable for the specimens that were destructively inspected during the specimen development phase of the Retirement-For-Cause (RFC) Program.

3. Interpreting a_m as a and substituting equation G-4 for log â into equation G-1 for the calculation of POD(a) gives

$$\begin{aligned} POD(a_t) &= P[\hat{a} > \hat{a}_{dec}] = P[\log \hat{a} > \log \hat{a}_{dec}] \\ &= P[b_0 + b_1 \log a_m + e > \log \hat{a}_{dec}] \\ &= P[b_0 + b_1 (\log a_t + \log c - h) + e > \log a_{dec}) \\ &= P[e - b_1 h > \log \hat{a}_{dec} - b_0 - b_1 (\log a_t + \log c)] \end{aligned}$$

Let $\xi = \varepsilon - \beta_1\eta$ and assume that ε and η are independent. Then

$$\sigma\xi = \sqrt{\delta^2 + \beta_1^2 \sigma^2_\eta}$$

4. Thus the variability observed about the â vs a_t relationship is inflated by an amount $\beta_1^2\sigma_\eta^2$. The POD(a) function is then, after simplification:

$$POD(a_t) = 1 - \phi\left[\frac{\log a_t + \log c - \left(\frac{\log \hat{a}_{dec} - \beta_0}{\beta_1}\right)}{\sigma\xi / \beta_1}\right] \qquad [G\text{-}5]$$

5. Very little experience has been acquired in the analysis of the relationship between the true and measured crack sizes. In the experiments conducted during 1985 - 1988 to evaluate the RFC NDE system, the value of σ_η, was observed to be significantly smaller than δ and the effect of scatter about the crack aspect ratio was negligible, and so equation G-3 is used.

G.3.2.3 Maximum likelihood estimators.

The estimates of the POD(a) parameters discussed in this document are maximum likelihood estimates (MLEs), which have several desirable statistical properties. Two are especially Important.

1. MLEs are sufficient statistics. That is, for a given underlying statistical model, knowing the MLE is just as good as knowing the actual sample data, as far as knowing the true values of the model parameters is concerned.

2. MLEs themselves have known statistical properties. For large samples this distribution is very nearly normal, and centered at the true parameter values.

G.3.2.3.1 Normal behavior of likelihood.

Because this normal behavior is fundamental to much of the analysis of NDE data, a brief discussion of likelihood is in order. Likelihood is analogous to probability, but with a subtle twist: A probability distribution describes the behavior of the data, given the distribution's parameters $\underset{\sim}{\theta}$. By comparison, the likelihood describes the behavior of the parameters, given the data. The data are considered fixed, since they have already been observed; it is the model Parameters, then, which vary according to the given statistical model. This is written as $L(\underset{\sim}{q};\ \underset{\sim}{C})$ where the undermark indicates a matrix of values. The mathematical formulation of the likelihood and its corresponding probability density are identical; they differ only in whether it is the data which are considered fixed (likelihood) or the parameters which are fixed (probability).

G.3.2.3.1.1 Variance-covariance matrix.

The variance - covariance matrix, which summarizes the behavior of the maximum likelihood estimators, can itself be estimated from the sample data. Thus, the likelihood function provides not only the model parameters, but estimates of their variability as well.

G.3.2.3.1.2 Normal behavior of maximum likelihood (ML) parameter estimators.

The asymptotically normal behavior of the maximum likelihood parameter estimators is exploited to provide confidence bounds for POD(a) curves (G.3.4) and to make statistical comparisons between and among different inspections (Appendix H).

G.3.2.4 Parameter estimation â vs. a.

To determine the relationship, POD(a), it is necessary to estimate β_0 , β_1 , and δ of equation G-2. For uncensored data, these can be determined using the familiar least-squares regression equations.

G.3.2.4.1 Parameter estimates for censored data.

When some observations are censored and therefore no â value exists, the regression approach becomes untenable. That is because the true location of the observation is unknown other than being less than the noise threshold or greater than the system signal saturation level. Since the true location is unknown, the difference between the observation and the model is also unknown. The equations based on minimizing this (squared) deviation are therefore unworkable.

a. In this circumstance, the method of maximum likelihood can be used to obtain parameter estimates for the censored data.

b. Lawless (1982) discusses a generalized case of a normal parametric model where the data are right-censored. For data influenced by both right and left-censoring, order the data, so that $\hat{a}_1 < \hat{a}_2 < \ldots < \hat{a}_n$ and let index:

i = 1, …, m represent data obscured by system noise. ($\hat{a} < \hat{a}_{th}$)

i = m+ 1, ..., m+ r represent data for which a valid signal response exists, and

i = m+ r+ 1, ..., n represent saturated signal data, ($\hat{a} > a_{sat}$)

The likelihood of an observation at *z* is $\frac{1}{\delta}\phi(z)$ and the likelihood for the set of independent, uncensored, observations. is then;

$$L(\beta_0, \beta_1, \delta, \underset{\sim}{x}, \underset{\sim}{y}) = \prod_{i=m+1}^{m+r} \frac{1}{\delta}\phi(z)$$

c. Only slight modification of this definition is required to address censored observations. In the case of right-censored observations, the likelihood is simply the proportion of the distribution, centered at $y = \beta_0 + \beta_1 x$, which lies above the censoring value. $y_{sat.}$ Similarly, for left-censored data, the likelihood is the proportion of the distribution below y_{th}.

d. The complete likelihood for all three situations is then

$$L(\beta_0, \beta_1, \delta, \underset{\sim}{x}, \underset{\sim}{y}) = \prod_{i=1}^{m} (1 - Q(z_{th})) \prod_{i=m+1}^{m+r} \frac{1}{\delta}\phi(z) \prod_{i=m+r+1}^{n} Q(z_{sat})$$

e. The likelihood will reach a maximum when its first derivatives with respect to the model parameters approach zero. Since the logarithm is a monotonic function, the maximum of log likelihood will coincide with that of the likelihood itself. Taking the logarithm of $L(b_0, b_1, , d ; \underset{\sim}{x}, \underset{\sim}{y})$ greatly simplifies the subsequent differentiations by reducing the series of products to one of sums. The log likelihood is

$$\log L(b_0, b_1, , d ; \underset{\sim}{x}, \underset{\sim}{y}) = \qquad \text{[G-6]}$$

$$\sum_{i=1}^{m} \log\left(1 - Q\left(z_{th}\right)\right) - r \log \delta - \frac{1}{2\delta^2} \prod_{i=m+1}^{m+r} \left(y - \left(\beta_0 + \beta_1 x\right)\right)^2 + \prod_{i=m+r+1}^{n} \log Q\left(z_{sat}\right)$$

f. It is necessary to find b_0, b_1 and d such that the first partial derivatives of the log likelihood in equation G-6 are zero. The matrix of these partial derivatives is referred to as the score.

G.3.2.5 Estimation algorithm for â vs. a data.

The parameters which maximize the likelihood equation, G-6, are evaluated iteratively using the following equations.

a. The elements of the score, which was mentioned In the preceding section, are:

$$\frac{\partial \log L}{\partial \beta_0} = \frac{1}{\delta}\left\{\sum_R z + \sum_S v(z) - \sum_M W(z)\right\}$$

$$\frac{\partial \log L}{\partial \beta_1} = \frac{1}{\delta}\left\{\sum_R xz + \sum_S xV(z) - \sum_M xW(z)\right\}$$

$$\frac{\partial \log L}{\partial \delta} = \frac{1}{\delta}\left\{-r + \sum_R z^2 + \sum_S zV(z) - \sum_M zW(z)\right\}$$

where:

$V(z) = f(z)/Q(z)$

$W(z) = f(z)/[1 - Q(z)]$

b. The matrix of negative second partial derivatives of the likelihood equation with respect to the model parameters is called the Fisher information matrix. The information matrix is used by the iteration procedure for estimating values for b_0, b_1 and d which will maximize equation G-6. Its inverse is the variance-covariance matrix of the model parameters, which is used in placing confidence limits on the POD(a) relationship (see G.3.2).

c. Elements of the Fisher information matrix are estimated by:

$$\frac{-\partial^2 \log L}{\partial \beta_0^{\,2}} = \frac{r}{\delta^2} + \frac{1}{\delta^2}\sum_S \lambda(z) - \frac{1}{\delta^2}\sum_M \psi(z)$$

$$\frac{-\partial^2 \log L}{\partial \beta_0 \partial \beta_1} = \frac{1}{\delta^2}\sum_R x + \frac{1}{\delta^2}\sum_S x\lambda(z) - \frac{1}{\delta^2}\sum_M x\psi(z)$$

$$\frac{-\partial^2 \log L}{\partial \beta_0 \partial \delta} = \frac{2}{\delta^2}\sum_R z + \frac{1}{\delta^2}\sum_S V(z) + \frac{1}{\delta^2}\sum_S z\lambda(z) - \frac{1}{\delta^2}\sum_m W(z) - \frac{1}{\delta^2}\sum_M z\psi(z)$$

$$\frac{-\partial^2 \log L}{\partial \beta_1^{\,2}} = \frac{1}{\delta^2}\sum_R x^2 + \frac{1}{\delta^2}\sum_S x^2\lambda(z) - \frac{1}{\delta^2}\sum_M x^2\psi(z)$$

$$\frac{-\partial^2 \log L}{\partial \beta_1 \partial \delta} = \frac{2}{\delta^2}\sum_R xz + \frac{1}{\delta^2}\sum_S xV(z) + \sum_S xz\lambda(z) - \frac{1}{\delta^2}\left[\sum_M xW(z) - \sum_M xz\psi(z)\right]$$

$$\frac{-\partial^2 \log L}{\partial \beta_0^{\,2}} = -\frac{r}{\delta^2} + \frac{3}{\delta^2}\sum_R z^2 + \frac{2}{\delta^2}\sum_S zV(z) + \frac{1}{\delta^2}\sum_S z^2\lambda(z) - \frac{2}{\delta^2}\sum_M zW(z) - \frac{1}{\delta^2}\sum_M z^2\psi(z)$$

where,

$\lambda(z) = V(z)[V(z)\text{-}z]$

$\psi(z) = -W(z)[W(z)+z]$

d. The variance-covariance matrix of the log â vs log a regression parameter estimates is related to the Fisher information by

$$\text{Var}\left(\hat{\beta}_0, \hat{\beta}_1, \hat{\delta}\right) = \begin{bmatrix} V_{00} & V_{01} & V_{02} \\ V_{10} & V_{11} & V_{12} \\ V_{20} & V_{21} & V_{22} \end{bmatrix} = I\left[\left(\hat{\beta}_0, \hat{\beta}_1, \hat{\delta}\right)^T\right]^{-1}$$

e. The elements of this matrix are in terms of the log â vs log a relationship. It is necessary to convert this matrix to the corresponding 2 X 2 variance-covariance matrix of the POD(a) model parameter estimates.

f. Using a Taylor series expansion about the true values of μ and σ, the appropriate variance-covariance matrix of $\hat{\mu}$ and $\hat{\sigma}$ is given by:

$$Var(\hat{\mu},\hat{\sigma}) = \frac{1}{\hat{\beta}_1^{\,2}} T\, Var(\beta_0,\beta_1,\delta) T$$

and the transformation matrix T is defined by:

$$T = \begin{bmatrix} 1 & \mu & 0 \\ 0 & \sigma & -1 \end{bmatrix}$$

g. Performing the indicated matrix operations provides estimates of the variances and covariances of $\hat{\mu}$ and $\hat{\sigma}$ as

$$Var(\hat{\mu}) = \frac{1}{\hat{\beta}_1^{\,2}} \left[V_{00} + 2\hat{\mu} V_{01} + \hat{\mu}^2 V_{11} \right]$$

$$Var(\hat{\mu},\hat{\sigma}) = \frac{1}{\hat{\beta}_1^{\,2}} \left[\hat{\sigma} V_{01} - V_{20} - \hat{\mu} V_{12} + \hat{\mu}\hat{\sigma} V_{11} \right]$$

$$Var(\hat{\sigma}) = \frac{1}{\hat{\beta}_1^{\,2}} \left[V_{22} - 2\hat{\sigma} V_{21} + \hat{\sigma}^2 V_{11} \right]$$

h. Inverting this 2 X 2 variance-covariance matrix produces the 2 X 2 Fisher information matrix used to place lower bounds on POD(a) curves, as discussed later in Appendix G.

G.3.2.6 Newton-Raphson Iteration:

The Newton-Raphson iteration finds a zero of a function by (grossly) approximating the function with a tangent plane at a point, and solving directly for the zero of the plane. Then the function is evaluated at this zero point. If the function itself is not zero, the process is repeated using this new point as the reference. The function in this instance is the score vector, the derivatives of the likelihood with respect to the model parameters. When these derivatives are zero, the likelihood will be at its maximum. The coordinates of the zero point, $\left(\hat{\beta}_0,\hat{\beta}_1,\hat{\delta}\right)^T$, are therefore the maximum likelihood estimates for the model parameters.

Given $\left(\hat{\beta}_0,\hat{\beta}_1,\hat{\delta}\right)^T$, is the vector of parameter estimates after k iterations,

Let *U* $\left(\hat{\beta}_0,\hat{\beta}_1,\hat{\delta}\right)^T$ be the score vector, and

Let *I* $\left(\hat{\beta}_0,\hat{\beta}_1,\hat{\delta}\right)^T$ be the Fisher information matrix, as described above.

a. The Newton-Raphson procedure uses uncensored MLEs as initial guesses, and solves

$$\left(\hat{\beta}_0, \hat{\beta}_1, \hat{\delta}\right)_{k+1} = \left(\hat{\beta}_0, \hat{\beta}_1, \hat{\delta}\right)_k + \left[I\left[\left(\hat{\beta}_0, \hat{\beta}_1, \hat{\delta}\right)_k\right]\right]^{-1} U\left[\left(\hat{\beta}_0, \hat{\beta}_1, \hat{\delta}\right)_k\right]$$

Until $abs\left(\left[I\left[\left(\hat{\beta}_0, \hat{\beta}_1, \hat{\delta}\right)_k\right]\right]^{-1} U\left[\left(\hat{\beta}_0, \hat{\beta}_1, \hat{\delta}\right)_k\right]\right) \leq \xi_1$

Or until $U\left(\hat{\beta}_0, \hat{\beta}_1, \hat{\delta}\right) \leq \xi_2$

Where ξ_1 and ξ_2 are convergence criteria.

b. Examples:

1. Data in table IV and similar data for nine other inspections were analyzed using the parameter estimation procedure described here. The test was a One-Factor-at-a-Time design. (Designs of NDE demonstration and evaluation experiments are discussed in Appendix H.)

2. The inspections designated Al, B1, B2, B3 are repeated evaluations of the (unchanged) NDE system. The same operator performed all four inspections using the same eddy current probe. Next, the inspection probe, and therefore system calibration parameters, were changed and designated as inspection C. Inspections G and H changed the physical orientation of the fatigue-cracked specimens being inspected. All system parameters were identical to inspection C. Finally, a new operator performed inspections J1J2, J3. Results are summarized in table V. A representative plot of the POD vs a relationship (Test A1) is provided as figure 19.

TABLE V. Model parameters for semi--automated inspections.

Test	a_{50}	σ	β_0	β_1	δ	n_1
A1	0.00498	0.2693	7.5271	1.4195	0.3822	30; 3; 2
B1	0.00526	0.2343	7.7306	1.4733	0.3452	30; 3; 2
B2	0.00489	0.2642	7.9070	1.4863	0.3926	30; 3; 2
B3	0.00473	0.3070	7.3941	1.3812	0.4240	30; 3; 2
C	0.00474	0.1968	8.4873	1.5859	0.3120	30; 3; 4
G	0.00484	0.2549	7.6671	1.4384	0.3666	30; 3; 3
H	0.00503	0.3070	7.7186	1.4585	0.4477	30; 4; 2
J1	0.00557	0.2379	7.7638	1.4956	0.3558	30; 4; 3
J2	0.00520	0.2012	8.2517	1.5691	0.3157	30; 3; 4
J3	0.00596	0.4662	7.2437	1.4142	0.6594	30; 6; 1

Notes:
1. $a_{50} = e^{\mu}$, crack size at 50 % POD.
2. Inspections A1, B1, B2, and B3 are operator 1, repeat tests. Probe and system calibration, unchanged.

3. Inspection C changed probe.
4. Inspection G and H changed specimen orientations.
5. Inspection J1, J2. and J3 are operator 2, repeat tests.
6. n_1 = total observations, n_2 = data in noise, n_3 = saturations.

G.3.3 Hit/miss analysis.

Fluorescent penetrant testing, PT, magnetic particle testing, MT, and ultrasonic testing, UT, tend to be characterized by their binary nature: either the crack is detected (hit or 1) or it is not (miss or 0). Unlike eddy current inspection data for which some crack size information is available, PT MT and UT data are usually hit/miss only. This presents an analysis difficulty since it precludes using the â vs. a procedure because there is no â. The â vs. a analysis, discussed in detail previously, is based on a normal distribution of apparent size, â, or a crack of actual size a, the model parameters being estimated by maximizing the likelihood of the test results was based on this normal distribution. By comparison , PT, MT, and UT data is binomial in nature with detection probability given by POD(a). Maximum likelihood is used to estimate the parameters of the model. The idea in both cases is to select model parameter estimates such that the likelihood is maximized based on the model, given the actual data observed.

a. For hit/miss testing, the likelihood of *P*, based on a single observation, is:

$$L(P_i; a_i, x_i) = P_i^{x_i}(1 - P_i)^{1-x_i} \qquad \text{[G-7]}$$

where P_i, is the probability of detection of crack size a_i, and x_i is the inspection outcome, 0 for miss, 1 for hit. (Notice that when the exponent of *P*, is one, that of (*1 - P*$_l$) is zero, and so that factor, (1 - P_i^0) , reduces to multiplication by one. Similarly with P_i^x , when *x* is zero.) P_i is a function of crack size, a_i, and the log normal model can be used to relate P_i= POD(a_i) with crack size.

The model formulation is

$$P_i = \text{POD}(a_i) = 1 - Q(z_i) \qquad \text{[G-8]}$$

where

$Q(Z_i)$ is the standard normal survivor function,

$z_i = \left[\dfrac{\log a_i - \mu}{\sigma}\right]$, is the standard normal variate,

μ, σ are the location and scale parameters,

and $\underset{\sim}{\theta} = \begin{bmatrix}\mu \\ \sigma\end{bmatrix}$

b. The log odds function, which is an approximation to the log normal, is often suggested in similar situations to model binary data. The log normal model is used here to be consistent with the POD(a) model resulting from â vs. a data.

Recall that P_i the probability of detecting crack size a_i and is given as P_i= POD(a_i) in equation G-8. The outcome of the *i* th inspection, x_i, is either a one for a hit or a zero for a miss. d. The overall likelihood of having observed all the data is, then, the product of their individual likelihoods. So for hit/miss data the likelihood is

$$L(\underset{\sim}{\theta}; \underset{\sim}{a}, \underset{\sim}{x}) = \left[\prod_{i=1}^{h} P_i\right]\left[\prod_{j=1}^{n-h}(1- P_i)\right] \qquad \text{[G-9]}$$

where the likelihood of the *(h)* hits is the first term of equation G-9, and the second term is the likelihood of the (*n - h)* misses. (Note that *P*(miss) = 1 - *P*(hit).)

e. Now, values for μ and σ equation G-8 can be selected to maximize the likelihood, equation G-9. Taking the natural logarithm of equation G-9 changes the series of products into a series of sums. The log likelihood is given as equation G-10.

$$\log L\ (\underset{\sim}{0}\ ;\ \underset{\sim}{a}\ ,\ \underset{\sim}{x}\ ) = \sum_{i=1}^{h} \log P_i + \sum_{j=1}^{n-h} \log(1- P_j) \qquad \text{[G-10]}$$

f. Because the logarithm is a monotonic function, the maximum of the log likelihood will coincide with the maximum of the likelihood itself. Therefore equation G-10 can now be differentiated with respect to μ and σ, the derivatives set equal to zero, and the resulting two equations solved simultaneously. In practice it is convenient to perform these differentiations numerically rather than algebraically, as was done in the case of â vs. a. As with the â vs. a analysis, the negative second partial derivatives of the log likelihood provide the Fisher information matrix, used to place confidence bounds on the POD(a) relationship.

G.3.4 POD vs a confidence bounds.

a. Confidence bounds can be placed on the POD vs. a relationship by taking advantage of the asymptotically normal behavior of the maximum likelihood estimators (MLE). It is true that ML estimators, $\underset{\sim}{\hat{\theta}}$, have an asymptotically multivariate normal distribution with mean $\underset{\sim}{\theta}$ and variance-covariance matrix $\left[I\left(\underset{\sim}{\theta}\right)\right]^{-1}$ (Kendall and Stewart, 1961 or Cramer, 1946) and consequentially that

$$\Omega(\underset{\sim}{\theta}) = (\underset{\sim}{\hat{\theta}} - \underset{\sim}{\theta})^{T}\, I\ (\underset{\sim}{\theta})\ (\underset{\sim}{\hat{\theta}} - \underset{\sim}{\theta}) \qquad \text{[G-11]}$$

is asymptotically a chi-squared variable with k degrees of freedom for a k-parameter model. The expected Fisher information for a two parameter normal model is estimated as part of the ML parameter estimation procedure.

a. a. Since the POD model is a cdf, 1 - Q (x ; $\underset{\sim}{\theta}$), the Cheng and Iles (1983, 1988) method of placing confidence bounds on a cdf, can be applied to the POD equation.

b. Plot the cdf scale and location parameters, respectively, and define C to be their confidence region. From equation G-11 it is seen that as $[\mu, \sigma]^T$ vary about $[\hat{\mu}, \hat{\sigma}]^T$ within C, they

describe an elliptical boundary for a given Ω. As μ and σ move about within this region, the cdf (and therefore POD(a)) changes.

c. Now consider x_p, the p th quantile, which is defined by $P[x \le x_p] = 1 - Q(x_p ; \underset{\sim}{\theta}) = p$. For a fixed p allow $\underset{\sim}{\theta}$ to vary within C and examine the behavior of x_p.

d. For a normal cdf, the p th quantile is given by:

$$(x_p - \mu)/\sigma = Q^{-1}(1 - p) = t$$

and so

$$x_p = m + t\, s \qquad \text{[G-12]}$$

e. All combinations, $\underset{\sim}{\theta}$, within C, can be obtained from equation G-12 by holding p constant.

f. Now, x_p will achieve its extreme values along the boundary of C, as given by equation G-11. The largest log crack size, $x_p(\max)$, which satisfies both equations G-11 and G-12 can be calculated using the method of Lagrangian multipliers. The Lagrangian is:

$$g(x_p , h; \underset{\sim}{q}) = x_p + h W(m , s) \qquad \text{[G-13]}$$

where $\Omega(\mu, \sigma)$ is given by equation G-11 , x_p by equation G-12, and η is the Lagrangian multiplier. Differentiating equation G-13 with respect to θ and equating these to zero, then eliminating η, provides the necessary equations for determining $x_p(\max)$. By repeating the evaluation of $x_p(\max)$ for all p, the desired confidence band on POD(a) can be constructed. The 95% lower confidence bound on POD illustrated on figure 19 was determined in this fashion using the standard software.

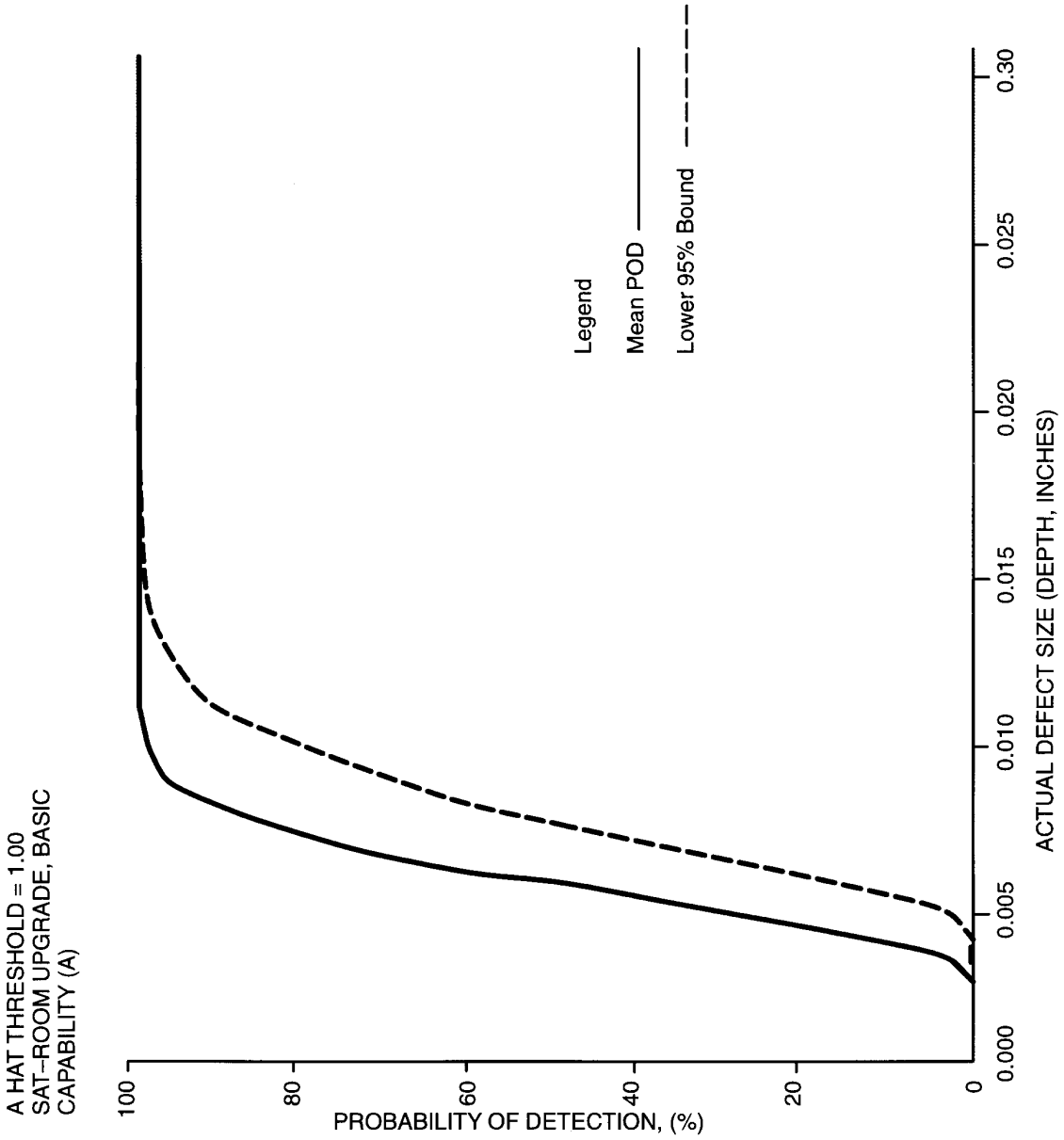


FIGURE 19. POD vs. A ECI data analysis PWA 1074 bolthole specimens.

APPENDIX H

ASSESSING SYSTEM CAPABILITY

H.1 SCOPE

H.1.1 Scope

This appendix addresses the methods for assuring that the estimated POD(a) curve is a valid representation of NDE system capability. It includes tests of model and data compliance, as well as statistical methods for comparing POD(a) relationships to assure that they can be combined to represent the entire NDE system.

H.1.2 Limitations.

The POD(a) characterization of capability is summarized by the model parameters, μ and σ, and represented by the resulting POD(a) curve. The lower bound, discussed in Appendix G, reflects the statistical uncertainty of the estimate of POD(a) function. The estimate and its lower confidence bound are compared with the system requirements as specified by the CDRL. In some instances these requirements will not have been met. Ancillary investigations described here may be required to isolate the cause(s) of inadequate system capability so that remedial action may be undertaken.

H.1.3 Classification.

H.2 APPLICABLE DOCUMENTS

This section is not applicable to this appendix.

H.2.1 Non-Government publications.

The following documents form a part of this appendix to the extent specified.

Johnson and Wichern, Applied Multivariate Statistical Analysis, 2nd ed., (Prentice Hall, 1988)

H.3 PROCEDURES

H.3.1 Statistical tests for model compliance.

Decisions made about the capability of the system to meet its requirements are based on the POD model. Before these decisions can be made, the "goodness" of the POD model may be assessed. If the model fails these tests, then the decisions made regarding the system through use of the model may be erroneous. The NDE reliability analyses are based on the assumption that the relationship between crack size and the probability of detection can be modeled by the cumulative lognormal distribution function. The analysis programs will usually (but not always) produce answers even if this assumption is not reasonable. Therefore, consideration may be given to the viability of the model in each new application. Different approaches to validating the model are required for the $\hat{a}$ vs. a data and hit / miss data.

H.3.1.1 $\hat{a}$ vs. a model compliance.

The cumulative lognormal function for POD(a) was derived by assuming that:

a. the mean of log $\hat{a}$ is a linear function of log a;

b. the regression residuals are normally distributed with zero mean; and,

c. the standard deviation of the residuals is constant for all values of a.

As a minimum, these assumptions may be subjectively evaluated by a visual examination of a plot of log â vs. log a for each data set. In general, regression analysis methods are robust with respect to the assumptions of normality and constant standard deviation of the residuals. There are also standard statistical tests of these assumptions which can be used to remove subjectivity from the validation of the assumptions. However, it should be noted that the tests for constant variance and normality of the residuals are relatively insensitive for the recommended minimum number of cracks in NDE reliability experiments. If any of the basic assumptions are not valid, the discrepancies may be noted on all reported parameter values and plots derived from the data using the standard analysis method.

When the log response signal is not linear with log crack size, it is likely to be concave downward at the larger crack sizes. Ignoring this type of nonlinearity results in values of a_{50} that are too small and values of σ that are too large. This combination of wrong parameter values will yield overestimates of POD at small crack sizes and underestimates of POD at large crack sizes. Restricting the range of crack sizes in the analysis may correct this difficulty when the linear range extends to crack sizes which produce very high probability of detections.

For the POD(a) model to be sensible, it is also necessary that the slope of the log â vs. log a line be positive. The standard computer program checks for a positive slope. If the slope of the log â vs. log a line is negative, the signal response is not an appropriate metric for making a hit / miss decision in the NDE system as the POD(a) function decreases with crack size. (If this occurs, the NDE system should not have reached the capability evaluation stage.) If the slope is positive but not significantly greater than zero, the lower confidence bound on the POD(a) function will not be monotonic and will eventually curve down. In this case the computer program will not produce a lower bound for the POD(a) function and will output the message 'INADEQUATE FIT TO THE POD MODEL'.

It should be noted that it is possible to develop a POD(a) function from different sets of assumptions regarding the â vs. a relation. However, these have not been implemented.

H.3.1.2 Hit/miss model compliance

Because 0 / 1 data cannot be easily plotted as decimal fractions, assessing the goodness-of-fit of the POD model is less straight forward than with â vs. a data. When there are several inspections of the same crack, a plot of the estimated POD(a) function can be superimposed on the observed detection proportions for each crack in the experiment. The comparison of model to data will be based on a subjective comparison of the fit. If only one inspection has been performed on each crack, the observed data will all be plotted at 0 or 1 and the comparison of model to data is difficult. If multiple inspections have been performed on each crack, there should be data points in the range of increase of the POD(a) function. In this case the subjective evaluation of the fit is easier.

There are two experimental situations in the hit/miss analysis which permit a less subjective evaluation of the cumulative lognormal model. If each crack in the experiment was inspected a large number of times or if a very large number of different cracks were used in the NDE reliability experiment, then the applicability of the model can be checked by the linearity of log of the odds of detection versus log of crack size.

$\log \frac{\mathrm{POD(a)}}{1+\mathrm{POD(a)}} = c_0 + c_1 \log a$, where c_0 and c_1 are the intercept and slope, respectively.

The cumulative lognormal distribution function is approximated by the log-odds model,

If a large number (say more than 20) inspections were performed on each crack, reasonable detection probabilities would be available for the cracks in the range of increase of the POD(a) function (assuming such crack sizes were in the experiment). Similarly, if a large number of different cracks (say more than 200) were used in the experiment, they could be grouped into independent size ranges and the detection probability assigned to the midpoint of each range. A plot of the log of the odds versus log crack size would provide an indication of the linearity of the relation (either subjectively or statistically evaluated).

There are other methods for evaluating goodness-of-fit for dichotomous data, and some statistical data analysis software packages, such as SAS, have algorithms for assessing goodness-of-fit for binary data.

H.3.2 Drawing conclusions from over POD (a)

The NDE evaluation experiment has been designed to establish the capability of the NDE system in terms of a representative POD(a) curve and its lower 95 percent confidence bound. The capability of the NDE system is then compared to the requirements as specified in the SOW/SRD. If the system fails to meet the requirements, a properly designed evaluation experiment may provide the information required to identify the source of the problem. If the evaluation experiment was not properly designed, it may be necessary to conduct additional experiments to isolate the cause(s) of the noncompliance.

The SOW/SRD capability requirements are typically expressed in terms of the flaw size, which corresponds to a high probability of detection. The requirement may be stated for the best estimate of the capability (as quantified by the POD function) or for a conservative capability evaluation (as quantified by the lower 95 percent bound on the POD function). The best estimate of the POD(a) function is completely determined from $\hat{\mu}$ and $\hat{\sigma}$, on the estimates of the parameters μ and σ. The lower 95 percent confidence bound depends both on μ and $\hat{\sigma}$ on the variance-covariance matrix which measures the statistical (sampling) variation in the estimates of μ and σ. The larger the number of flaws in the experiment, the closer the confidence is bound to the estimate.

The parameter μ defines the crack size which is detected 50 percent of the time, $a_{50} = \exp(\mu)$. This crack size is defined as the median detectable crack size of the system. Under the lognormal POD(a) model of this document, the crack size which is detected p percent of the time is given by $a_p = \exp(\mu) \exp(z_p \sigma)$, where z_p is the p th percentile of the standard normal distribution. For example, $a_{90} = \exp(\mu) \exp(1.282\ \sigma)$. If POD(a) is plotted against log a , increasing μ with σ fixed shifts the function to the right without changing its shape. Increasing σ with μ fixed, holds the location (the median detectability) but flattens the curve (larger flaw sizes are required to reach a fixed POD).

A system will fail to meet requirements if the POD(a) function (or its lower confidence bound) is too low at a specified crack size. To improve the capability, μ or σ will have to be reduced. (The confidence bound can be tightened by increasing the number of flaws in the evaluation experiment. Note, however, that the larger the value of σ, the more samples are required to achieve equivalent widths of the confidence bounds). The median detectability, $\exp(\mu)$, tends to be determined by decision thresholds while POD flatness, σ, tends to be determined by variation in system response when applied to flaws of the same size.

Taking measures to improve the system capability can be viewed at two levels: process optimization and process variation reduction. To provide an intuitive distinction between process optimization and process variation reduction, consider that any inspection process can be viewed as applying a stimulus to the structure and interpreting the "magnitude" of the response (in whatever form it may take). Different flaws of the same size and multiple inspections of the same flaw when inspected under absolutely identical conditions will produce different response magnitudes. Reducing the scatter in these response magnitudes is process optimization and leads to a smaller σ in the POD(a) function for that set of test conditions. When inspections of the same flaw are made for different inspection conditions, the magnitude of the inspection result will also vary, perhaps significantly. Since the different inspection conditions are all representative of the application, the effect of this variation may also be included in the capability experiment and its effect also shows up as an increase in σ. Reducing the scatter in response magnitudes that results from different test conditions is process variation reduction.

Inspection process optimization should have been performed prior to the evaluation experiment and, in fact, could have been accomplished using designed experiments as discussed herein. The optimization process leads to the definition of the test procedures (4.3.3) and provides the basis for demonstrating that the system is in a state of statistical control (4.3.4).

However, process optimization cannot be based on fixing all factors which might influence probability of detection. Some factors will inherently change during the application of the system. For example, apparently identical probes do produce different responses when applied to the same flaw and different inspectors do have different levels of proficiency at applying the inspection stimuli and interpreting the response. Probes and inspectors have their own POD(a) functions for the system and the scatter of these functions is the process variation. These latter types of factors should have been accounted for in the design of the evaluation experiment. If so, their effect on the POD function can be determined and, if significant, can indicate a direction for improving the process.

H.3.3 Analysis of date from one-factor-at-a-time experiments.

While the overall goal of an NDE demonstration is to describe the system capability with a single POD(a) relationship, it is often necessary to compare individual POD(a) curves. The implicit assumption in using a single curve to represent an entire NDE system is that the influences of system parameters such as inspector or probe are random and of the same order as system "noise" or random error. By comparing POD(a) curves, the hypothesis that the individual curves each represent the same NDE system capability can be tested statistically. Data can then be combined to produce a single POD(a) curve which represents the entire NDE system.

H.3.3.1 Comparing two POD(a) curves.

One of the useful properties of maximum likelihood estimators (cf. Appendix G.3.2.3), such as those describing the POD(a) relationship, is that they are asymptotically normally distributed as the sample size increases. These normal characteristics can be used to compare two POD(a) curves.

Let $\underset{\sim}{\overline{X}}_1 = (\mu_1, \hat{\sigma}_1)^T$ and $\underset{\sim}{\overline{X}}_2 = (\mu_2, \hat{\sigma}_2)^T$ be the estimated inspection behavior for curves 1 and 2 respectively.

APPENDIX H

If M_1 and M_2 are the true mean vectors, then the expected difference between

$\overline{X}_1$ and $\overline{X}_2$ is M_1 - M_2 and the expected value of the variance-covariance matrix is the sum of the individual covariances.

$$\text{Cov}\,(\underset{\sim}{\overline{X}}_1) + \text{Cov}\,(\underset{\sim}{\overline{X}}_2) = \underset{\sim}{\Sigma}_1 + \underset{\sim}{\Sigma}_2 \qquad \text{[H-1]}$$

By the central limit theorem

$$\left(\underset{\sim}{\overline{X}}_1 - \underset{\sim}{\overline{X}}_2\right) \sim N_p\left[\left(\underset{\sim}{M}_1 - \underset{\sim}{M}_2\right), \left(\underset{\sim}{\Sigma}_1 - \underset{\sim}{\Sigma}_2\right)\right] \qquad \text{[H-2]}$$

where N_p, indicates a *p*-variate normal population. Since there are 2 parameters in the POD model, $p = 2$. Under the null hypothesis, both POD curves represent the same (unknown) actual capability, $(\mu, \sigma)^T = .$

Thus, $\underset{\sim}{M}_1 = \underset{\sim}{M}_2 = \underset{\sim}{M}$

If the curves are similar, the statistical distance between them should be small. The squared statistical distance from $\left(\underset{\sim}{\overline{X}}_1 - \underset{\sim}{\overline{X}}_2\right)$ to ($\underset{\sim}{M}_1$ - $\underset{\sim}{M}_2$) = 0 is

$$T^2 = \left(\underset{\sim}{\overline{X}}_1 - \underset{\sim}{\overline{X}}_2\right)^T \left[\underset{\sim}{\Sigma}_1 + \underset{\sim}{\Sigma}_2\right]^{-1} \left(\underset{\sim}{\overline{X}}_1 - \underset{\sim}{\overline{X}}_2\right) \qquad \text{[H-3]}$$

which is analogous to the square of the *t* statistic in univariate analysis. When the sample size is large, T^2 has an approximate chi-square distribution with two degrees of freedom, x_2^2 .

Now, $\underset{\sim}{\Sigma}^{-1}$ is the inverse of the variance-covariance matrix of the model parameters μ and σ, and is called the Fisher information matrix. Further, the observed Fisher information is the negative of the matrix of second partial derivatives of the log likelihood function taken with respect to the model parameters, and so is computed as part of the maximum likelihood parameter estimation procedure.

To evaluate equation H-3, $\underset{\sim}{S}$ is computed for each curve by inverting its information matrix. The resulting two variance-covariance matrices are added, as in equation H-1, and the resulting matrix is inverted. This 2 x 2 matrix is then premultiplied by the 1 x 2 transpose of the matrix of differences between the parameters of curve 1 and curve 2, and postmultiplied by the 2 x 1 matrix of differences. The result of equation H-3 is then compared with the appropriate critical x^2 statistic, $x_2^2 = 5.99$, for a 95% confidence ellipse.

If $T^2 > x_2^2$ the null hypothesis is not supported by the data, and curves 1 and 2 would be considered statistically different.

Example

Table VI provides the $\underset{\sim}{\overline{X}}_1$, $\underset{\sim}{\overline{X}}_2$, $\underset{\sim}{\Sigma}_1$ and $\underset{\sim}{\Sigma}_2$, matrices for semi-automated eddy current inspections A1 and J3 in table V to illustrate the calculations comparing those two inspections. The T^2 test can be performed by any handheld calculator which supports matrix arithmetic; no special software is required.

T^2 for inspection J3 , (second operator, third inspection) was larger than the critical x^2 value of 5.99, and so differed significantly from test A1, the first inspection performed. All 10 inspection capabilities are plotted on figure 20, and J3 appears unlike the others.

Table VI. Calculation comparing inspection A1 with J3.

$$\underset{\sim}{\bar{X}}_{A1} = \begin{bmatrix} \log(0.004979) \\ 0.2693 \end{bmatrix} \qquad \underset{\sim}{\Sigma}_{A1} = \begin{bmatrix} 0.0102813 & -0.0014460 \\ -0.0014460 & 0.0017786 \end{bmatrix}$$

$$\underset{\sim}{\bar{X}}_{J3} = \begin{bmatrix} \log(0.005965) \\ 0.4664 \end{bmatrix} \qquad \underset{\sim}{\Sigma}_{J3} = \begin{bmatrix} 0.0026594 & -0.0069121 \\ -0.0069121 & 0.0080443 \end{bmatrix}$$

$$T^2 = \left[\underset{\sim}{\bar{X}}_{A1} - \underset{\sim}{\bar{X}}_{J3}\right]^T \left[\underset{\sim}{\Sigma}_{A1} + \underset{\sim}{\Sigma}_{J3}\right]^{-1} \left[\underset{\sim}{\bar{X}}_{A1} - \underset{\sim}{\bar{X}}_{J3}\right] T^2 = 24.78$$

$T^2 > x^{2(2,0.05)} = 5.99$ Reject H_0

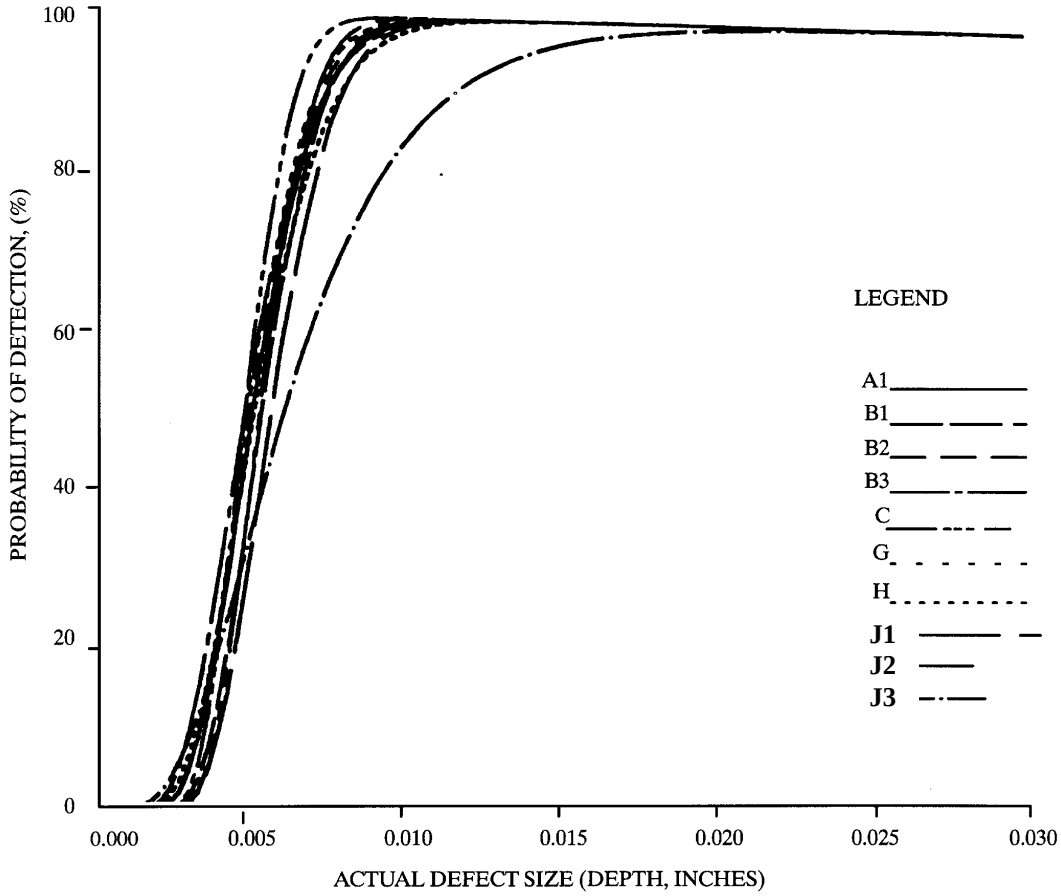


FIGURE 20. Composite plot for semi-automated inspections showing inspection J3 to be different.

H.3.3.2 Comparing many POD(a) curves.

The T^2 test compares one POD(a) relationship with another, and the preceding example compared inspection J3 to A1. The selection of A1 as the standard against which another inspection was compared was quite arbitrary. To avoid an arbitrary choice of a standard inspection, it is desirable to compare all POD(a) curves with each other simultaneously. Since there are two model parameters, μ and σ, the comparison may consider both parameters, and their possible interactive behavior.

This is accomplished by again exploiting the normal behavior of the model parameters and using a statistical procedure called Multivariate ANalysis Of VAriance, MANOVA. Although a thorough discussion is beyond the scope of this document and the arithmetic for its implementation is messy, the underlying idea is simple; compare the variation within the POD(a) relationships with the variation exhibited between inspections. This is done by taking the ratio of the magnitude of the within variation to the magnitude of the overall total (within plus between) variation.

The determinant of the variance-covariance matrix is called the generalized sample variance and is a convenient single value which summarizes the magnitude of the variation in S. So the magnitude of the variability within inspections, |W|, is the determinant of the sum of the covariance matrices of the model variability within inspections, |W|, is the determinant of the sum of the covariance matrices of the model parameters times the sample size (the number of specimens used to produce the individual POD(a) curves).

$$|W| = \left| n\left[\underset{\sim}{\Sigma}_1 + \underset{\sim}{\Sigma}_2 + \underset{\sim}{\Sigma}_2 + \ldots \underset{\sim}{\Sigma}_g\right] \right| \qquad \text{[H-4]}$$

where *g* is the number of groups, that is, the number of POD(a) curves being compared, and n is the number of specimens being inspected.

The multiplication by *n* converts $\underset{\sim}{\Sigma}$ from a matrix of mean squares and cross-products to one of summed squares and cross-products, SSC. It is the SSC which will be used in the test statistic, Λ^*, to be described later

The variability between inspections is estimated from the model parameters themselves as the sum of squares and cross-products.

$$\underset{\sim}{B} = \sum_{i=1}^{g} \left(\underset{\sim}{\overline{X}}_i - \underset{\sim}{\overline{\overline{X}}}\right)\left(\underset{\sim}{\overline{X}}_i - \underset{\sim}{\overline{\overline{X}}}\right)^T \qquad \text{[H-5]}$$

Where *g* is the number of groups and $\underset{\sim}{\overline{\overline{X}}}$ is the mean of the $\underset{\sim}{\overline{X}}$ vectors.
The magnitude of the total variability is the determinant or the sum of the within and between matrices: $|B + W|$.
The ratio of the magnitude of within variability to total variability is called Wilks's Lambda, Λ^*.

$$\Lambda^* = \frac{\left|\underset{\sim}{W}\right|}{\left|\underset{\sim}{B} + \underset{\sim}{W}\right|}$$

[H-6]

This text statistic is related to F for a two parameter model by

$$\left[\frac{N-g-1}{(g-1)}\right]\left[\frac{1-\sqrt{\Lambda^*}}{\sqrt{\Lambda^*}}\right] \sim F_{2(g-1)}, 2(N-g-1), \text{a} \qquad \text{[H-7]}$$

where *n* is the number of specimens, *g* is the number of groups, and N= *ng* is the total number of specimen inspections.

If Λ^* is too small, that is, if the total variation is large compared with individual variation, then the between-inspections variability cannot be explained by chance alone. If the differences cannot be explained by happenstance, the curves may be significantly different.

For a discussion of MANOVA and other related topics, see Johnson and Wichern, Applied Multivariate Statistical Analysis, 2nd ed., 1988, Prentice Hall.

APPENDIX H

Example

The inspections in table V were compared using a MANOVA, which showed them to differ significantly. Removing inspection J3 and performing a second MANOVA on the remaining nine inspections showed no difference among them. Inspection J3 is statistically different from the others. These results are summarized in table VII.

TABLE VII. Mean vectors and covariance matrices for inspections in Table V, Appendix G.

$$\underset{\sim}{\overline{X}}_{A1} = \begin{bmatrix} \log(0.004979) \\ 0.2693 \end{bmatrix} \qquad \underset{\sim}{\Sigma}_{A1} = \begin{bmatrix} 0.0102813 & -0.0014460 \\ -0.0014460 & -0.0017768 \end{bmatrix}$$

$$\underset{\sim}{\overline{X}}_{B1} = \begin{bmatrix} \log(0.005263) \\ 0.2344 \end{bmatrix} \qquad \underset{\sim}{\Sigma}_{B1} = \begin{bmatrix} 0.007634 & -0.0009093 \\ -0.0009093 & -0.0012713 \end{bmatrix}$$

$$\underset{\sim}{\overline{X}}_{B2} = \begin{bmatrix} \log(0.004893) \\ 0.2642 \end{bmatrix} \qquad \underset{\sim}{\Sigma}_{B2} = \begin{bmatrix} 0.00106600 & -0.0014810 \\ -0.0014810 & -0.0016570 \end{bmatrix}$$

$$\underset{\sim}{\overline{X}}_{B3} = \begin{bmatrix} \log(0.004732) \\ 0.30702 \end{bmatrix} \qquad \underset{\sim}{\Sigma}_{B3} = \begin{bmatrix} 0.0145000 & -0.0022900 \\ -0.0022900 & -0.0023640 \end{bmatrix}$$

$$\underset{\sim}{\overline{X}}_{C} = \begin{bmatrix} \log(0.004741) \\ 0.1968 \end{bmatrix} \qquad \underset{\sim}{\Sigma}_{C} = \begin{bmatrix} 0.068270 & -0.0006593 \\ -0.0006593 & -0.0009042 \end{bmatrix}$$

$$\underset{\sim}{\overline{X}}_{G} = \begin{bmatrix} \log(0.004843) \\ 0.2549 \end{bmatrix} \qquad \underset{\sim}{\Sigma}_{G} = \begin{bmatrix} 0.0100950 & -0.0012590 \\ -0.0012590 & -0.0015520 \end{bmatrix}$$

$$\underset{\sim}{\overline{X}}_{H} = \begin{bmatrix} \log(0.005031) \\ 0.3070 \end{bmatrix} \qquad \underset{\sim}{\Sigma}_{H} = \begin{bmatrix} 0.0013824 & -0.0020959 \\ -0.0020959 & -0.0024362 \end{bmatrix}$$

$$\underset{\sim}{\overline{X}}_{J1} = \begin{bmatrix} \log(0.005567) \\ 0.2380 \end{bmatrix} \qquad \underset{\sim}{\Sigma}_{J1} = \begin{bmatrix} 0.0007952 & -0.000884 \\ -0.0008884 & -0.0013570 \end{bmatrix}$$

$$\underset{\sim}{\overline{X}}_{J2} = \begin{bmatrix} \log(0.005202) \\ 0.2012 \end{bmatrix} \qquad \underset{\sim}{\Sigma}_{J2} = \begin{bmatrix} 0.0063446 & -0.0006381 \\ -0.0006381 & -0.0009398 \end{bmatrix}$$

$$\underset{\sim}{\overline{X}}_{J3} = \begin{bmatrix} \log(0.005965) \\ 0.4664 \end{bmatrix} \qquad \underset{\sim}{\Sigma}_{J3} = \begin{bmatrix} 0.0026594 & -0.0069121 \\ -0.0069121 & -0.0080443 \end{bmatrix}$$

TABLE VIII. One-way MANOVA comparing 10 inspections in Table V, Appendix G.

$$W = \begin{bmatrix} 2.11820 & -0.55737 \\ -0.55737 & 0.66913 \end{bmatrix} \qquad B = \begin{bmatrix} 0.04966 & 0.02852 \\ 0.02852 & 0.05380 \end{bmatrix}$$

$$\text{Wilks } \Lambda^* = \frac{\left|\underset{\sim}{W}\right|}{\left|\underset{\sim}{B}+\underset{\sim}{W}\right|} = 0.8595$$

$$\left[\frac{300-10-1}{(10-1)}\right]\left[\frac{1-\sqrt{\Lambda^*}}{\sqrt{\Lambda^*}}\right] = 2.525$$

$F_{18,\infty.01} = 1.94$ Reject H_0

TABLE IX. One-way MANOVA excluding inspection J3 in Table V, Appendix G.

$$W = \begin{bmatrix} 2.03842 & -0.035001 \\ -0.035001 & 0.042781 \end{bmatrix} \qquad B = \begin{bmatrix} 0.02298 & -0.00462 \\ -0.00462 & 0.01264 \end{bmatrix}$$

$$\text{Wilks } \Lambda^* = \frac{\left|\underset{\sim}{W}\right|}{\left|\underset{\sim}{B}+\underset{\sim}{W}\right|} = 0.9583$$

$$\left[\frac{270-9-1}{(9-1)}\right]\left[\frac{1-\sqrt{\Lambda^*}}{\sqrt{\Lambda^*}}\right] = 0.700$$

$F_{16,\infty.01} = 2.01$ Do **Not** Reject H_0

H.3.4 Analysis of date from factorial experiments.

The statistical tests discussed in the previous section may indicate that performance of a particular inspection differs from those against which it is being compared. They do not, however, provide specific information as to the cause of the difference. To do this, the overall observed variance may be partitioned into its constitutive components. The resulting analysis will then permit assignment of causes for differing NDE capabilities, and thus allow for remedial action. It may be noted that, in general, the components of variance cannot be determined unless the experiment was planned to accomplish this. It is very important, therefore, that proper consideration be given to this goal before any experimentation is carried out, and before any data are collected. See 4.3, Demonstration design, and Appendix E, Test Program Guidelines. The methods discussed previously were developed to compare inspection systems using data not specifically gathered for that purpose. A designed experiment can provide more engineering information from a given number of tests than is available from the one-factor-at-a-time data presented in tables VII, VIII, and IX. The following sections describe methods which can be used with data from a statistically designed experiment.

H.3.4.1 Factorial experimental design.

In any NDE demonstration there will be a certain amount of variation from inspection to inspection. With the proper demonstration design, this variation can be partitioned into components of variance, each component being assignable to a specific cause, or factor. In some instances, interactions among the factors influencing NDE capability can also be identified. Furthermore, the resulting estimates of the model parameters $\hat{\mu}$ and $\hat{\sigma}$, will be more precise because they are based on the average behavior of several inspections. These types of demonstration designs are called Factorial Designs, because they can identify the factors causing (nonrandom) variation.

Example

The â vs. a data in table X were part of a demonstration designed to assess the influence on POD of different operators, different probes, and different positions of the piece being inspected using a semi-automated ET system. Data in table X and similar data for eight other inspections were analyzed using the maximum likelihood parameter estimation procedure described in this document.

The NDE demonstration was a factorial test to evaluate the influence on POD(a) of three different OPerators (OP), three PRobes (PR), and two POSitions (POS) of the workpiece being inspected. Results are summarized in table XI.

APPENDIX H

TABLE X. â vs a data for web/bore surface flaws, semi-automated inspection.

a	â	a	ä	a	â
0.001	(1.0)	0.009	1.60	0.015	10.10
0.003	(1.0)	0.009	4.40	0.016	11.00
0.003	(1.0)	0.010	5.10	0.019	15.00
0.006	3.800	0.010	6.60	0.022	22.00
0.007	3.000	0.011	6.00	0.029	29.00
0.007	2.900	0.011	8.40	0.031	38.00
0.008	3.900	0.012	5.80	0.042	31.00
0.008	3.600	0.013	57.40	0.065	49.00
0.009	2.200	0.014	2.20	0.100	80.30

Notes:
1. a is crack size in inches
2. â is apparent size (see text)
3. *,** censored observations:
   * unknown, below $\hat{a}_{th}$ = 1.0
   ** unknown, above $\hat{a}_{sat}$ = 20.0

TABLE XI. Model parameters for semi-automated inspections.

OP	PR	POS	A_{50}	σ	α	β	δ	n_1	n_2	n_3
1	1	1	0.00326130	0.235297	8.0673	1.4090	0.33153	25	3	0
1	2	1	0.00335512	0.260288	8.0807	1.4184	0.36918	26	3	0
1	3	2	0.00337838	0.201442	8.2139	1.4435	0.29078	25	3	0
2	1	2	0.00335999	0.400897	7.9109	1.3889	0.55680	24	4	0
2	2	1	0.00354285	0.393517	8.1534	1.4449	0.56860	24	4	0
2	3	1	0.00339956	0.399634	8.0139	1.4099	0.56343	24	4	0
3	1	1	0.00302999	0.233559	7.9871	1.3773	0.65326	25	3	0
3	2	2	0.00336885	0.331408	7.8785	1.3839	0.45862	25	3	0
3	3	1	0.00337758	0.260116	8.1646	1.4348	0.34904	25	3	0

Notes:
1. a_{50}, crack size at 50% POD
2. n_1 = total observations, n_2 = data in noise, n_3 = saturations

H.3.4.2 Effect of NDE process parameters on μ and σ individually.

The methods presented here can be used to compare POD(a) relationships which result from either â vs a data, or hit / miss data. They are straightforward applications of well known statistical procedures and can be performed by many commercially available statistical packages.

Often a quick comparison of the individual model parameters, considered separately, is informative. An ANalysis Of VAriance, ANOVA, is performed which considers only one model parameter at a time.

The statistical ANOVA model is $y = \bar{y} + OP_i + PR_j + POS_k + \varepsilon_{ijk}$, where y is the model parameter (either μ and σ) being evaluated, and $\bar{y}$ is average parameter response, and

i= 1.. I, the number of operators

j= 1...J, the number of probes

k= 1...k, the number positions, and

e_{ijk} is the random error.

The experiment has been designed so that an unambiguous test can be performed to determine if a difference between operators, between probes, or between positions is statistically significant. The test used is an F test. The statistic has the form $F = s_1^2 / s_2^2$ where s_1^2 and s_1^2 are two independent mean squares. This method assumes that the data comes from a normal distribution. Since μ and σ are MLE's, this is a reasonable assumption. This assumption is necessary particularly for small sample sizes.

The F statistic is used to test hypothesis of the form $H_0 : \sigma_1^2 = \sigma_2^2$. That is, is the variance attributed to a specific cause equal to the variance due to random causes. If σ_1^2 is greater than σ_2^2, then the variation in the response between the levels of a factor (eg: operator, position, or probe) is greater than the experimental error. The ratio of estimates of these two components, F, should be approximately equal to one if the hypothesis is true, and greater than one if the data do not support the hypothesis.

TABLE XII. Analysis of variance table.

Source	Df	SS	MS	*F*
OP	I-1	$JK\sum_{i=1}^{I}\left(\bar{y}_{i\ldots}-\bar{y}_{\ldots}\right)^2$	$S_1^2 = ss_{OP}/df_{op}$	S_1^2/S^2
PR	J-1	$IK\sum_{j=1}^{J}\left(\bar{y}_{.j.}-\bar{y}_{\ldots}\right)^2$	$S_2^2 = SS_{PR}/df_{pr}$	S_2^2/S^2
POS	K-1	$IJ\sum_{k=1}^{K}\left(\bar{y}_{..k.}-\bar{y}_{\ldots}\right)^2$	$S_3^2 = SS_{POS}/df_{pos}$	S_3^2/S^2
Error	subtract	subtract	S^2	
Total	IJK-1	$\Sigma\Sigma\Sigma\left(y_{ijk}-\bar{y}_{\ldots}\right)^2$		

Example

Using the data in table XI, the ANOVA for μ is

TABLE XIII. ANOVA for model parameter μ.

Source	D*F*	Type III SS	*F* Value	Prob > *F*
OP	2	0.00000005	2.36	0.24253
PR	2	0.00000007	3.63	0.15803
POS	1	0.00000000	0.35	0.59767

As *F* increases, *p* decreases. The larger the differences between levels in a factor, the larger the value of *F*. The larger the F, the greater the incredibility associated with $H_0 : \sigma_1^2 = \sigma_2^2$. A measure of this incredibility is the probability that an *F* as large as the observed *F* could have occurred if H_0 were true. This probability is called a *p*-value associated with the observed *F*, in practice, *p*-values of $p = 0.10$ or $p = 0.05$ are considered significant. In table XIII, PRobe is the most significant variable although it is not statistically significant at the usual confidence levels (10%, 5%, or 1%).

The ANOVA for σ is:

TABLE XIV. ANOVA for model parameter, σ.

Source	DF	Type III SS	*F* Value	Prob > *F*
OP	2	0.04439593	20.21	0.01817
PR	2	0.00319839	1.46	0.36154
POS	1	0.00040217	0.37	0.58785

Here the *p*-value for operators is $p = 0.01817$ indicating a statistically significant difference in the levels of operator.

H.3.4.3 Analysis of the means.

To perform the ANOVA , the mean was calculated for each level of each variable. Once a significant difference has been detected by the ANOVA , the average values for each level of a factor (the mean) are examined. These values are examined to determine the magnitude of the difference between them and to determine if a variable which is statistically significant is practically significant. For example, it may be that a difference in m is statistically significant, but upon examining the average values it is found that the largest difference between the averages is 0.001. Although this difference is statistically significant, it is not practical to differentiate to the 0.001 level. Also, large differences which are not statistically significant should be investigated. It should be determined if the lack of significance is due to having not included a significant variable in the experiment or if the sample size for the experiment was not large enough.

Example

Table XV summarizes the analysis of means for the example used throughout H.3.4. Given parameters are the variable, the level of the variable, and the model parameter of interest (either μ or σ). Here, a statistically significant difference (DIFF) is represented for a group by a different letter of the alphabet.

TABLE XV. Analysis of means.

OP	μ	DIFF	σ	DIFF
1	0.00333	A	0.23234	B
2	0.00344	A	0.39802	A
3	0.00326	A	0.27503	B
PR				
1	0.00322	A	0.28992	A
2	0.00342	A	0.32840	A
3	0.00339	A	0.28706	A
POS				
1	0.00332	A	0.29707	A
2	0.00337	A	0.31125	A

The means indicate that there is only one significant difference: that due to OP for the parameter σ. Remember that this test is done at an $\alpha = 0.05$ level of significance. It may be that a more, or less, strict level is required.

H.3.4.4 Effect of NDE process parameters on μ and σ jointly.

Data from factorial designs can be analyzed using a MANOVA procedure similar to the one described in H.3.3.2 . However, there is a fundamental difference. In the one-factor-at-a-time data it was possible only to conclude that all ten inspections were not the same; no further breakdown as to the influence of operator, eddy current probe, experimental setup, or other cause, was possible. With factorial design, the data are balanced so that the influence of each factor can be identified by its contribution to the total sum of squares (a sort of statistical distance between an individual observation and the average for that condition). The MANOVA procedure is available in many commercially available statistical analysis software packages.

APPENDIX H

A MANOVA simultaneously compares the variation in both model parameters, μ and σ, which results from a given factor (or combination of factors) with the random variation observed in the inspection system. This random, or error, component of variance can be estimated from the variance-covariance structure of the data. The analysis can be greatly simplified, however, by using instead the variation attributed to the highest order interaction. For example, the interaction among operator, probe, and position of the workpiece. It is unlikely that this interaction would be as influential as the main effects (eg: operator, probe, position, by themselves) or as the second order interactions (eg: operator-probe, operator-position, probe-position). Confounding this third order interaction with random error greatly simplifies the subsequent MANOVA because the individual variance-covariance matrices would not have to be evaluated as part of the analysis. Even with a packaged program, keying in many large variance-covariance matrices is tedious work. The simplified procedure requires only the model parameters themselves and that they have resulted from a factorial NDE demonstration design.

Example

A multivariate analysis of variance (MANOVA) was performed on the data resulting from the factorial design summarized in Table XI. Wilks's Lambda was computed as the criterion, and an *F* test was performed.

TABLE XVI. MANOVA for model parameters, μ and σ (H.3.4.4)

Factor	*F*	*p*
OP	**3.88**	**0.10868**
PR	**1.40**	**0.37674**
POS	**0.20**	**0.83561**

Overall operator has an effect on the POD with both μ and σ considered simultaneously. Changing μ moves the POD curve horizontally. Changing σ varies the shape of the curve, but not its central location. The MANOVA calculations test if these combined effects are significant in showing a difference among operators, probes, or positions.

H.3.4.5 Components of variation.

The components of variation can be decomposed into variation due to each factor (OP, PR, POS, error) . Basically, the mean square for each factor is not an expression of variance for that factor alone, but is a function of that factor and possibly other factors.

The components of variation in μ and σ for each factor can be found by substituting the estimate of error V(error) = 0.00326027 and setting each equal to its EMS value. Table XVII illustrates these calculations for this example.

TABLE XVII. MANOVA for model parameters, μ and σ (H3.4.5)

Source	Type III Expected Mean Square
OP	V (error) + 3V (OP)
PR	V (error) + 3V (PR)
POS	V (error) + 4V (POS)

Sometimes negative components of variance occur due to rounding or general lack of significance of any variable. In this case the components are set equal to zero.

APPENDIX J

EXAMPLE DATA REPORTS

J.1. SCOPE

J.1.1 Scope.

This appendix presents sample data sheets for reporting test matrices and the results of individual inspections. Examples of summary results are also included for reference.

J.2. APPLICABLE DOCUMENTS.

This section is not applicable to this appendix.

J.3. PROCEDURES

J.3.1 Test matrix

Figures 21 and 22 are examples of two methods for summarizing the description of a capability evaluation test matrix. For this example it was assumed that the assessment of an ET system was to include the effects of two operators, two probes, and two replications. Figure 21 is essentially a list of the combinations of the levels of the test matrix. Figure 22 is a table of the test factor combinations and shows the levels of all of the factors being evaluated. Although figure 22 more clearly displays the experimental design, this format becomes unwieldy if the experiment contains more than four factors or more than three levels of the factors.

J.3.2 Individual test results

Figure 23 is an example data sheet for a permanent record of the individual test results of an NDE evaluation. The results from each inspection of the specimen set under a defined set of conditions are presented in the column for the specific test.

J.3.3 Anaylsis results

Figures 24 and 25 present examples of the â vs. a and hit / miss analyses, respectively. In both of the examples, the analysis provided complete sets of parameter estimates.

Examples of the plots required in the results summary are presented on figures 26 through 29. The POD(a) functions with 95 percent confidence limits for the analyses of figures 24 and 25 are presented on figures 26 and 27, respectively. These figures illustrate the minimum information that may be included on all plots of the POD(a) function. Figure 28 presents the log â vs. log a plot for the analysis of figures 24 and 26. The POD(a) function and the observed detections for the hit / miss analysis of figures 25 and 27 are presented on figure 29.

APPENDIX J

EXPERIMENTAL DESIGN DATA SHEET

DATE: _____ EXPERIMENT ID NUMBER: _____
NDE SYSTEM: _____ SPECIMEN SET: _____

ORGANIZATION: _____

OBJECTIVE: To evaluate Station 1 of the RFC system for two randomly selected operators, probes, and replications in a complete factorial experiment

Test Identification	Operator Number	Probe Number	Replication Number
111	1	1	1
112	1	1	2
121	1	2	1
122	1	2	2
211	2	1	1
212	2	1	2
221	2	2	1
222	2	2	2

Randomization: The eight sets of inspections were conducted in a random order.

FIGURE 21. Example data sheet for describing the experimental design - list format.

APPENDIX J

EXPERIMENTAL DESIGN DATA SHEET

DATE: _____ EXPERIMENT ID NUMBER_____
NDE SYSTEM:_____ SPECIMEN SET:_____

ORGANIZATION:_____

OBJECTIVE: To evaluate Station 1 of the RFC system for two randomly selected operators, probes and replications in a complete factorial experiment

Table of Test Identification Numbers

	Operator 1	Operator 2
Probe 1 - Rep 1	111	211
- Rep 2	112	212
Probe 1- Rep 1	121	221
- Rep 2	122	222

Randomization: The eight sets of inspections were conducted in a random order.

FIGURE 22. Example data sheet for describing the experimental design - table format.

APPENDIX J

TEST RESULT DATA SHEET

Page ______ of _______

DATE: ________________________________ EXPERIMENT ID NUMBER: ______________________
NDE SYSTEM: _____________________________________ SPECIMEN SET: ______________________
ORGANIZATION: ____________________________

Crack ID	Crack Size		Test Identification						
	Length	Depth							
Number of Extra Indications									

FIGURE 23. Example data sheet for test results.

APPENDIX J

AHAT VS A POD ANALYSIS
VERSION 2.3b

DATE: 30 JUL 90

IDENTIFICATION: FILE = RFC2WBIN.DAT
DATA SET = WBIN100
INSPECTIONS = A C D

REGRESSION ANALYSIS

MODEL: LN(AHAT) = B0+B1*LN(A)

CRACK SIZE RANGE: 1.00 TO 100.

NUMBER OF UNCENSORED CRACKS: 25

RECORDING THRESHOLD: 70. NUMBER OF CRACKS BELOW THRESHOLD: 2

SATURATION LEVEL; 4095. NUMBER OF CRACKS AT SATURATION: 1

PARAMETER ESTIMATES

PARAMETER	ESTIMATE	SE
INTERCEPT(B0) -	3.06	0.300
SLOPE(B1) -	1.44	0.116
RESIDUAL ERROR -	0.417	0.593E-01

REPEATABILITY ERROR: 0.268

POD MODEL PARAMETER ESTIMATES

SIGMA: 0.328

INSPECTION

THRESHOLD	A50	A90	A90/95	V11	V12	V22
70.0	2.29	3.48	4.65	0.212E-01	-0.325E-02	0.193E-02
100.	2.93	4.46	5.79	0.162E-01	-0.276E-02	0.193E-02
200.	4.74	7.22	8.99	0.888E-02	-0.180E-02	0.193E-02
270.	5.84	8.89	11.0	0.665E-02	-0.139E-02	0.193E-02
300.	6.29	9.57	11.8	0.599E-02	-0.124E-02	0.193E-02
350.	7.00	10.7	13.1	0.517E-02	-0.103E-02	0.193E-02

FIGURE 24. â vs a analysis.

APPENDIX J

HIT/MISS POD ANALYSIS
LOGNORMAL MODEL
VERSION 2.3

DATE: 30-JUL-90

IDENTIFICATION: FILE = PADMOD.PF
DATA SET = SET2FPI
INSPECTIONS = 1 2 3 6 9

NUMBER OF VALID CASES: 36

CRACK SIZE RANGE: 8.0 TO 275.0

THRESHOLD: 0.5

MAXIMUM LIKELIHOOD ESTIMATES:

MU-HAT = 4.62 SIGMA-HAT = 0.630

PERCENTILE ESTIMATES:

A50 = 101. A90/50 = 227. A90/95 = 0.730E+04

ESTIMATED VARIANCE/COVARIANCE MATRIX OF THE
MAXIMUM LIKELIHOOD ESTIMATES:

	MU-HAT	SIGMA-HAT
MU-HAT	0.286E-01	0.466E-02
SIGMA-HAT	0.466E-02	0.483E-01

FIGURE 25. Hit/miss analysis.

APPENDIX J

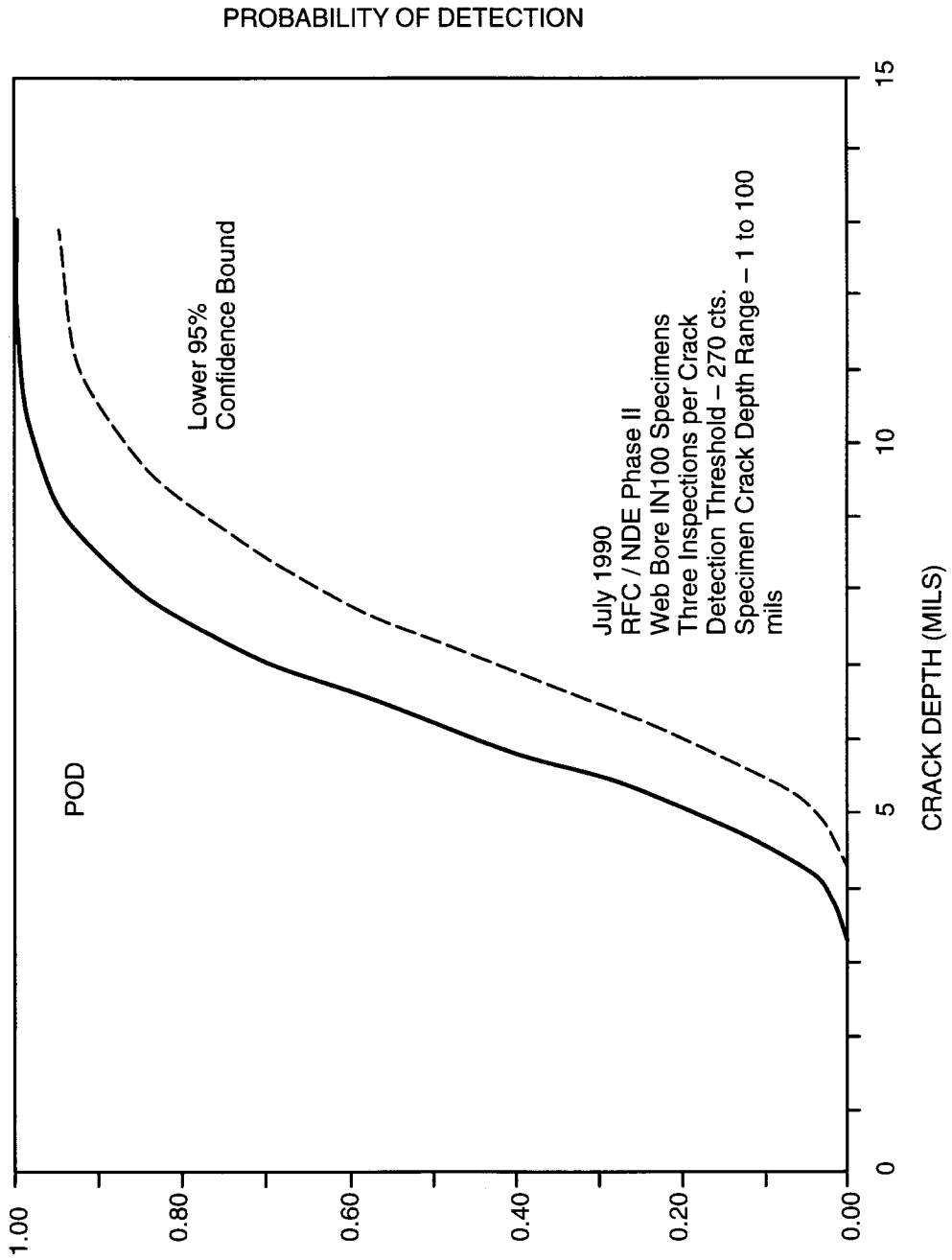


FIGURE 26. POD(a) for â vs a analysis.

APPENDIX J

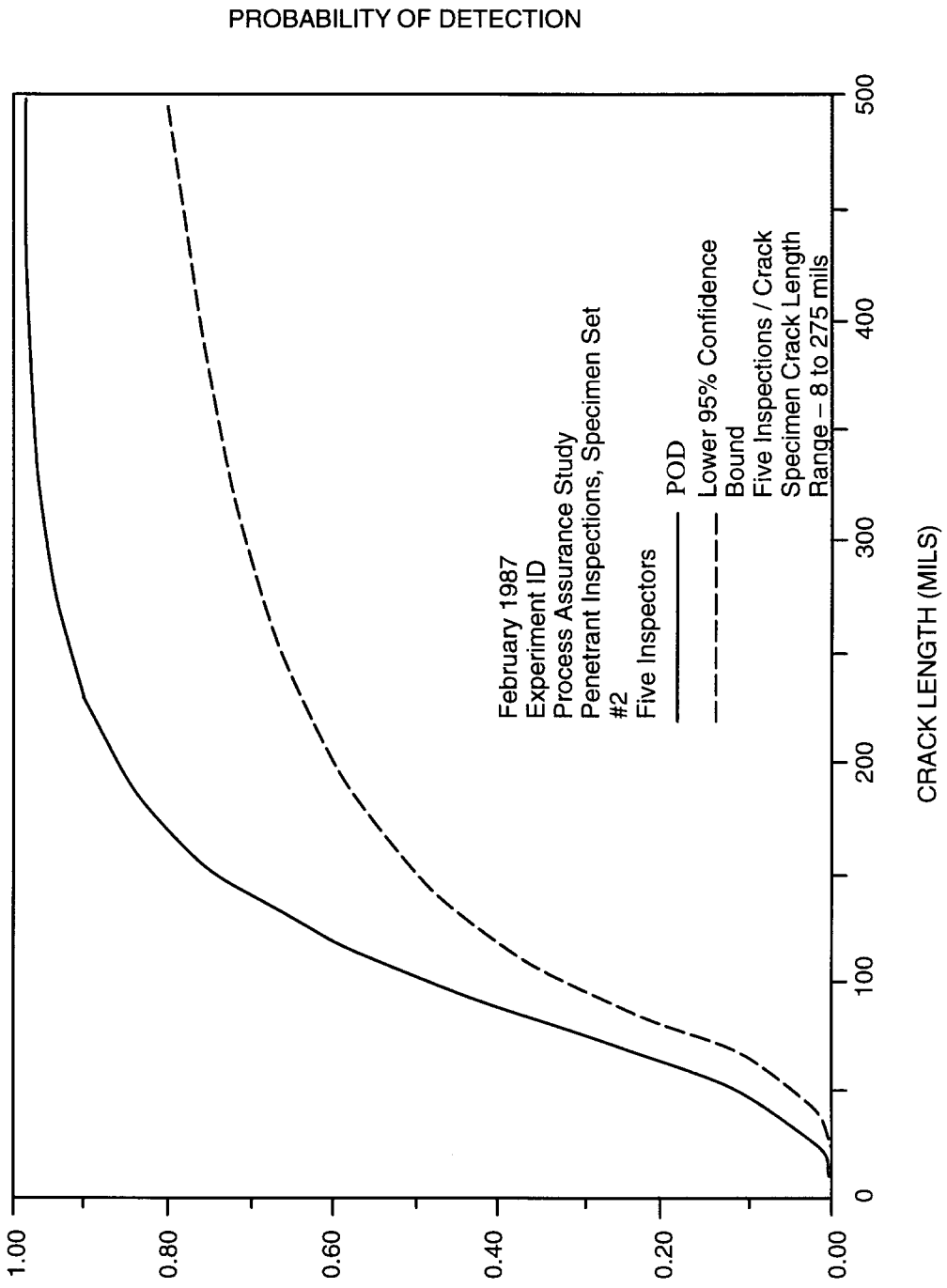


FIGURE 27. POD(a) for hit/miss analysis.

APPENDIX J

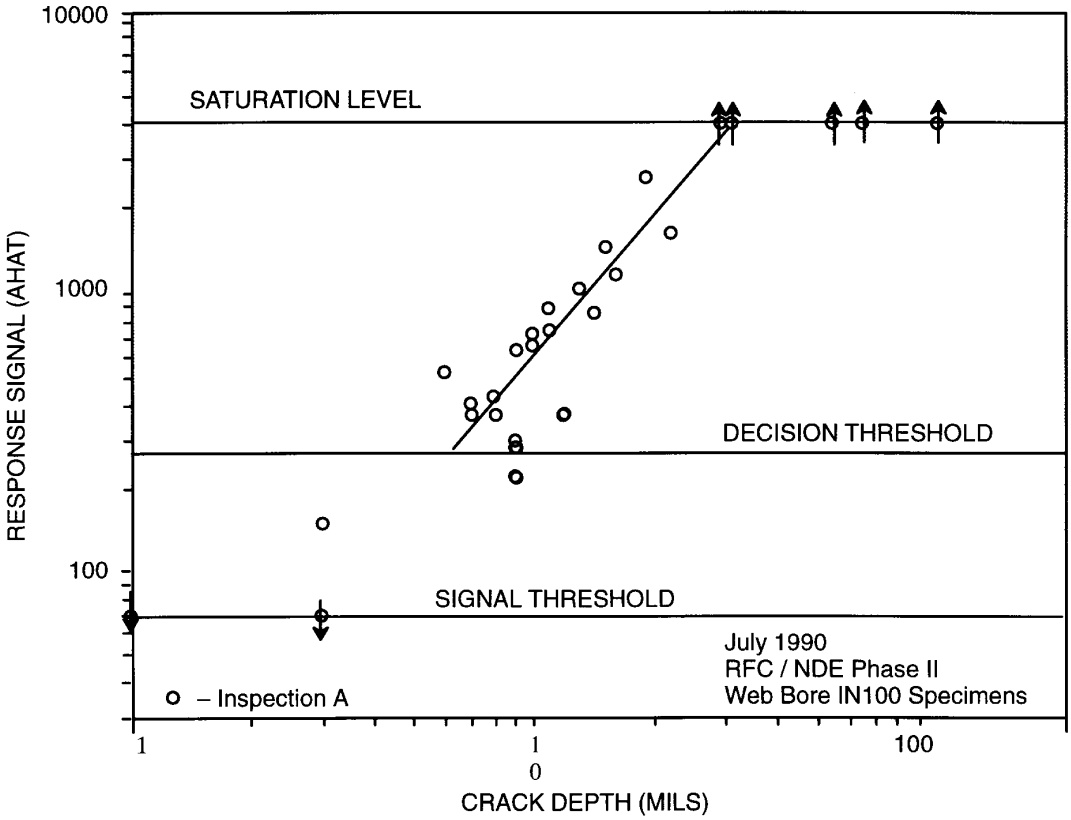


FIGURE 28. Log â vs log a for â vs a analysis.

APPENDIX J

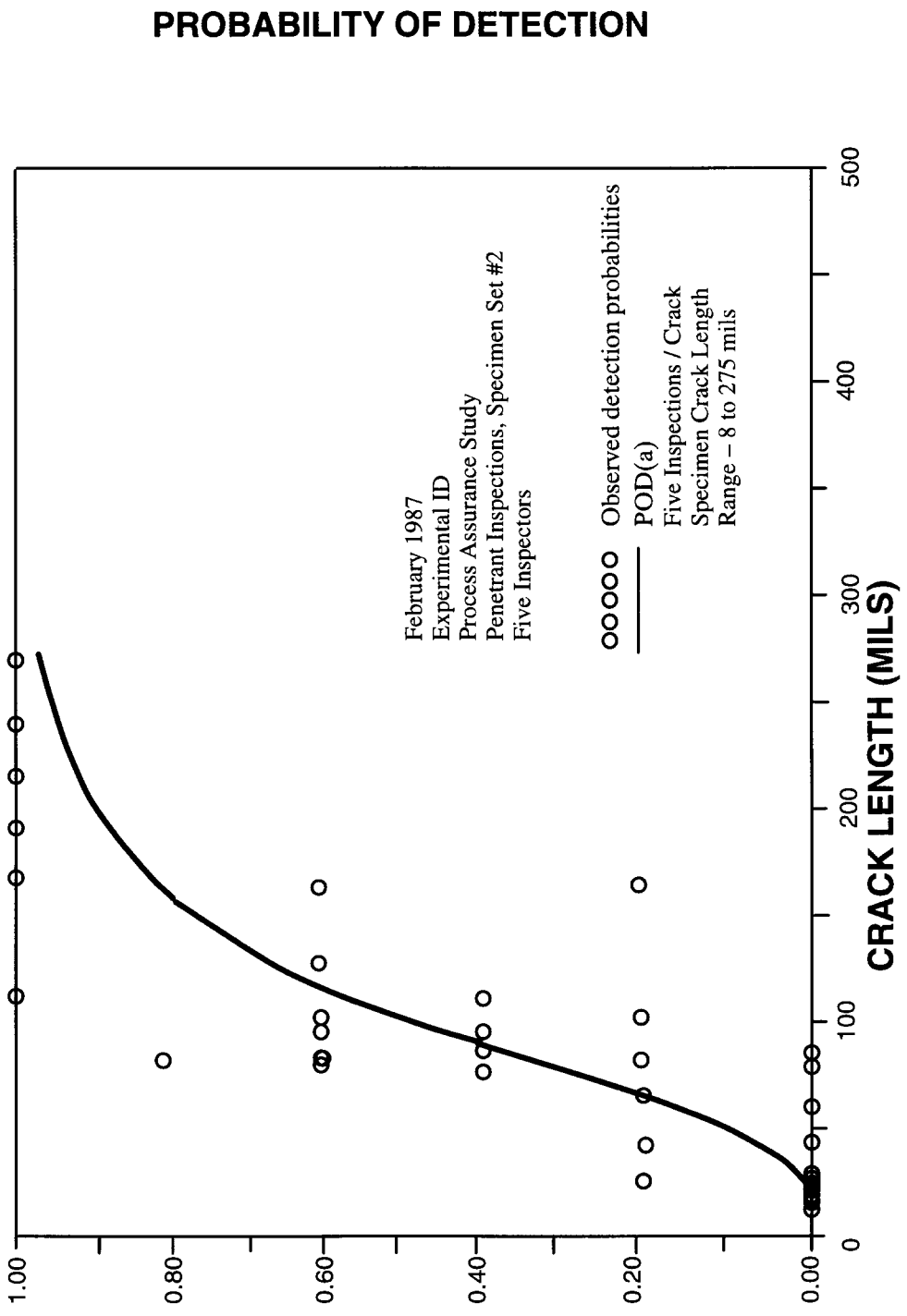


FIGURE 29. Observed detections and POD for hit/miss analysis.

APPENDIX J

Custodian:
Army-MR
Navy-AS
Air Force - 11

Review activities:
Army- SH
Air Force - 99

Preparing Activity:
Air Force - 11

(Project NDTI-0221)

STANDARDIZATION DOCUMENT IMPROVEMENT PROPOSAL

INSTRUCTIONS

1. The preparing activity may complete blocks 1, 2, 3, and 8. In block 1, both the document number and revision letter should be given.
2. The submitter of this form may complete blocks 4, 5, 6, and 7.
3. The preparing activity may provide a reply within 30 days from receipt of the form.
   NOTE: This form may not be used to request copies of documents, nor to request waivers, or clarification of requirements on current contracts. Comments submitted on this form do not constitute or imply authorization to waiver any portion of the referenced document(s) or to amend contractual requirements.

I RECOMMEND A CHANGE:	1. DOCUMENT NUMBER MIL-HDBK-1823	2. DOCUMENT DATE (YYMMDD) 990430

3. DOCUMENT TITLE
NONDESTRUCTIVE EVALUATION SYSTEM RELIABILITY ASSESSMENT

4. NATURE OF CHANGE (Identify paragraph number and include proposed rewrite, if possible. Attach extra sheets as needed.)

5. REASON FOR RECOMMENDATION

6. SUBMITTER

a. NAME (Last, Middle Initial)	b. ORGANIZATION	
c. ADDRESS (include Zip Code)	d. TELEPHONE (Include Area Code) (1) Commercial (2) AUTOVON *(If applicable)*	e. DATE SUBMITTED *(YYMMDD)*

8. PREPARING ACTIVITY

a. NAME	b. TELEPHONE (Include Area Code	
ASC/ENSI (AF-11)	(1) Commercial (937)255-6281	(2) AUTOVON 785-6281
c. ADDRESS *(Include Zip Code)* BLDG 560 2530 LOOP ROAD W WRIGHT-PATTERSON AFB OH 45433-7101	IF YOU DO NOT RECEIVE A REPLY WITHIN 45 DAYS, CONTACT: Defense Quality and Standardization Office 5203 Leesburg Pike, Suite 1403, Falls Church, VA 22041-3466 Telephone (703) 756-2340 AUTOVON 289-2340	