NUREG/CR-6227
PNNL-9433

Performance Demonstration Tests for Eddy Current Inspection of Steam Generator Tubing

Prepared by
R. J. Kurtz, P. G. Heasler, C. M. Anderson

Pacific Northwest National Laboratory

Prepared for
U.S. Nuclear Regulatory Commission

AVAILABILITY NOTICE

Availability of Reference Materials Cited in NRC Publications

Most documents cited in NRC publications will be available from one of the following sources:

1. The NRC Public Document Room, 2120 L Street, NW., Lower Level, Washington, DC 20555-0001
2. The Superintendent of Documents, U.S. Government Printing Office, P. O. Box 37082, Washington, DC 20402-9328
3. The National Technical Information Service, Springfield, VA 22161-0002

Although the listing that follows represents the majority of documents cited in NRC publications, it is not intended to be exhaustive.

Referenced documents available for inspection and copying for a fee from the NRC Public Document Room include NRC correspondence and internal NRC memoranda; NRC bulletins, circulars, information notices, inspection and investigation notices; licensee event reports; vendor reports and correspondence; Commission papers; and applicant and licensee documents and correspondence.

The following documents in the NUREG series are available for purchase from the Government Printing Office: formal NRC staff and contractor reports, NRC-sponsored conference proceedings, international agreement reports, grantee reports, and NRC booklets and brochures. Also available are regulatory guides, NRC regulations in the *Code of Federal Regulations,* and *Nuclear Regulatory Commission Issuances.*

Documents available from the National Technical Information Service include NUREG-series reports and technical reports prepared by other Federal agencies and reports prepared by the Atomic Energy Commission, forerunner agency to the Nuclear Regulatory Commission.

Documents available from public and special technical libraries include all open literature items, such as books, journal articles, and transactions. *Federal Register* notices, Federal and State legislation, and congressional reports can usually be obtained from these libraries.

Documents such as theses, dissertations, foreign reports and translations, and non-NRC conference proceedings are available for purchase from the organization sponsoring the publication cited.

Single copies of NRC draft reports are available free, to the extent of supply, upon written request to the Office of Administration, Distribution and Mail Services Section, U.S. Nuclear Regulatory Commission, Washington, DC 20555-0001.

Copies of industry codes and standards used in a substantive manner in the NRC regulatory process are maintained at the NRC Library, Two White Flint North, 11545 Rockville Pike, Rockville, MD 20852-2738, for use by the public. Codes and standards are usually copyrighted and may be purchased from the originating organization or, if they are American National Standards, from the American National Standards Institute, 1430 Broadway, New York, NY 10018-3308.

DISCLAIMER

NUREG/CR-6227
PNNL-9433

Performance Demonstration Tests for Eddy Current Inspection of Steam Generator Tubing

Manuscript Completed: June 1995
Date Published: May 1996

Prepared by
R. J. Kurtz, P. G. Heasler, C. M. Anderson

Pacific Northwest National Laboratory
Richland, WA 99352

J. Muscara, NRC Project Manager

Prepared for
Division of Engineering Technology
Office of Nuclear Regulatory Research
U.S. Nuclear Regulatory Commission
Washington, DC 20555-0001
NRC Job Code B2097

Abstract

This report describes the methodology and results for development of performance demonstration tests for eddy current (ET) inspection of steam generator tubes. Statistical test design principles were used to develop the performance demonstration tests. Thresholds on ET system inspection performance were selected to ensure that field inspection systems would have a high probability of detecting and correctly sizing tube degradation. The technical basis for the ET system performance thresholds is presented in detail. Statistical test design calculations for probability of detection and flaw sizing tests are described. A recommended performance demonstration test based on the design calculations is presented. A computer program for grading the probability of detection portion of the performance demonstration test is given.

Contents

Figures

Tables

Executive Summary

Statistically-based performance demonstration qualification tests have been developed for eddy current (ET) inspection of steam generator tubing. The purpose of the performance demonstration tests is to provide a mechanism for determining which ET inspection systems conducting inspections in the field can reliably detect and size all of the known forms of tube damage that occur in operating steam generators. For this work the goal of steam generator tube inservice inspection was to identify most or all of the defective tubes which could fail during reactor operation. An extensive data base on the failure pressure of degraded steam generator tubes as a function of flaw type and size was utilized to define a defective tube as one with degradation ≥ 75% through-wall. Information from a study on the reliability of ET systems to detect and size service-induced tube degradation, coupled with results from an effort to develop and evaluate the effectiveness of inservice inspection sampling plans was used to establish thresholds on probability of detection (POD) performance, flaw sizing accuracy, and false call rate. The performance thresholds were selected so that a passing ET system would have a 90% probability of detecting and plugging a defective tube, assuming the tube was inspected, and the plugging limit was 40% through-wall degradation as measured by the ET system. Thresholds were also established for degraded, but not defective tubes, because current and proposed inservice inspection sampling plans rely on detection of degraded tubes to trigger additional inspection, and to call attention to conditions which may require corrective actions to mitigate additional tube damage. The POD, flaw sizing, and false call rate performance thresholds were used in statistical test design calculations to determine the appropriate number and size distribution of flawed steam generator tube samples that would be needed in a steam generator tube bundle mockup to ensure reliable ET inspection system performance. Binomial calculations and Monte Carlo simulations were performed for mockups containing different numbers of flawed tube samples to determine the probability of an acceptable ET system failing the test and conversely, the probability of an unacceptable ET system passing the test. Limits were established so that ET systems possessing acceptable performance characteristics would pass the test a high percentage of the time but ET systems with unacceptable performance characteristics would have only a small chance (≤ 5%) of passing. For the POD test a mockup of 360 flawed tube samples composed of all relevant flaw types would be needed to meet the performance goals selected. For the flaw sizing test only about 170 flawed tube samples would be necessary to establish acceptable sizing performance.

Previous Reports in Series

Allen, R. P., R. L. Clark, and W. D. Reece. 1984. "Steam Generator Group Project Task 6 - Channel Head Decontamination." NUREG/CR-3841, PNL-4712, Pacific Northwest Laboratory, Richland, Washington.

Alzheimer, J. M., R. A. Clark, W. E. Anderson, J. C. Crowe, and M. Vagins. 1977. "Steam Generator Tube Integrity Program. Quarterly Report, January 1-March 31, 1977." NUREG-0359. Pacific Northwest Laboratory, Richland, Washington.

Alzheimer, J. M., R. A. Clark, C. J. Morris and M. Vagins. 1979. "Steam Generator Tube Integrity Program Phase I Report." NUREG/CR-0718, PNL-2937, Pacific Northwest Laboratory, Richland, Washington.

Bickford, R. L., R. A. Clark, P. G. Doctor, H. E. Kjarma and C. A. LoPresti. 1984. "Eddy Current Round Robin Test on Laboratory Produced Intergranular Stress Corrosion Cracked Inconel Steam Generator Tubes." NUREG/CR-3561, PNL-4695, Pacific Northwest Laboratory, Richland, Washington.

Bowen, W. M., P. G. Heasler, and R. B. White. 1989. "Evaluation of Sampling Plans for In-Service Inspection of Steam Generator Tubes: Modeling of Eddy-Current Reliability Data, Analytical Evaluations, and Initial Monte Carlo Simulations." NUREG/CR-5161, PNL-6462, Vol. 1. Pacific Northwest Laboratory, Richland, Washington.

Bradley, E. R., P. G. Doctor, R. H. Ferris, and J. A. Buchanan. 1988. "Steam Generator Group Project - Task 13 Final Report - NDE Validation." NUREG/CR-5185, PNL-6399, Pacific Northwest Laboratory, Richland, Washington.

Clark, R. A. and P. G. Doctor. 1985. "Surry Steam Generator Program and Baseline Eddy Current Examination," Vol. 4, p. 260-285, in *Proceedings of the Twelfth Water Reactor Safety Research Information Meeting.* NUREG/CP-0058, Vol. 4. Pacific Northwest Laboratory, Richland, Washington.

Clark, R. A., P. G. Doctor, and R. H. Ferris. 1986. "Surry Steam Generator - Examination and Evaluation," Vol. 2, p. 409-430 in *Proceedings of the Thirteenth Water Reactor Safety Research Information Meeting,* NUREG/CP-0072, Vol. 2. Pacific Northwest Laboratory, Richland, Washington.

Clark, R. A. 1984. "Status and Progress of Research on a Removed-from-Service Steam Generator." NUREG/CP-0048, Vol. 4, p. 98-116, in *Proceedings of the Eleventh Water Reactor Safety Research Information Meeting,* Pacific Northwest Laboratory, Richland, Washington.

Clark, R. A. and M. Lewis. 1985. "Steam Generator Group Project Annual Report - 1984." NUREG/CR-4362, PNL-5417, Pacific Northwest Laboratory, Richland, Washington.

Clark, R. A. and M. Lewis. 1985. "Steam Generator Group Project Annual Report - 1983." NUREG/CR-4361, PNL-5017, Pacific Northwest Laboratory, Richland, Washington.

Clark, R. A. and M. Lewis. 1984. "Steam Generator Group Project Annual Report - 1982." NUREG/CR-3581, PNL-4693, Pacific Northwest Laboratory, Richland, Washington.

Clark, R. A. and M. Lewis. 1984. "Steam Generator Group Project Semi-Annual Progress Report July - December 1982." NUREG/CR-3580, PNL-4692, Pacific Northwest Laboratory, Richland, Washington.

Clark, R. A. and R. L. Bickford. 1984. "Steam Generator Tube Integrity Program Leak Rate Tests - Progress Report." NUREG/CR-3562, PNL-4629, Pacific Northwest Laboratory, Richland, Washington.

Clark, R. A., R. L. Bickford, A. S. Birks, R. L. Clark, P. G. Doctor, R. H. Ferris, L. K. Fetrow, W. D. Reece, E. B. Schwenk, and G. E. Spanner. 1985. "1984 Branch Annual Report - Steam Generator Integrity Program." NUREG-0975, Vol. 3, PNL-SA-13015, Pacific Northwest Laboratory, Richland, Washington.

Clark, R. A. et al. 1984. "1983 Branch Annual Report - Steam Generator Integrity Program." NUREG-0975, Vol. 2, PNL-SA-11720, Pacific Northwest Laboratory, Richland, Washington.

Doctor, P. G., A. S. Birks, R. H. Ferris, H. Harty and G. E. Spanner. 1988. "Steam Generator Group Project - Task 7: Final Report - Post-Service Baseline Eddy-

Current Examination." NUREG/CR-5087, PNL-5940, Pacific Northwest Laboratory, Richland, Washington.

Doctor, P. G., J. A. Buchanan, J. M. McIntyre, P. J. Hof and S. S. Ercanbrack. 1984. "Steam Generator Group Project Progress Report on Data Acquisition/Statistical Analysis." NUREG/CR-3579, PNL-3955, Pacific Northwest Laboratory, Richland, Washington.

Doctor, P. G., H. Harty, R. H. Ferris and A. S. Birks. 1988. "Steam Generator Group Project - Task 9: Final Report - Nondestructive Evaluation Round Robin, Comments Vol. I: Description and Summary Data and Vol. II." NUREG/CR-4849 Vol. I and II, PNL-5868, Pacific Northwest Laboratory, Richland, Washington.

Kurtz, R. J., R. A. Clark, E. R. Bradley, W. M. Bowen, P. G. Doctor, F. A. Simonen, and R. H. Ferris. 1990. "Steam Generator Tube Integrity Program/Steam Generator Group Project - Final Summary Report." NUREG/CR-5117, PNL-6446, Pacific Northwest Laboratory, Richland, Washington.

Kurtz, R. J., and R. A. Clark. 1988. "Compendium and Comparison of International Practice for Plugging, Repair, and Inspection of Steam Generator Tubing." NUREG/CR-5016, CSNI No. 140, PNL-6341, Pacific Northwest Laboratory, Richland, Washington.

Kurtz, R. J., J. M. Alzheimer, R. L. Bickford, R. A. Clark, C. J. Morris, E. A. Simonen and K. R. Wheeler. 1988. "Steam Generator Tube Integrity Program Phase II Final Report." NUREG/CR-2336, PNL-4008, Pacific Northwest Laboratory, Richland, Washington.

Kurtz, R. J., R. L. Bickford, W. M. Bowen, E. R. Bradley, P. G. Doctor, R. H. Ferris, L. K. Fetrow, and F. A. Simonen. 1988. "Surry Steam Generator Program - Final Results - Evaluation of NDE Performance and Technical Basis for Improved ISI and Plugging Criteria," Vol. 2, p. 429-457, in *Proceedings of the Fifteenth Water Reactor Safety Research Information Meeting*, USNRC Conference Proceeding NUREG/CP-0091, Vol. 2. Pacific Northwest Laboratory, Richland, Washington.

Kurtz, R. J., M. Lewis and R. A. Clark. 1987. "Steam Generator Group Project Annual Report - 1985." NUREG/CR-4848, PNL-5771, Pacific Northwest Laboratory, Richland, Washington.

Kurtz, R. J. and R. A. Clark. 1986. "1985 Branch Annual Report - Steam Generator Integrity Program." NUREG-0975, Vol. 4, PNL-5772, Pacific Northwest Laboratory, Richland, Washington.

Kurtz, R. J. 1987. "1986 Branch Annual Report - Steam Generator Integrity Program." NUREG-0975, Vol. 5, PNL-6110, Pacific Northwest Laboratory, Richland, Washington.

Kurtz, R. J. et al. 1988. "1987 Branch Annual Report - Steam Generator Integrity Program." NUREG-0975, Vol. 6, PNL-6401, Pacific Northwest Laboratory, Richland, Washington.

Kurtz, R. J., P. G. Heasler, and D. B. Baird. 1994. *Evaluation of Sampling Plans for Inservice Inspection of Steam Generator Tubes: Comprehensive Analytical and Monte Carlo Simulation Results for Several Sampling Plans*. NUREG/CR-5161, Vol. 2. Pacific Northwest Laboratory, Richland, Washington.

Morris, C. J., G. H. Lyon, T. J. Davis and C. B. Perry. 1980. "Eddy-Current Inspection of INCONEL-600 Steam Generator Tubes at the Tube Sheet." NUREG/CR-1626, PNL-3472, Pacific Northwest Laboratory, Richland, Washington.

Reece, W. D., G. R. Hoenes and M. A. Parkhurst. 1984. "Steam Generator Group Project Progress Report - Task 3 Health Physics." NUREG/CR-3578, PNL-4711, Pacific Northwest Laboratory, Richland, Washington.

Schwenk, E. B. and K. R. Wheeler. 1984. "Steam Generator Group Project Task 10 - Secondary Side Examination." NUREG/CR-3843, PNL-5033, Pacific Northwest Laboratory, Richland, Washington.

Schwenk, E. B. 1987. "Steam Generator Group Project Task 10 - Secondary Side Examination Final Report." NUREG/CR-4850, PNL-6045, Pacific Northwest Laboratory, Richland, Washington.

Sinclair, R. B. 1984. "Secondary Side Photographic Techniques Used in Characterization of Surry Steam

Generator." NUREG/CR-3094, PNL-5053, Pacific Northwest Laboratory, Richland, Washington.

Wheeler, K. R., P. G. Doctor, L. K. Fetrow and M. Lewis. 1984. "Steam Generator Group Project Task 8 - Selective Tube Unplugging." NUREG/CR-3842, PNL-4876, Pacific Northwest Laboratory, Richland, Washington.

1.0 Introduction

Eddy current (ET) inservice inspections of steam generator tubing are routinely performed as an element in the overall defense-in-depth strategy for ensuring the structural and leak-tight integrity of the reactor coolant pressure boundary. The main objectives of these inspections are to detect evidence of tube degradation so that corrective action(s) may be taken to mitigate tube damage, and to identify most or all degraded tubes that could fail before the next inspection. To attain these objectives a reliable inservice inspection (ISI) must be performed.

To ensure the reliability of these ISIs, a performance demonstration qualification test has been developed. A performance demonstration test should duplicate (as closely as possible) the conditions that would exist for the ET systems in the field on real steam generators. For this report, an ET inspection system was taken as the ET personnel, equipment, and procedure in combination. During the performance demonstration test, the ET system should have no more information available than in the field. The ET system should inspect tubing containing realistic flaws and should be graded on how reliably flaws can be found and sized. The test should not be designed to evaluate intermediate steps in the inspection process, but should concentrate on the ultimate outputs, proper flaw detection and sizing.

This report describes the methodology used for developing a performance demonstration test based on statistical design principles. Implementation of a performance demonstration test based on these statistical design principles will help to ensure that field inspection systems have a high probability of detecting and correctly sizing tube degradation. The first step in the development process is to formulate the basic performance demonstration objectives as hypotheses tests of the form:

H_0: The ET system is unacceptable

versus

H_1: The ET system is acceptable

Statistical design calculations can then determine a proper pass/fail threshold and the most efficient grading scheme for the hypothesis test.

Although this general strategy is quite straightforward, several important issues have to be resolved before a workable test is actually constructed. These issues include:

1. How should ET system performance (or reliability) be quantified?

2. What performance "thresholds" should ET systems exceed to be considered qualified?

3. How is a test constructed to ensure with a high level of confidence that ET systems exceed the selected performance thresholds?

The above list of issues produces a framework for the construction of a performance demonstration test. Information relating to the first issue is contained in Section 2 of this report, the second issue is addressed in Section 3 and the third issue is addressed in Sections 4 through 6. It is important to note that before a statistical test for performance demonstration can be constructed, these issues must be resolved. To construct a statistical test, one must describe the test objectives in quantitative terms. Related to these issues is the matter of measuring test performance (as distinguished from ET system performance). Statisticians have standard measures for evaluating the performance of a test (called Type I and Type II errors). Consequently, after a workable test has been constructed and effort is directed on improving the test or determining the appropriate sample size of the test, it becomes important to calculate Type I and II errors for the prospective tests.

2.0 Quantification of ET System Performance

A reliable ET inspection system must perform two tasks, it must detect flaws a high percentage of the time and then accurately size them. Because of this, evaluation of detection reliability is usually separated from sizing. Detection performance is quantified by means of a probability, which is most commonly called probability of detection (POD). Sizing performance has typically been quantified in a less standard manner. Generally speaking, sizing performance is usually described by some sort of regression model which relates true flaw size to measured flaw size. Parameters, defined in terms of this regression model, are then used to measure sizing performance.

2.1 Probability of Detection

Probability of detection is defined as the probability the ET system will detect a flaw of a certain size, s, and is denoted by $POD(s)$. With the use of POD, an ET system's performance can be summarized by a curve, as illustrated in Figure 1 below. This POD curve completely describes the two types of errors an inspection system can make during the task of detection. A "significant" flaw may be missed or good material may be called flawed. If C_0 is the size of a significant flaw, then $1 - POD(C_0)$ represents the chances of committing the first error, while $POD(0)$ represents the chances of committing the second error (i.e., the false call probability = $POD(0)$).

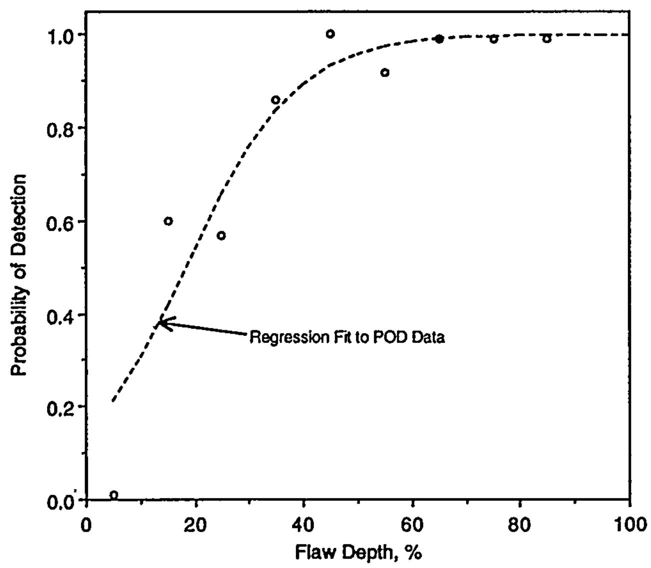

Figure 1. Example POD Curve

Inspections will produce binary data which can be used to estimate the POD. For example, the points displayed in Figure 1 represent binomial data obtained from inspections. Each point in this figure represents estimated POD for a collection of flaws grouped over a 10% size interval. Such data can be fitted to any parameterized family of curves as the points are in Figure 1. In Figure 1, the points have been fit to a logistic curve, perhaps the most popular type of curve used to model POD.

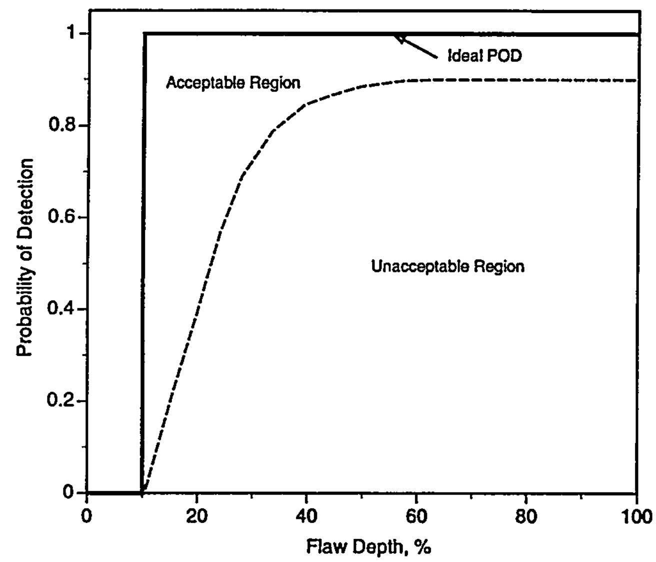

Figure 2. Ideal POD Performance

POD curves can be used to easily describe acceptable inspection system performance. This concept is illustrated schematically in Figure 2. Ideal inspection performance is represented by a step function as illustrated in Figure 2. For this step function, POD = 0 for flaws of no interest and POD = 1 for flaws of "significance." An inspection system with an ideal POD curve will never make a mistake; no false calls will be made and no "significant" flaws will be missed.

In the case of steam generator tube inspection, it is important to note that "significant" flaws are not just the ones large enough to threaten tube integrity. Detection of flaws smaller than the "critical" size is important because steam generator sampling strategies rely on detection of tube degradation at an early stage to aid in identifying defective tubes. (The term "critical' in the context of this report refers to a flaw severe enough to cause failure of the tube by leak or rupture. Failure could be by growth of a flaw through the tube wall resulting in a small leak, or failure could be by rupturing of a tube resulting in a large leak. Leaks of any size were considered sufficient to classify a tube as failed.) For example, U.S. NRC Regulatory Guide 1.83 (U.S. NRC 1975) and the most recent edition of the EPRI Inspection Guidelines (EPRI 1992) have criteria which trigger additional inspection when flaws less than 40% through-wall (TW) are detected. The rationale for this strategy is that detection of flaws less than the plugging limit indicates the presence of a problem in the steam generator. Depending on the numbers of degraded tubes discovered, additional inspection may be required. Further, the location of all degraded tubes must be recorded and these tubes included in the sample set for the next inspection. This strategy is based on the idea that detection of tube degradation below rejectable limits calls attention to a potentially significant condition that needs to be followed and aids the process of identifying tubes with rejectable flaws by causing more tubes in the steam generator to be examined.

In order to specify POD curves that are "acceptable", it is therefore natural to designate a region that is "close" to the ideal step function as illustrated by the dotted lines in Figure 2. POD curves that do not fall within this region would be considered unacceptable and any system that has such a POD would be considered unqualified.

Difficult compromises are involved in the determination of this region, however. The more stringent it is made, the more likely that no existing inspection system can satisfy its requirements and the more likely resources will be required to develop new detection techniques. On the other hand, the less stringent it is made, the greater the post-inspection tube failure probability. Since the purpose of ISI is to catch flaws before tube integrity is threatened, it is important for field inspection systems to exhibit POD performance as close to the ideal POD behavior as possible. This is needed to minimize the probability for tube leaks and ruptures, which could affect steam generator safety and reliability.

2.2 Flaw Sizing Regression Model

Sizing performance is generally evaluated through a regression model. Most commonly, flaw sizing is assumed to obey a regression relationship of the form:

$$M(t_i) = \beta_1 + \beta_2 * t_i + e_i \tag{1}$$

where $M(t_i)$ represents the measured size of a flaw with true size t_i. According to this regression model, measured and true sizes are related to each other in a linear manner as defined by the parameters β_1 and β_2.

The error term e_i is assumed to be a normal deviate with mean 0 and constant standard deviation of σ. Although these assumptions are not often explicitly stated, they are necessary if the regression results are to give an adequate description of sizing performance. With some sizing procedures, the error distribution is skewed, and the shape of the distribution is dependent on flaw size. The error distribution for small flaws typically has a heavy right-hand tail while the situation is reversed for large flaws. In other words, small flaws tend to be oversized while large flaws are typically undersized.

Ideal sizing performance should fit a regression model of the form:

$$M(t_i) = 0 + 1 * t_i \tag{2}$$

In other words, ideal sizing performance exhibits $\beta_1 = 0$, $\beta_2 = 1$, and $\sigma = \text{stdev}(e) = 0$. Therefore, acceptable sizing performance should fulfill the following criteria:

1. σ should be suitably small.

2. $\beta_1 \approx 0$ and $\beta_2 \approx 1$ so there is little bias in the measurements.

3. The linear regression model should be a reasonable description of the data.

When the sizing data does not fit a linear regression model, this must be considered unacceptable performance. This may result in a non-linear relationship between the true and measured sizes or a non-normal error distribution.

A single parameter known as the mean-square-error (MSE) can be used to ensure that all of the above requirements for acceptable sizing performance are met. In fact, if one requires that mean square error is less than c^2, that is,

$$MSE(t) = E(M(t) - t)^2 < c^2 \qquad (3)$$

then the following bounds on the regression parameters must hold:

1. $\sigma < c$

2. $\beta_1 < c$ and

3. $|\beta_2 - 1| < c/t$

The MSE is therefore a very concise parameter for specifying acceptable sizing performance. In order to provide a reasonable requirement for sizing performance, one should only require a low MSE for flaws in the range from 10% to 100% TW. When flaws are smaller than 10% TW and very difficult to size, it could be an unreasonable requirement for ET systems to produce a low MSE..

3.0 ET System Performance Thresholds

3.1 Degraded and Defective Tubes

As discussed in Section 2.0, a reliable ET inspection system must detect and accurately size "significant" flaws a high percentage of the time. In addition, an ET inspection system must also possess acceptable reliability to detect and size smaller flaws which are not "significant" but which serve to call attention to conditions which may require corrective actions to mitigate further tube damage. In this section definitions of degraded and defective tubes to provide the basis for establishing the ET system performance thresholds are given.

A defective tube is one which contains a flaw of such severity that the tube is unacceptable for continued service at the end of the next operating cycle. A degraded tube is one which contains a flaw of lesser severity than a defective tube. For this work a defective tube was defined as one with TW degradation severe enough to cause tube failure under main-steam-line-break loading conditions. To determine the flaw severity which would result in a tube being classified as defective, test data on tube failure pressure as a function of flaw size and geometry were utilized.

Relevant failure pressure data have been published by Alzheimer, et al. (1979) and Kurtz, et al. (1988) on mechanically and chemically flawed specimens of Inconel 600 tubing. Three types of flaws were used to simulate known or postulated defects occurring in operating steam generators. Electro-discharge machined (EDM) slots or stress corrosion cracks (SCC) were used to simulate crack-type flaws, mechanically or chemically-induced uniform thinning to simulate general wall thinning, and mechanically or chemically-induced elliptical wastage to simulate localized wall thinning. Tube specimens were burst and collapse tested at steam generator operating temperatures under controlled loading conditions. Three sizes of Inconel 600 tubing used in the United States were tested, as well as three different strength levels of a single tubing size. In all, more than 600 specimens were tested. From the data, constitutive equations were developed relating tube failure pressure to flaw size and morphology. The constitutive equations were initially developed from the failure pressure data obtained from tubes with machined flaws. These equations were then validated by tests run on tubes with chemically induced flaws. A final validation of the constitutive equations was then performed by conducting burst tests on tubes with actual service-induced flaws that were removed from a retired steam generator. The equations developed for burst loading conditions were the following:

For EDM slots:

$$\frac{\Delta P}{\Delta P_o} = 1 - \frac{a}{t} + \frac{a}{t}\exp\left\{\frac{-0.373*L}{\sqrt{Rt}}\right\} \qquad (4)$$

For mechanically and chemically-induced uniform thinning:

$$\frac{\Delta P}{\Delta P_o} = \left[1 - \frac{a}{t}\right]^{\left\{1 - \exp\left(\frac{-0.142*L}{\sqrt{R(t-a)}}\right)\right\}} \qquad (5)$$

For mechanically and chemically-induced elliptical wastage:

$$\frac{\Delta P}{\Delta P_o} = \left[1-\frac{a}{t}\right]^{0.626} \qquad (6)$$

where

$\frac{\Delta P}{\Delta P_o}$ = ratio of defected/undefected (normalized) tube burst pressure
a = defect depth
t = wall thickness
R = inner radius of tube
L = defect length.

Figure 3 gives a plot of Equation 4 as a function of slot length for three fixed slot depths. The plot was generated for tubes with nominal dimensions of 0.875 in. OD by 0.050 in. wall thickness. The data points shown are the normalized burst test data for 0.875 x 0.050 tubing. Note the decreasing influence of slot length on normalized burst pressure for slots greater than ≈ 0.8 in. Note scatter of the data points about the calculated lines is due, in part, to variation in slot depth for a fixed slot length. Figure 4 shows a plot of Equation 4 for constant slot lengths. The same burst test data shown in Figure 3 were plotted in Figure 4. Note that for long flaws (≈ 1.5 in.) the data points indicate the burst pressure decreases linearly with slot depth. For short slots (≈0.25-0.5 in.), the dependence of the burst pressure on slot depth is linear with depth for shallow slots,

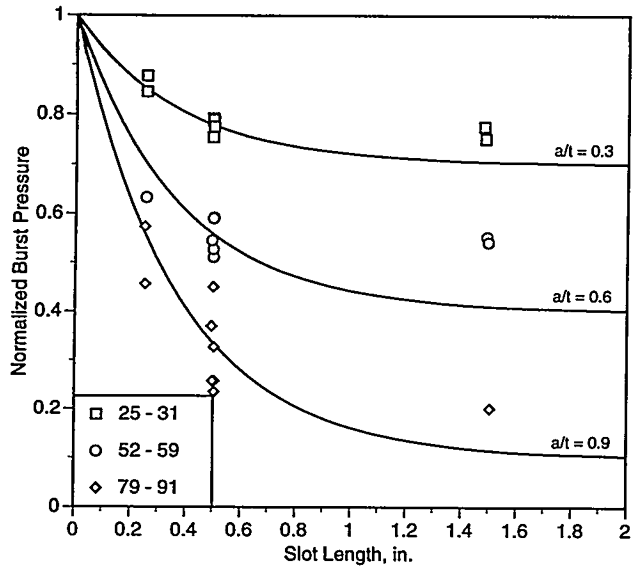

Figure 3. Normalized Burst Pressure vs. Slot Length for 0.875 in. OD x 0.050 in. Wall Thickness Steam Generator Tubing Tested at 600°F. Lines are plot of Equation 4.

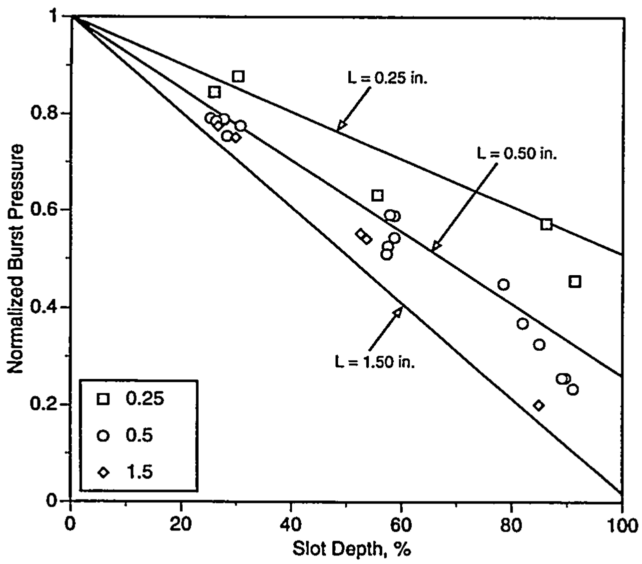

Figure 4. Normalized Burst Pressure vs. Slot Depth for 0.875 in. OD x 0.050 in. Wall Thickness Steam Generator Tubing Tested at 600°F. Lines are plot of Equation 4.

but then appears to fall off more rapidly with increasing depth for slots greater than about 80% through-wall.

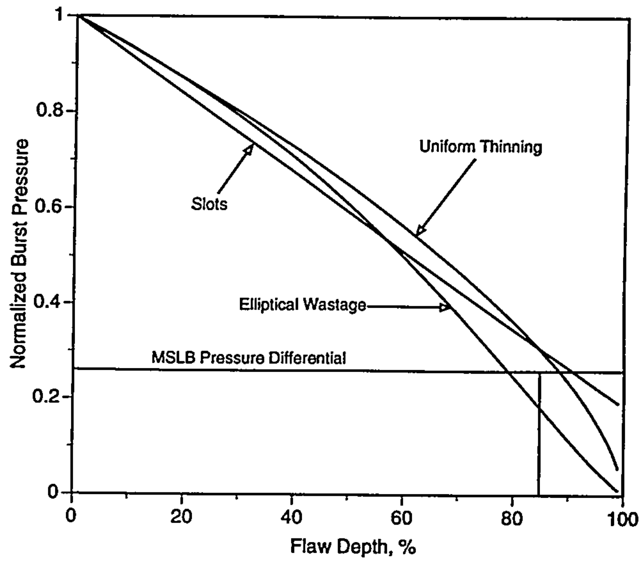

Figure 5. Plots of Equations 4, 5, and 6, for 0.875 in. OD x 0.050 in. Wall Thickness Tube. Flaw Length ≥ 0.875 in.

The burst test constitutive equations were used to develop a definition of an unacceptable flaw which was used in the development and evaluation of ISI sampling plans and performance demonstration qualification criteria. Figure 5 shows a plot of Equations 4, 5, and 6 for an 0.875 x 0.050 tube with a flaw 0.875 in. long. Note the length of all elliptical wastage flaws was always ≈1.5 in. It is evident from this plot that an 85% TW flaw represents an average depth for all flaw types that would fail under main-steam-line-break loading conditions (≈2600 psi pressure differential). If a flaw growth rate of 10% per operating cycle is assumed, then a tube with an actual flaw ≥ 75% TW flaw could fail under main-steam-line-break loading conditions by the end of the next operating period. This level of degradation was used to define an unacceptable (i.e., defective) tube condition requiring tube plugging or repair.

3.2 Information on ET Inspection Reliability

To guide the selection of POD and flaw sizing performance thresholds, prior research results (Kurtz et al. 1990), on the reliability of ET inspection techniques to detect and size flaws in laboratory and service-degraded tubes were utilized. The most extensive and realistic ET reliability data base was obtained from round robin examinations of the retired-from-service Surry 2A steam generator. This section summarizes the salient results and conclusions from that research effort.

In the Surry study four round robins, involving 25 teams, were conducted to determine the reliability of conventional multi-frequency ET and alternative NDE methods. To validate the *in situ* NDE results, more than 550 tube segments were removed from the generator. Pitting and wastage were the predominant tube defects found. The most severe pitting/wastage degradation was located on the hot-leg in the region 0 to 2 in. above the top of the tube sheet (TTS) where TW depths ranged up to 87%. Wide variations in the distribution and depth of degraded areas were observed both axially and circumferentially within the corroded region of the hot-leg TTS specimens. These variations in defect distributions were a major factor in the variability of the ET depth estimates. In general, the TW depth from pitting/wastage-type degradation in specimens from other regions of the generator was less than 20% TW.

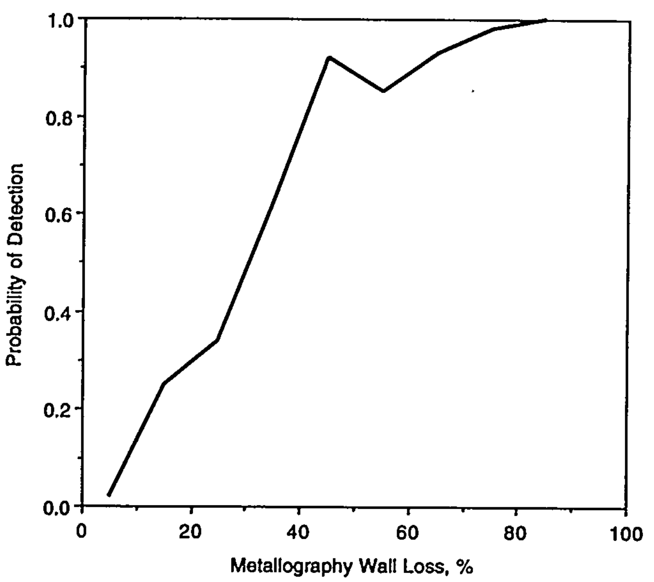

Figure 6. Average POD for Teams Inspecting Retired-from-Service Surry 2A Steam Generator Using Zetec MIZ-12/DDA-4 Equipment

Estimates of the POD were obtained by matching the ET inspection results with data from both visual and destructive metallographic analysis of the removed specimens. For each "true flaw size" category, the number of non-zero ET indications divided by the total number of flaws in that size category was used as a POD estimate. Results of these analyses indicated that POD depended on flaw severity. The curve shown in Figure 6 gives the average POD performance for seven teams employing conventional Zetec MIZ-12 multi-frequency inspection and DDA-4 analysis equipment. The curve was based on metallographic measurement of the maximum TW depth for defects from all regions of the steam generator combined. It should be noted that almost all of the data pertain to pitting/wastage type defects and the estimated POD values may not be appropriate for other forms of degradation, such as stress corrosion cracking (SCC) or intergranular attack. The oscillatory behavior of the curves is due to the relatively small numbers of specimens in each of the incremental TW depth categories.

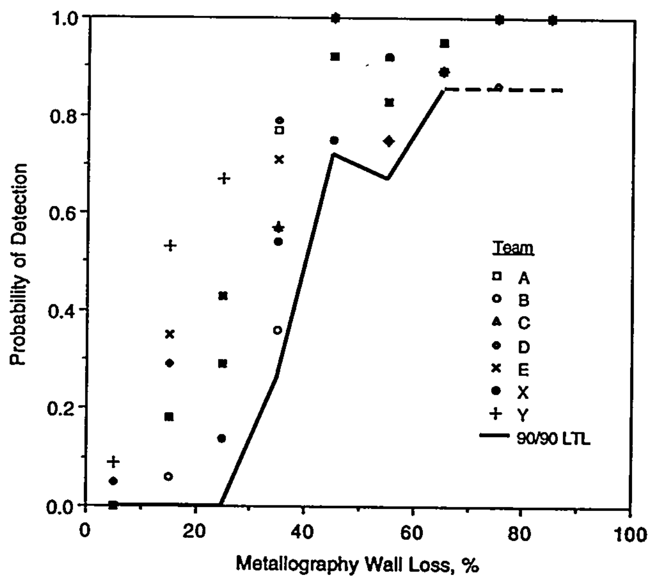

Figure 7. Individual POD Estimates for the Same Seven Teams Used to Develop Figure 6. Curve is 90/90 Lower Tolerance Limit.

Figure 7 is a plot of the individual POD estimates for the same seven teams used to develop Figure 6. The curve in Figure 7 is an approximate 90/90 lower tolerance limit (LTL) for these teams. The teams are assumed to be typical of the total population of teams performing ISI; therefore, if each team in the total population of teams performing ISI had inspected the round robin tube set, we can be 90% confident that 90% of the individual team POD values would be above the LTL. Note the portion of the curve extending from 65% to 85% TW depth is flat because the number of specimens with defects in this range is not adequate to provide a meaningful estimate of the LTL. Thus, the LTL at 65% TW depth was extended as a conservative approximation of the LTL for TW depths ≥ 65%.

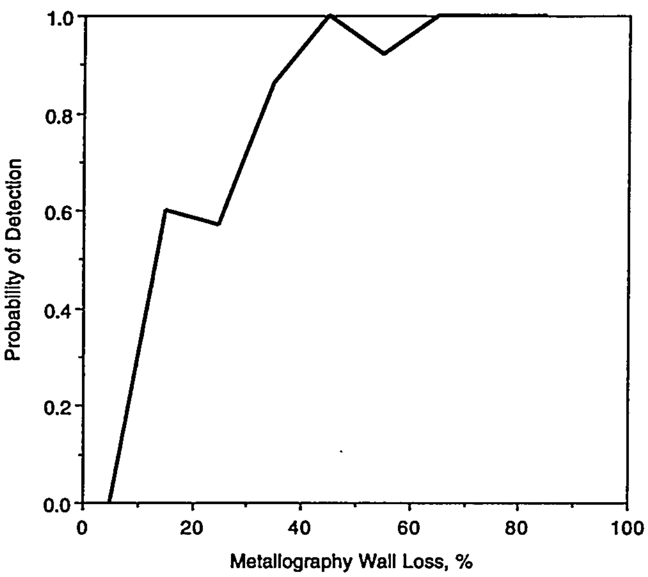

Figure 8. Best POD Performance for Teams Inspecting Retired-from-Service Surry 2A Steam Generator

As shown in Figure 8, an apparent improved POD performance was observed for one team that employed alternative inspection methods. The POD curve for this team increased more rapidly at small flaw depths and was higher above 40% TW depth than the POD curves for other teams. This team employed specially developed frequency mixes to enhance the signal-to-noise ratio and computer data screening techniques to minimize the possibility of a missed indication.

Wide variations in the reported ET depth estimates were observed between specimens with similar TW depths and also within the same specimen for data from different inspection teams. The team-to-team variations for a given specimen appear to result from differences in analysis procedures or the analyst's interpretation of the complex ET patterns. For the same team, the

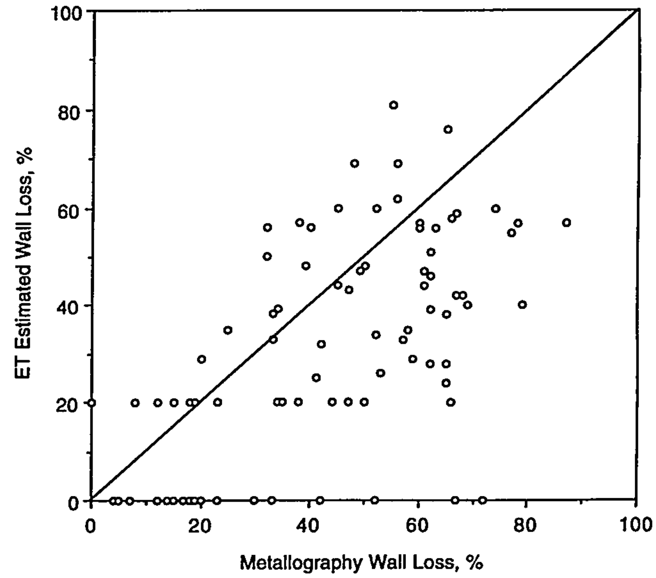

Figure 9. Typical ET Sizing Performance for Teams Inspecting Retired-from-Service Surry 2A Steam Generator and Using Zetec MIZ-12/DDA-4 Equipment

defect morphology and distribution within the corroded region was considered the major cause for variations between specimens with similar TW depths. However, dents and deposits near the defects also contributed to the sizing variations. In general, teams tended to undersize pitting/wastage type degradation, especially for severely degraded specimens. Figure 9 shows the relationship between ET estimated defect depth and metallographic results for a typical inspection team using conventional multi-frequency inspection equipment and procedures. The best correlation observed is shown in Figure 10. This team used alternative inspection equipment and specially developed frequency mixes to augment their conventional inspection data to achieve improved sizing accuracy and precision.

3.2.1 SCC Mini Round Robin

To supplement the ET reliability information obtained from the round robins on the steam generator, an additional round robin was performed to provide information on the reliability of ET techniques to detect and size SCC (Kurtz et al. 1988). A tube bundle consisting of tubes with laboratory-produced SCC of varying part TW depths and lengths was assembled and sent to several firms that routinely conduct ISI of steam gener-

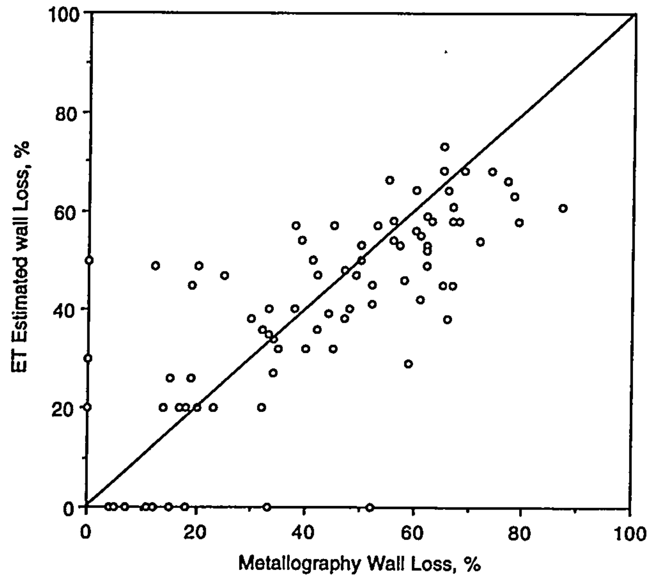

Figure 10. Best Observed ET Sizing Performance for Teams Inspecting Retired-from-Service Surry 2A Steam Generator

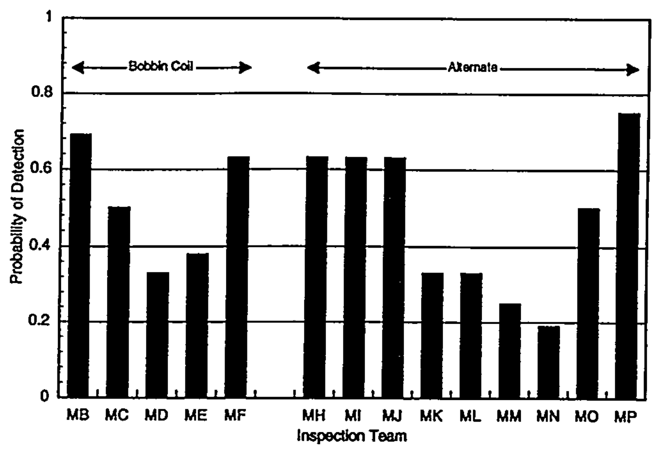

Figure 11. POD Results for SCC Mini Round Robin

ators. Some of the tubes were coated with a 1- to 2-mil nonuniform layer of copper to simulate the deposits found on tubes removed from the retired-from-service steam generator.

Each round robin participant performed a standard bobbin-coil inspection with 100 kHz and 400 kHz frequencies and any other frequencies of their choice. Each was also asked to inspect the tube bundle with any alternative technique desired. Typically, specially designed bobbin-coil and rotating pancake-coil probes were used for the alternative inspections.

Results indicated that the average POD of SCC by conventional bobbin-coil and alternative inspection techniques was low (see Figure 11). The average POD for teams using conventional inspection techniques was 0.51. The average POD for teams using alternative inspection techniques was 0.47, but for teams using rotating pancake or array coil techniques, the average POD was slightly higher at 0.63. The reliability of the various ET techniques to determine SCC length and depth was neither accurate nor precise. The alternative inspection results did not show improved sizing capability compared to conventional bobbin-coil techniques.

3.3 Evaluation of ISI Sampling Plans

Another significant consideration in the development of performance demonstration qualification criteria is the level of ET reliability needed to ensure the ISI sampling plan is effective. In prior research work (Bowen, Heasler, and White 1989), an evaluation and comparison of a number of sampling plans for ISI of steam generator tubes was performed. This section gives the significant results from this work.

The primary criterion for comparing sampling plans was the probability of detecting and either plugging or repairing defective tubes. For a single tube, the probability of detecting and either plugging or repairing a defective tube is a function of two other probabilities: 1) the POD, and 2) the conditional probability, denoted by PEL, that the ET reading will exceed the plugging limit. Both the POD and PEL are functions of the true size and type of flaw. They also depend upon the capability and reliability of the ET inspection system.

The ET reliability data from the retired-from-service generator were used to develop empirical models of POD and PEL as functions of true flaw size for tubes with pitting/wastage type flaws. Because multiple inspection teams were involved, the statistical modeling yielded a range of estimated POD and PEL values for each specified flaw size. These ranges of values were utilized with probability theory and Monte Carlo simulation techniques to evaluate and compare several sampling plans.

An analytical evaluation was performed in which it was assumed that defective tubes tend to occur in clusters of degraded and defective tubes. For this analysis a cluster configuration consisting of one defective tube surrounded by four degraded but not defective tubes was assumed. This cluster configuration was chosen because it would be harder to detect than a larger cluster or a cluster that includes more than one defective tube; it should, therefore, provide conservative (lower bound) results. Of course, in an actual generator, a cluster could have a different shape and composition than were assumed for this analysis.

Two systematic/sequential sampling plans were evaluated in this phase of the study. In these sampling plans an initial systematic sample of either 20% or 40% of the tubes was selected and each tube in the sample was inspected. When an ET indication due to degradation was observed, inspection continued in the region immediately surrounding the suspect tube until a two tube "buffer zone" consisting of tubes with no ET indications was observed. Each tube with an ET indication exceeding the plugging limit was considered to be plugged or repaired. As a basis for comparison, 100% inspection was evaluated; however, the analytical results for 100% inspection did not depend on the clustering assumption.

Results of the analytical evaluation indicated that even with 100% inspection, most teams that inspected the Surry steam generator cannot detect and plug more than 65% of the defective tubes present. Analysis of sampling plans demonstrated that if the clustering assumption holds, and if the POD for the degraded tubes in the cluster was at least 0.7, then 40% systematic/sequential sampling was nearly as effective as 100% inspection for detecting and plugging defective tubes. This was true for any PEL value. However, 20% systematic/sequential sampling was found to be significantly inferior to both 100% inspection and the 40% systematic/sequential scheme.

Monte Carlo simulation analyses were also conducted to further evaluate and compare the effectiveness of various sampling plans for detecting and plugging defective tubes. Tube maps representing several distributions of degraded and defective tubes were considered. Two POD models, two ET sizing models, and one plugging limit were considered in various combinations with the tube maps. This work has been described in a recent report (Kurtz, Heasler, and Baird 1994).

There were six tube flaw distributions used in the simulations. Figures 12 to 17 show the tube maps used in the study. Tube maps 21 and 20 had 3 and 12 isolated defective tubes, respectively. Tube maps 1A, 6A, 8A and 13A had varying degrees of "defective tube clustering" (from low to high) and varying numbers of defective tubes (from low to high). The phrase "defective tube clustering" refers to the number of degraded or defective tubes surrounding a particular defective tube. A range of defective tube clustering was considered to test the dependency of sampling plan effectiveness on the cluster assumptions.

Table 1. ET Sizing Models Used in Monte Carlo Simulations

Model	Equation	SD*	Description
1	14.5+0.46 (x)	16	Average U.S. Team
2	12.6+0.68 (x)	10	Best Team

*SD = Standard Deviation

A comparison of the POD models used against the average and best POD curves from the Surry round robin study is shown in Figure 18. Table 1 presents the relevant parameters of the sizing models considered. For each combination of POD curve/ET sizing model, 25 applications of the various sampling plans were simulated. Summary results were plotted for comparison.

Eight sampling plans were evaluated: 100% inspection; the standard technical specification (STS); and six plans consisting of the combination of sampling type (systematic or random) and initial sample size (20%, 33.3% and 40%). It should be noted the initial sample size and expansion rules employed by the STS sampling plan are considerably different than the systematic sampling plans summarized here. A more detailed comparison

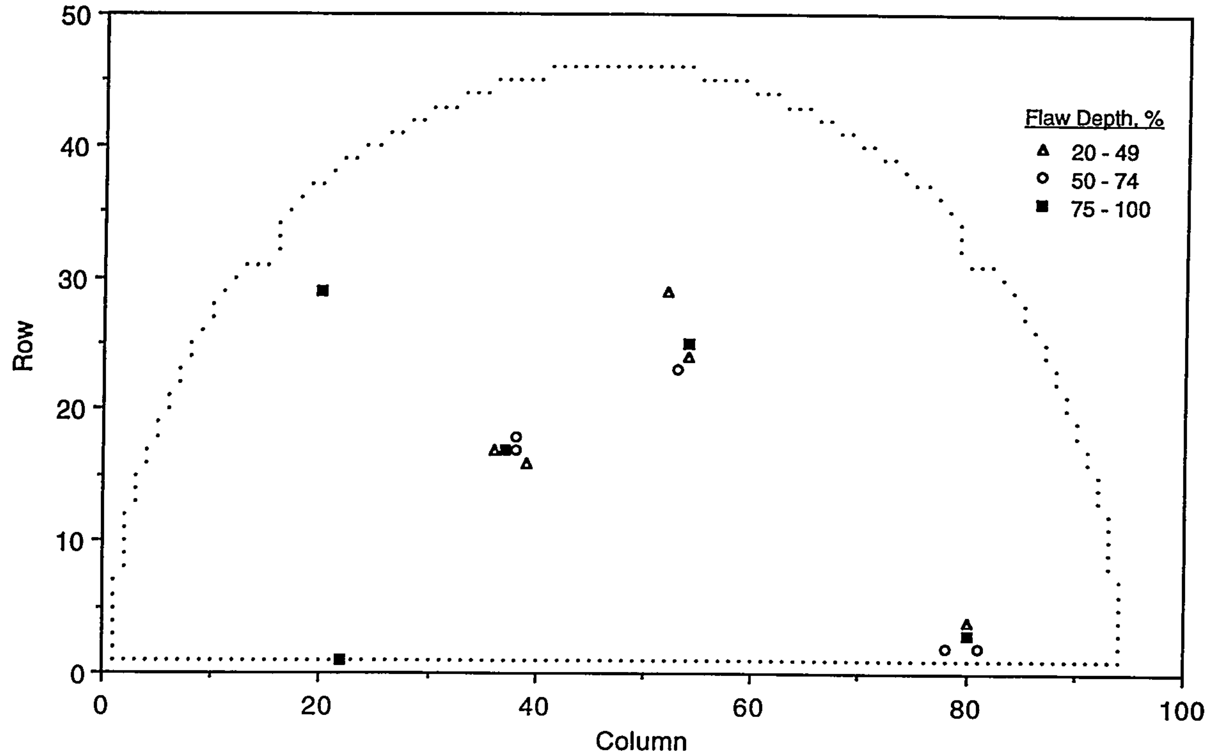

Figure 12. Tube Map 1A

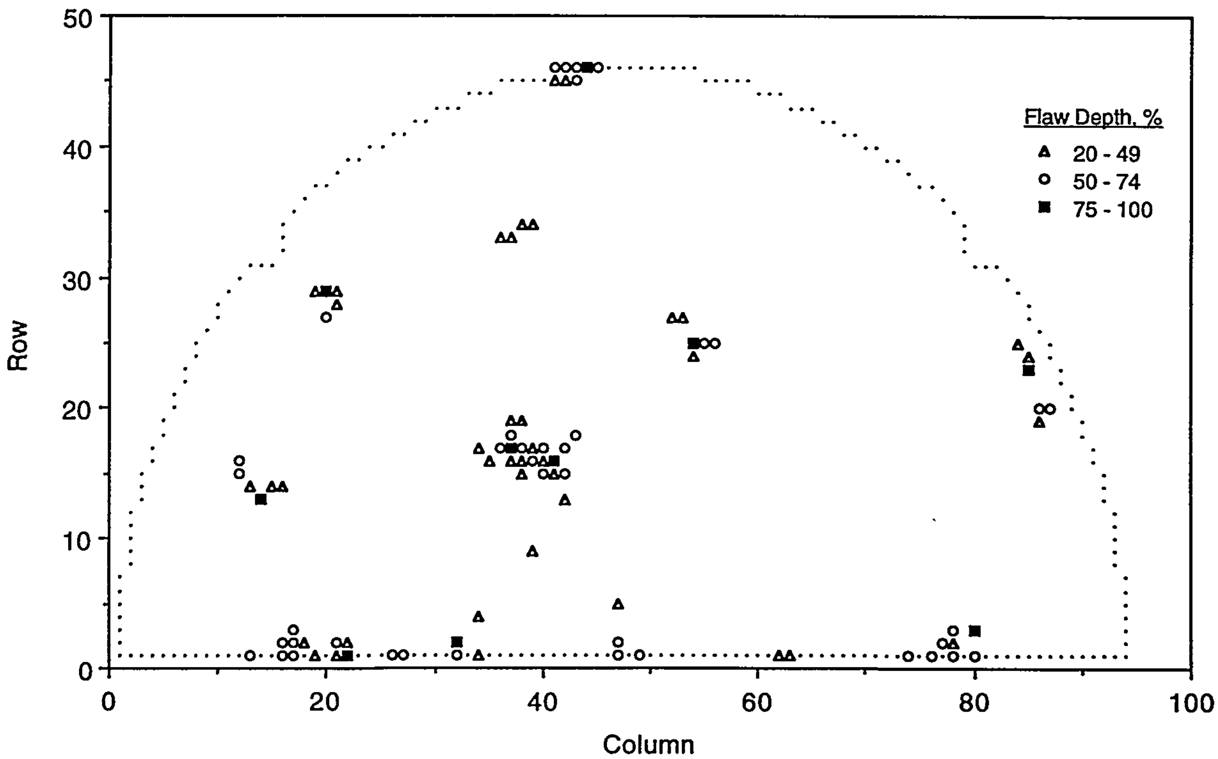

Figure 13. Tube Map 6A

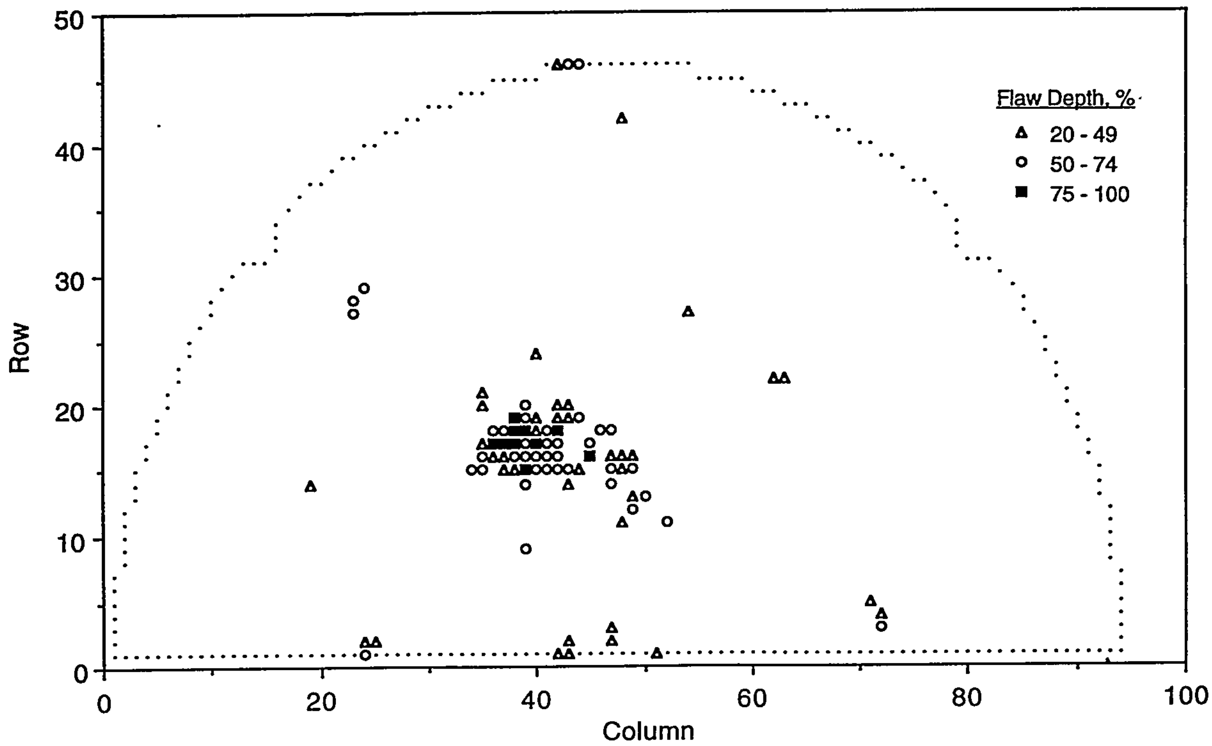

Figure 14. Tube Map 8A

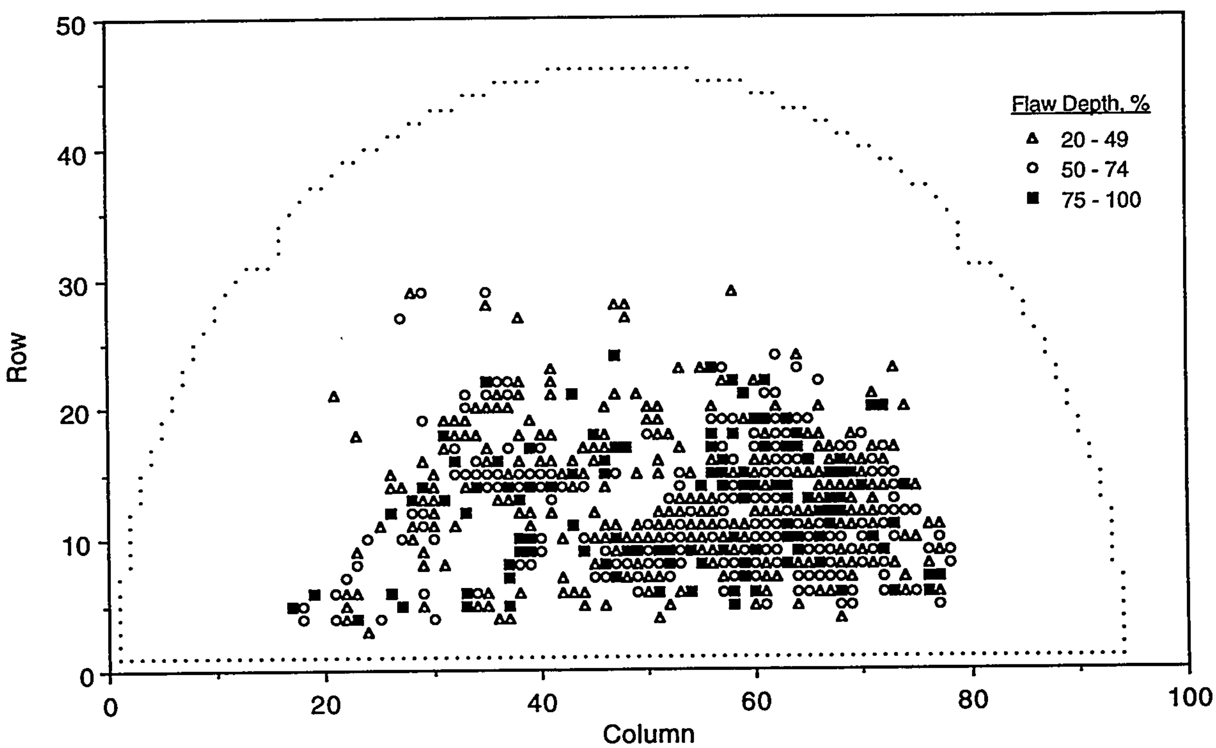

Figure 15. Tube Map 13A

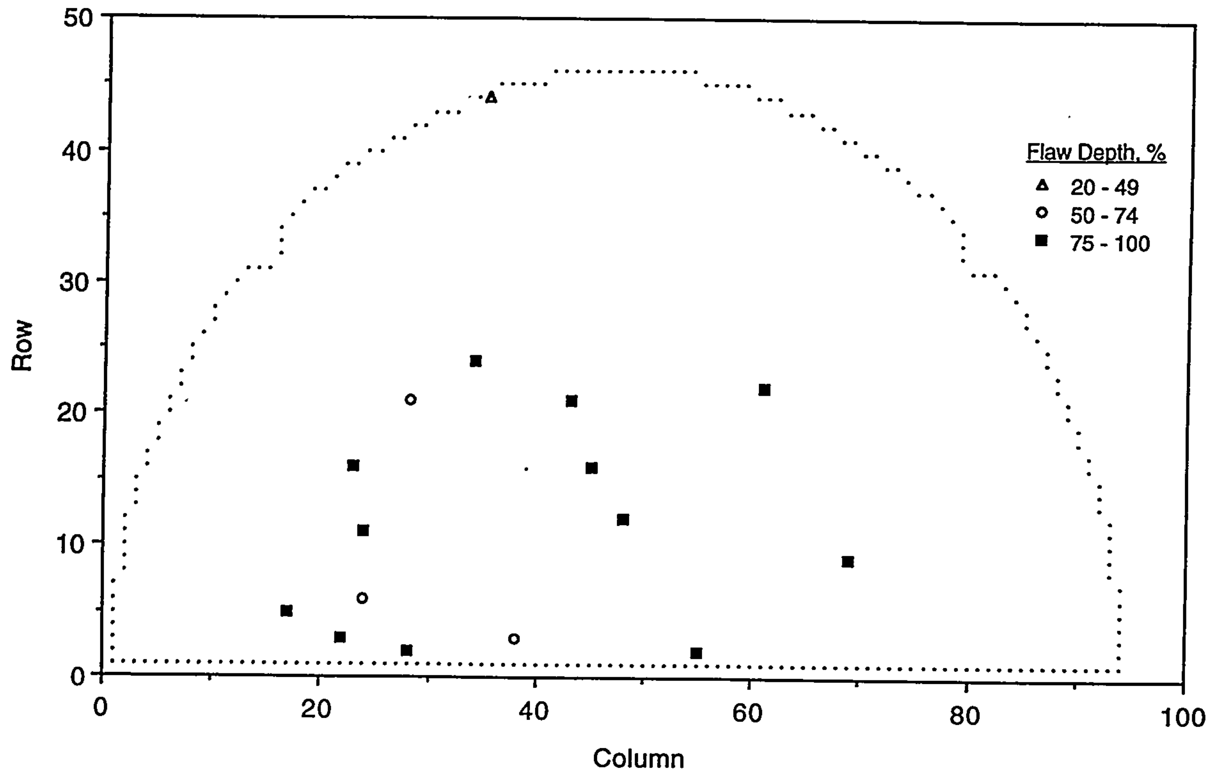

Figure 16. Tube Map 20

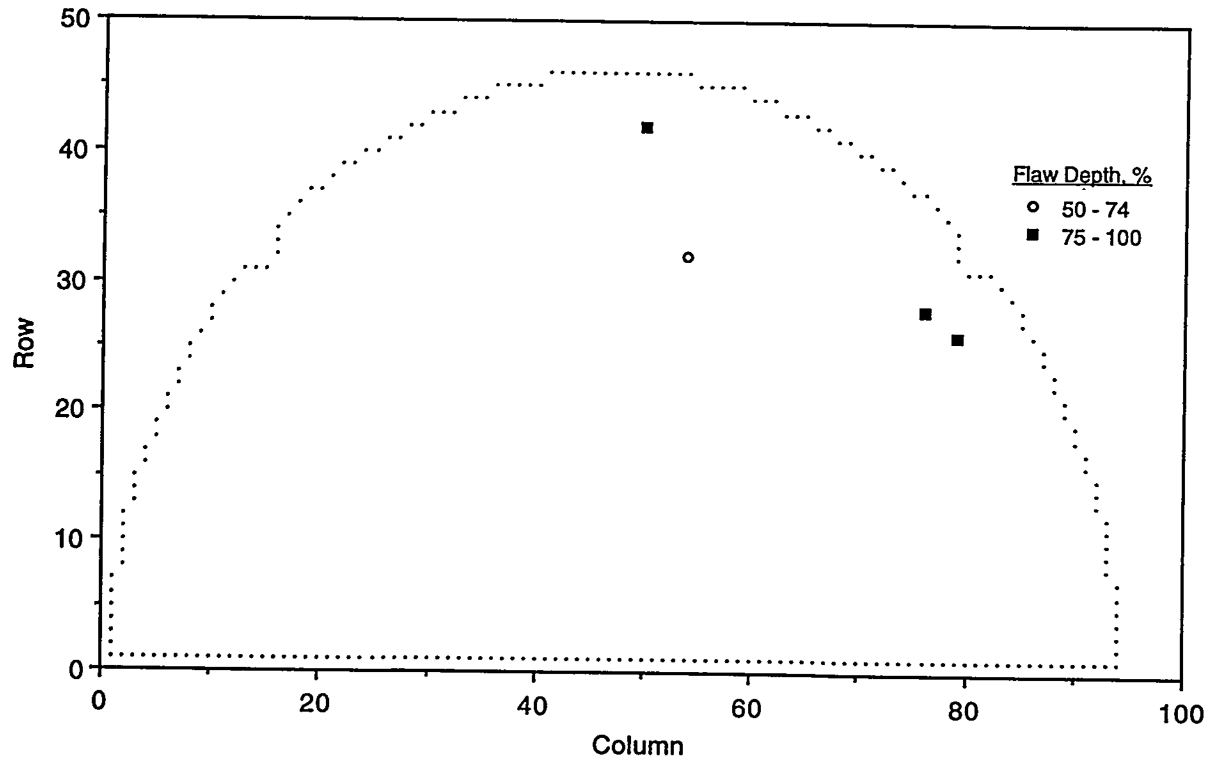

Figure 17. Tube Map 21

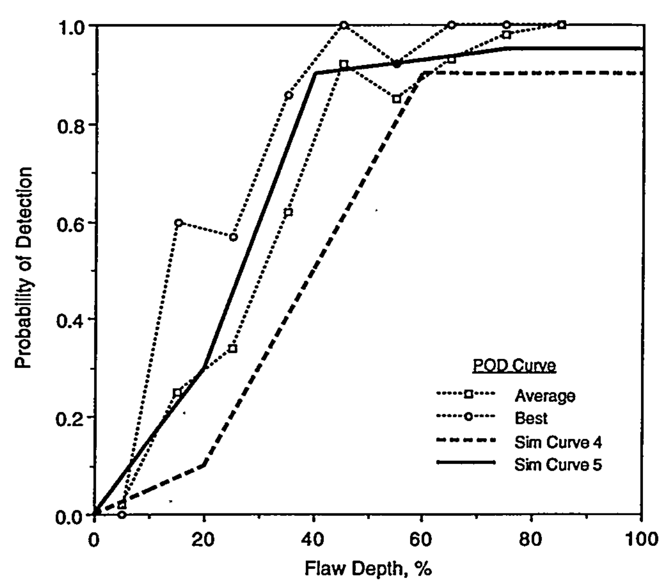

Figure 18. Comparison of Surry Round Robin POD with Simulation Curves 4 and 5

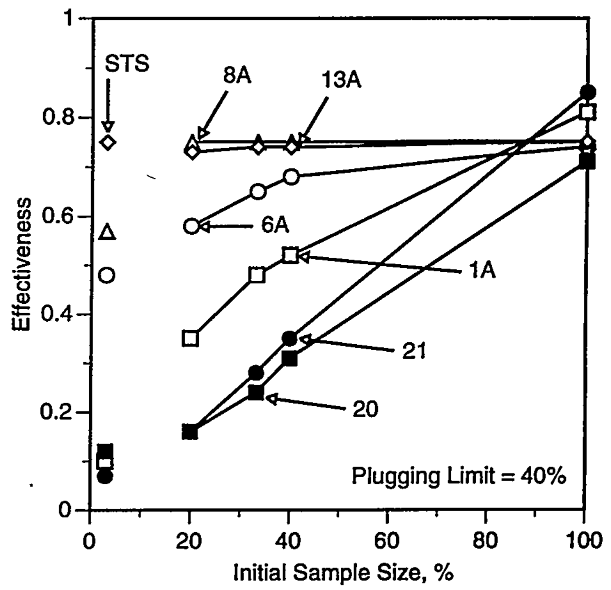

Figure 20. Sampling Effectiveness for Various Distributions of Tube Degradation Using POD Curve 4 and ET Sizing Model 1

of the differences is given in the recent report by Kurtz, Heasler, and Baird (1994).The principal measure of plan performance for detecting and plugging tubes was the sampling plan effectiveness. The sampling plan

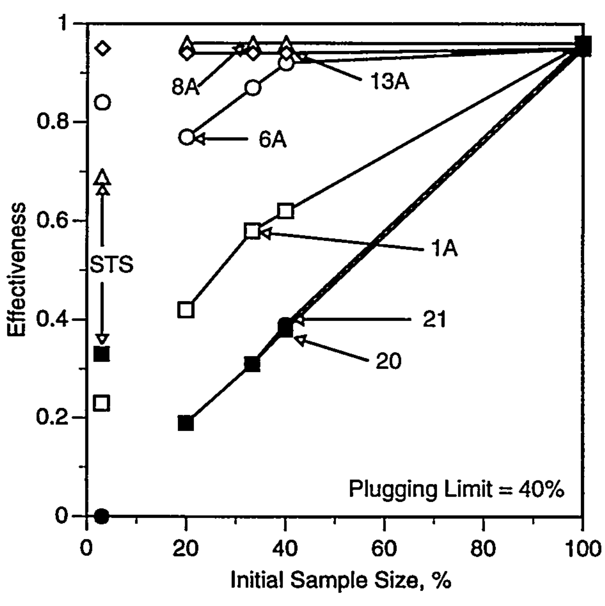

Figure 19. Sampling Plan Effectiveness for Various Distributions of Tube Degradation Using POD Curve 5 and ET Sizing Model 2

effectiveness was defined as the ratio of the average number of defective tubes plugged to the total number of defective tubes in the tube map. The effectiveness parameter provided a means for comparing the plugging capability of various sampling plans across the different tube maps.

The results of the Monte Carlo simulations support the conclusions reached from the analytical evaluation and provide valuable additional insights. The average effectiveness values for the systematic sampling plans (similar results were obtained for the random plans) along with results for the standard technical specification and 100% inspection are presented in Figures 19 and 20 for the best and worst combination of flaw detection and sizing capability, respectively. When defective tubes were grouped in one large cluster with degraded tubes, all the sampling/inspection schemes were equally effective (except for the standard technical specification). However, when defective tubes were isolated, then the sampling plan effectiveness was approximately equal to the initial systematic sample size. Thus, for isolated defective tubes, large sample sizes were the most effective. For intermediate levels of clustering, large sample sizes were more effective than smaller ones. There was an indication that systematic sampling plans were better than random sampling plans of equal initial size, but this result was not statistically significant. Comparing Figures 19 and 20 shows that improving the POD and flaw sizing capability to the best observed in the NDE round robins yielded improved effectiveness for all of the sampling/inspection schemes.

3.4 ET System Performance Thresholds

The purpose of the performance demonstration test is to provide a mechanism to ensure that field inspection systems (i.e. personnel, equipment and procedure) can reliably detect and size flaws in steam generator tubing. For this development effort the goal of steam generator tube ISI was to identify most or all defective tubes which could fail during reactor operation. The sampling plan evaluation work indicated that a 40% systematic, sequential sampling strategy was almost as effective as 100% inspection for identifying defective tubes, assuming some clustering of tube degradation. The sampling strategy relies on two key concepts to achieve this high level of effectiveness. First, a relatively large, uniformly distributed initial sample is used to provide a reasonable probability of finding isolated defective tubes, and second, detection of tube degradation of any severity triggers second-stage inspection to aid in finding defective tubes which may be in close proximity. In order for this sampling strategy to be effective, relatively high flaw detection reliability is needed even when degradation is < 75% TW.

Based on the above results, the POD curve denoted "unacceptable" and shown in Figure 21 was selected to define unacceptable POD performance. The defining points for the unacceptable POD curve are listed in Table 2. This particular unacceptable POD curve was selected so that ET systems possessing performance characteristics at or below the unacceptable curve would fail the test a high percentage of the time, and ET systems with performance characteristics similar to the acceptable curve plotted in Figure 21 and listed in Table 3 would pass the POD test a high percentage of the time. A team with "acceptable" POD performance would have a ≥ 95% probability of detecting a defective tube (flaws ≥ 75% TW) and ≥ 90% probability of detecting flaws ≥ 40% TW.

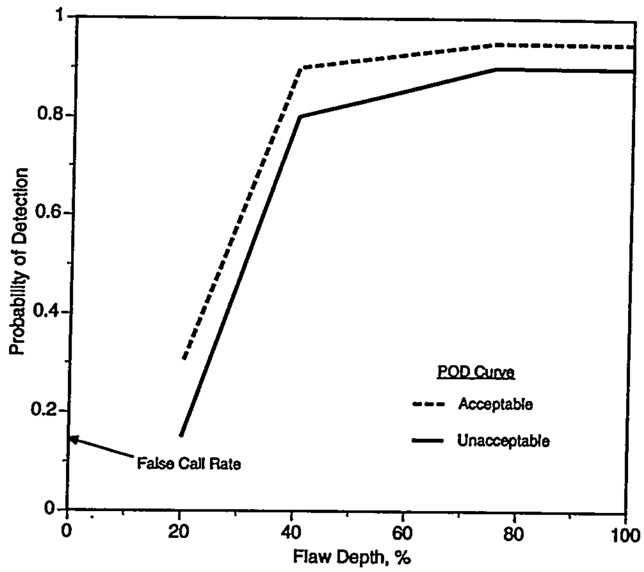

Figure 21. Base Case Threshold POD Performance Characteristics

Table 2. Defining Points for Unacceptable POD Performance

Through-Wall Flaw Depth, %	Threshold POD Curve
0 (Blank)	> 0.15
20	< 0.15
40	< 0.80
75	< 0.90
100	< 0.90

There are two mistakes that can be made when using results from a performance demonstration test to determine the acceptability of an ET system. The first mistake is that an ET system is called acceptable when it is really unacceptable. The probability of making this type of mistake is called the Type I error. The second mistake is that an ET system is called unacceptable when it is really acceptable. The probability of making this type of mistake is called the Type II error. The probability of correctly identifying an acceptable ET system is called the power of the test.

Table 3. Defining Points for Acceptable POD Performance

Through-Wall Flaw Depth, %	Acceptable POD Performance
0 (Blank)	< 0.15
20	0.30
40	0.90
75	0.95
100	0.95

For this test, as with any hypothesis testing problem, limits on the Type I and Type II errors are specified. These limits and the pass/fail thresholds determine the final sample size requirements. The acceptable and unacceptable ET thresholds (given in Tables 2 and 3) were used with a Type I error of 10% and Type II error of 7% to determine sample sizes for 20%, 40%, and 75% TW flaws. Sample sizes were chosen for 0, 20%, 40%, and 75% TW flaws so that the overall Type I and Type II errors would be 0.01% and 25%, respectively. Monte Carlo simulations (described in Section 5), were utilized to investigate the actual Type I and Type II errors and to see the effects on the error of using alternative acceptable/unacceptable POD curves and sample sizes.

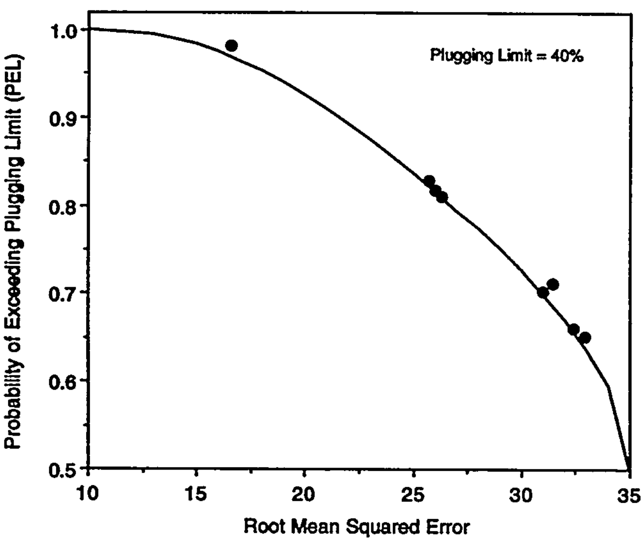

Figure 22. Curve is Theoretical PEL vs. √MSE for 40% Plugging Limit and Defective Tube Defined as ≥ 75% TW Degradation. Data points represent actual performance of Surry round robin teams.

Similar thresholds were selected on the MSE to establish appropriate controls on flaw sizing performance. Figure 22 gives a plot of the root mean squared error (RMSE) versus PEL for teams participating in the Surry round robins. The reader should recall that PEL is the probability of a flaw being sized by an ET system in excess of the plugging limit (40% TW) when the tube is truly defective (i.e., with degradation ≥ 75% TW). Also shown in Figure 22 are results of a theoretical calculation of RMSE versus PEL. From this plot, a value of RMSE = 20 was selected to represent unacceptable sizing performance since this value of RMSE would yield a PEL of about 0.93. In other words, ET systems with RMSE ≥ 20 should fail the sizing test a high percentage of the time. A value of RMSE = 17 was chosen to represent acceptable sizing performance since this would produce a PEL of about 0.96. The sizing test was designed so that ET systems with RMSE ≤ 17 would pass the test a high percentage of the time. It should be emphasized that these values of RMSE were selected on the basis of a 40% TW plugging limit and the definition of a defective tube given in Section 3.1. If another plugging limit is used then different values of RMSE must be specified. The effect of plugging limit on PEL as a function of RMSE is illustrated in Figure 23. The results in this figure show that for a constant PEL, the RMSE required decreases significantly with an increasing plugging limit. For a PEL of 0.95 and plugging limit of 40%, the required RMSE is about 18.2%. For plugging limits of 50%, 60%, and 70%, the required RMSE at the same PEL level are 13.0%, 7.8%, and 2.6%, respectively.

POD and flaw sizing performance characteristics were selected so that a passing ET system would possess an overall ≥ 90% chance of detecting and plugging a defective tube, provided the tube was inspected. This is readily apparent since the acceptable POD performance for defective tubes is ≥ 95% and the acceptable sizing performance is PEL ≥ 95% which results in a joint probability of detecting and correctly calling a tube defective when the flaw size is ≥ 75% of about 90%. In addition, the POD performance of an ET system likely to pass the POD test would be about 90% for

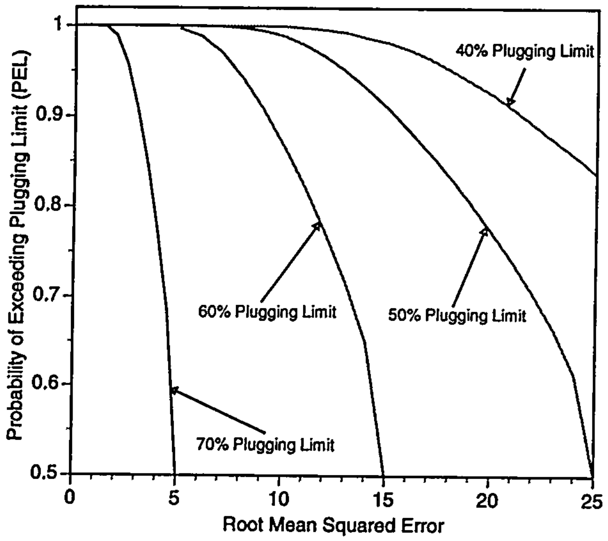

Figure 23. PEL vs. RMSE for Various Plugging Limits

flaws ≥ 40% TW. Sections 5 and 6 of the report present the detailed statistical calculations that were performed to develop performance demonstration tests to meet these design objectives.

4.0 Recommended Performance Demonstration Test

This section describes the recommendations for the performance demonstration test, including a description of the number of tubes to be inspected, the distribution of the flaw sizes, and the methods for grading the POD and sizing performance of the ET systems. The statistical details that were used as the basis for this section are presented in Sections 5 and 6.

4.1 General Structure of Performance Demonstration Test

For a performance demonstration test, the flaw types and locations should simulate those found in operating steam generators. Specifically, the specimen set should be unknown to the personnel taking the test in order for the results to be indicative of ET system reliability. An effective means for simulating the flaws and conditions found in real steam generators would be to construct a tube bundle mockup. Use of a mockup would provide the needed flexibility for evaluating the reliability of new NDE techniques and procedures. To be realistic the mockup must simulate conditions which affect ET inspection reliability such as steam generator internal structure, tubesheet sludge accumulations, deposits on tube surfaces, crevice deposits, and tubing geometry variations.

The matrix of flawed tubes included in the mockup should represent those flaw types and locations associated with known tube damage mechanisms such as:

(a) Wastage/Thinning
(b) Pitting
(c) Fretting/Wear
(d) Stress Corrosion Cracking initiated on either the ID (PWSCC) or OD (ODSCC) of the tube wall surface at various locations
(e) Intergranular Attack (IGA)
(f) Erosion-Corrosion
(g) Fatigue Cracking

Where appropriate, the mockup should combine flaws with other conditions which affect flaw detection and sizing reliability. The mockup should include, but not be limited to the following conditions:

(a) Tube expansion transitions created by rolling, hydraulic or kinetic, methods
(b) Tube bend transitions
(c) Tube support structures and tubesheet simulations
(d) Antivibration bars or spacers
(e) Tubesheet sludge
(f) Crevice deposits
(g) Deposits on tube surfaces

Since all flaw types can not be detected and sized with the same degree of proficiency, it is appropriate to weight the mix of flaw types contained in the mockup toward the most difficult flaw type. This helps to ensure that passing ET systems should be able to detect and size "easier" flaw types with equal or greater proficiency than demonstrated during the test. Consequently, the majority of the flaws in the mockup should be cracks representative of typical morphologies, orientations, and locations. Not only are cracks the most difficult flaw type to detect and size accurately, but they are also the most prevalent form of tube degradation occurring at this time.

Table 4. Minimum Number of Flaws and Blank Grading Units for Tube Mockup

Through-Wall Flaw Depth, %	Number of Samples
0 (Blank)	100
10-30	60
31-60	90
61-90	200
91-100	10

The recommended number and depth range of flaws to be incorporated in the mockup is given in Table 4. These numbers were derived to produce approximate Type I and Type II errors of 10% and 7%, respectively, for flaw detection at each individual flaw size. The statistical basis for these numbers is described in Section 5.

The length range for a given flaw depth and type should bound the range of lengths observed in service-induced flaws. The objective should be to produce a sample set with a realistic range of flaw lengths. In addition, the sample set should be designed to be a challenging test for the inspection system. Toward this end, the sample set should be weighted toward shorter rather than longer flaws, unless a technical basis for including long flaws exists. For example, flaws longer tha : the coil spacing of a bobbin probe represent a difficult inspection problem due to low signal levels when the probe is being operated in differential mode. A rationale exists for including flaws of this type because they challenge the ability of the inspection system and also decrease tube integrity the most. In general, an acceptable length range for most flaw types would be 0.1 inches to 0.5 inches with the caveats noted above.

Assessment of POD and flaw sizing reliability requires knowledge of the true dimensions of each flaw. The processes used for producing and sizing flaws for the mockup should be validated (with respect to size) by destructive metallographic analysis of specimens. Since it is impractical to destructively measure all test specimen flaws, the group of flaws incorporated in the mockup should be nondestructively characterized prior to use for performance demonstrations. Destructive measurements should be made periodically on a percentage of the flaws to verify the accuracy of the techniques used to provide the nondestructive flaw characterization data.

4.2 POD Test Grading Methods

As described in Section 1, the performance demonstration objectives must be formulated as a hypothesis test. The hypotheses are defined in terms of two threshold values. The form of the hypotheses for this test will be:

H_0: $POD_{\text{ET System}}(s) \leq POD_U(s)$ for all flaw sizes $s \geq 20\%$ TW

$POD_{\text{ET System}}(0) \geq POD_U(0)$ for blanks

versus

H_1: $POD_{\text{ET System}}(s) \geq POD_A(s)$ for all flaw sizes $s \geq 20\%$ TW

$POD_{\text{ET System}}(0) \leq POD_A(0)$ for blanks

The $POD_U(s)$ identifies failing performance at each flaw size s (see Table 2) and $POD_A(s)$ identifies passing performance at each flaw size s (see Table 3). The blank specimens are considered to include a flaw of size 0 and dealt with in the same way as the other "flaws" as a way of incorporating false call information into the demonstration test. Systems with an unacceptable POD should fail the test a high percentage of the time and those with acceptable POD to pass a high percentage of the time. The specifics of how to grade the ET system and decide between hypotheses are discussed in Section 5, but are outlined in the following paragraph.

The ET system is graded by 1) estimating the POD curve, 2) calculating 80% confidence limits for the estimated curve (note the 80% confidence limit consists of a ±10% interval about the estimated POD curve), and 3) comparing the lower confidence limit to the threshold (unacceptable) curve, designated as $POD_U(s)$ and shown in Figure 21. A passing POD curve is one with a 80% lower confidence limit which is greater than the curve shown in Figure 21 over the interval 20% to 100% TW. The POD curve and 80% confidence limits may be estimated using the computer program described in Appendices A and B. In addition, the false call rate must be less than or equal to 12% to pass the test.

4.3 Flaw Sizing Test Grading Methods

The sizing test is graded by calculating the root mean squared error of the depth measurements. The minimum number of flaws required is 170, and their sizes should be uniformly distributed over the interval 10% to 90% (with 10 of the total number of flaws being 91% to 100% TW). A subset of the detection test specimen set may be used for this test. A large percentage of the flaws used in the sizing test should be crack-type flaws representative of typical morphologies, orientations, and locations observed in the field. The grading criteria for this and larger sample sizes is given in Table 5. Acceptable RMSE values for 40%, 50%, 60%, and 70% plugging limits are presented in Table 5. The statistical background for this testing method is found in Section 6.

Table 5. Number of Flaws and Critical RMSE Values for Sizing Test

# of Flaws	Acceptable RMSE, %			
	40%*	50%	60%	70%
170	18.20	13.00	7.80	2.50
200	18.35	13.11	7.86	2.62
250	18.52	13.23	7.94	2.65

* = plugging limit on eddy current scale

5.0 Statistical Background for the POD Test

This section provides the specific statistical background for the performance demonstration POD test. The objectives of the performance demonstration POD test have been expressed as hypotheses in terms of two threshold values in Section 4.2.

The specifics of how to decide between hypotheses are discussed in the subsections that follow. First a description of the test is given, then a flaw size distribution is determined. The flaw size distribution will be used as the basis of a Monte Carlo simulation to determine the power of the POD test.

5.1 General Description of Calculations

To evaluate the detection performance of an ET system, the basic strategy is to present the ET inspection system with n flaws that have sizes s_i, $i=1,2,3..n$. These flaws are included within a large set of specimens, such as a tube bundle mockup, which also contains blank (unflawed) specimens. By comparing the inspection results to the true state of the specimens, it is possible to summarize the detection results with a binary variable, Y_i which describes whether or not the ith flaw was detected. (i.e. $Y_i = 1$ if the ith flaw was detected and $Y_i = 0$ if it was not).

The detection test will be constructed so as to use the binary data to estimate the ET system's POD curve and then "compare it" to the POD_U shown in Figure 21 and listed in Table 2. Since the estimated curve for the inspection system cannot be exact, we will surround the curve by a Type I confidence bound and only fail ET systems whose lower confidence bound is at or below the thresholds listed in Table 2.

To construct this test, the most widely used procedure for analyzing binary data is employed, that of logistic regression. The term logistic regression actually refers to a general algorithm that can be used to fit curves to binary data.

A form of the logistic regression curve for this test contains three independent and unknown parameters (it can be generalized to contain any number of parameters), which give the curve enough flexibility to approximate the threshold curve defined by the values in Table 2. The mathematical form of the curve can be expressed as;

$$POD(s;\beta) = \begin{bmatrix} logit(\beta_0+\beta_1 s) \ for \ s<40\% \\ logit(\beta_2+\beta_3 s) \ for \ s\geq 40\% \end{bmatrix} \qquad (8)$$

and the parameters are constrained so that the curve is continuous at 40% TW flaw size. In other words, the above formulation produces a "linear" logistic curve with a possible "kink" in the curve at 40% TW flaw depth. This formulation is explained in more detail in the Appendix B.

5.2 Approximate Flaw Size Distribution Calculations

In order to determine the approximate number of specimens needed for the performance demonstration tests, we examined binomial tests at fixed flaw sizes (20%, 40%, and 75% TW). It is recognized that in an actual performance demonstration test, ET systems would be exposed to a continuum of flaw sizes. Actual flaw sizes would range from a low of 10% up to TW. Flaw lengths would also be variable. However, the sample size determination for the binomial tests should behave approximately like the logistic test since the logistic test also considers binary data as the response, but on a flaw by flaw basis.

A sample size at each TW depth listed in Table 2 must be determined. To determine the sample size n that satisfies a particular set of Type I and II requirements, one must solve the following two binomial equations:

$$Type\ I \geq \sum_{i=1}^{n} \binom{N}{i} (P_U)^i (1-P_U)^{N-i} \qquad (9)$$

and

$$Type\ II \geq 1 - \sum_{i=1}^{n} \binom{N}{i} (P_A)^i (1-P_A)^{N-i} \qquad (10)$$

These equations were solved iteratively and the results for TW depths 20%, 40%, and 75% are presented in Table 6 for Type I = 10% and Type II = 7%. These values of the Type I and Type II error were selected to give a low overall probability of passing for an unacceptable ET system and a relatively high overall probability of passing for an acceptable ET system. Discrete points from the POD_U curve listed in Table 2 and a POD_A curve representing acceptable performance and given in Table 3 were used in these calculations.

Table 6. Sample Sizes for Type I = 10% and Type II = 7%

Through-Wall Flaw Depth, %	Number of Samples
0 (Blank)	100
20	57
40	94
75	203
100*	10

*Type I and Type II limits above do not apply to this category.

The number of blanks to be examined was chosen to be 100 to represent approximately 1/3 of the total number of flaws of size 20% to 75% TW. In general, it is desirable to have 1/3 to 1/2 of the total number of flaws be blanks. Using Equations 9 and 10 to calculate the number of specimens with flaws of 100% TW would give 203. The number of flaws selected to represent the 100% TW category was 10. This number was selected rather than the 203 specimens calculated from Equations 9 and 10 because flaws in the 61%-90% TW range comprise about 56% of the total number of flaws in the test. The large number of flaws in this category serves to indirectly establish flaw detection performance for 100% TW flaws because experience indicates that 100% TW flaws are usually easier to detect than shallower flaws.

5.3 Evaluation of the True Errors of the POD Test

Monte Carlo simulation techniques were utilized in order to evaluate the true errors of the POD test derived in Section 5.2. A fixed sample of flaw sizes was produced according to the sample sizes determined in Section 5.2 for Type I = 10% and Type II = 7%. Specifically, there were 100 blanks, 60 flaws randomly distributed between flaw sizes 10% and 30% TW, 90 flaws randomly distributed between flaw sizes 31% and 60% TW, 200 flaws randomly distributed between flaw sizes 61% and 90% TW, and 10 flaws 91%-100% TW.

There were five unacceptable POD curves and four acceptable POD curves used in the simulations, each representing the true POD of an ET system that might be participating in the performance demonstration test. The "base case" POD for a unacceptable ET system is the pass/fail threshold, POD_U. The "base case" POD for an acceptable ET system is POD_A, for all non-zero flaw sizes and a 5% false call rate. These are listed in Table 7.

There were four other unacceptable ET systems considered. Unacceptable System(US) #2 represents a system that has a better POD than the base case for all flaw sizes, but has an unacceptable false call rate. US #3 represents a system that handles false calls and large flaw sizes well, but has a difficult time detecting the smaller flaw sizes; i.e., performs like the base case for 20% and 40% TW. US #4 represents a system that handles false calls and small flaw sizes well, but has a difficult time detecting the larger flaw sizes; i.e., performs like the base case for 75% and 100% TW. US #5 represents a system that performs like the base case for all flaw sizes except one (20% TW for this case) where it does well.

The probability of detection for the acceptable systems were chosen to represent systems that we would expect to pass during the demonstration tests. These simulations will also help identify any biases that are introduced to the test through the estimation procedure. Acceptable System (AS) #2 represents a system whose POD is above POD_U but slightly worse than POD_A except for flaws of size 100% TW. AS #3 has a constant ability to detect flaws of size 40% TW and greater. It was of interest to see if the modeling techniques would provide confidence bounds that would fail this team a high percentage of the time. AS #4 represented a system whose POD for flaws of size 20% TW was much greater that the pass/fail threshold. This was another test of the modeling techniques.

Table 7. POD Curves Used in Monte Carlo Simulations

Flaw Size, %TW	POD_U	Unacceptable					Acceptable			
		Base	2	3	4	5	Base	2	3	4
0% (blank)	15	15	20	5	5	15	5	5	5	5
20%	15	15	30	15	30	40	30	25	30	65
40%	80	80	90	80	85	80	90	85	95	90
75%	90	90	97	97	90	90	95	95	95	97
100%	90	90	97	99	90	90	95	99	95	97

For each of the true PODs the following steps were taken.

1) The regression parameters for the true POD were calculated based on five knot points at 0, 0.2, 0.4, 0.75, and 1.0.

2) The true POD was calculated for the flaw and then compared to a random uniform value. If the POD value was larger than the random uniform value, then the flaw was designated as found. If the POD value was smaller, then the flaw was not found.

3) A "new" POD curve was calculated based on three knot points at 0, 0.4, and 1.0 TW with the simulated test data and then compared at five points to the POD_U.

4) The POD curve failed if it failed at every knot point. The simulations were run 1000 times and the percentage of times the simulated POD curve did not fail was tabulated.

A compilation of the Monte Carlo results using 70%, 80%, and 90% confidence bounds on the simulated POD curves is given in Tables 8, 9, and 10, respectively.

Based on the information in Table 9, an 80% confidence bound with a sample size of 100 blanks and 360 non-zero flaws provides at most a 2% chance of passing a unacceptable system (Type I error) and approximately a 63% chance of passing the base case acceptable system. These results do not justify reduced sample sizes since that would in turn reduce the probability of passing an acceptable team.

Table 8. Monte Carlo Results Using 70% Confidence Bounds on Simulated POD, Fraction Passing, Number of Simulations = 1000

Teams	Through-Wall Depth (%)					Overall
	0	20	40	75	100	
Base Case	0.610	1	0.002	0.060	0.778	0
US #2	0.066	1	0.370	0.998	1	0.034
US #3	1	0.860	0.018	1	1	0.016
US #4	1	0.992	0.534	0.074	0.282	0.042
US #5	0.236	1	0.116	0.052	0.452	0.002
Base Case	1	1	0.828	0.936	0.924	0.744
AS #2	1	0.974	0.294	0.990	1	0.286
AS #3	1	1	0.996	0.986	0.842	0.838
AS #4	0.984	1	1	1	0.938	0.922

Table 9. Monte Carlo Results Using 80% Confidence Bounds on Simulated POD, Fraction Passing, Number of Simulations = 1000

Teams	Through-Wall Depth (%)					Overall
	0	20	40	75	100	
Base Case	0.508	1	0.002	0.044	0.678	0
US #2	0.044	1	0.272	0.996	1	0.022
US #3	1	0.816	0.008	0.998	1	0.006
US #4	0.998	0.992	0.416	0.042	0.198	0.014
US #5	0.148	1	0.066	0.030	0.356	0
Base Case	1	1	0.764	0.894	0.870	0.628
AS #2	1	0.966	0.198	0.966	1	0.182
AS #3	1	1	0.988	0.970	0.744	0.730
AS #4	0.972	1	0.998	1	0.894	0.866

Table 10. Monte Carlo Results Using 90% Confidence Bounds on Simulated POD, Fraction Passing, Number of Simulations = 1000

Teams	Through-Wall Depth (%)					Overall
	0	20	40	75	100	
Base Case	0.344	1	0	0.016	0.528	0
US #2	0.024	1	0.150	0.982	0.992	0.006
US #3	0.996	0.710	0.006	0.996	1	0.004
US #4	0.992	0.986	0.270	0.018	0.108	0.002
US #5	0.080	1	0.038	0.006	0.216	0
Base Case	1	0.998	0.626	0.780	0.754	0.392
AS #2	0.998	0.954	0.104	0.928	0.982	0.094
AS #3	0.996	1	0.972	0.908	0.616	0.586
AS #4	0.928	1	0.996	0.996	0.754	0.686

6.0 Sizing Test

Test design calculations were also performed to determine the number of test specimens required to demonstrate that ET system capability for flaw sizing would result in about a 95% chance of calling a tube defective when the true flaw size was ≥ 75% TW and the plugging limit was 40% TW. It was assumed that the objective of the test is to distinguish between two hypotheses of the form:

H_0: $MSE > (\sigma_u)^2$

versus

H_1: $MSE < (\sigma_a)^2$

where sizing performance is measured by the MSE (see Section 2.2). The MSE is defined by the formula

$$MSE = \frac{1}{n} \sum_{i=1}^{n} (M(t_i) - t_i)^2 \qquad (11)$$

where $M(t_i)$ is the measured flaw size and t_i is the true flaw size. The MSE for a particular flaw size is related to the bias and standard deviation according to the formula

$$MSE = \sigma^2 + B^2 \qquad (12)$$

The test is defined in terms of two threshold values σ_u, which identifies unacceptable performance and σ_a, which identifies acceptable performance. The relationship of these thresholds to PEL is discussed in Section 3.4 and plotted in Figure 22. The test was designed so that ET systems with unacceptable MSE would fail the test a high percentage of the time and those with acceptable MSE would pass the test a high percentage of the time.

The test is conducted by having the ET system size n flaws. It should be noted that there would not have to be two separate performance demonstration tests, one for detection and one for sizing. ET systems could inspect one tube mock-up and be required to report both detections and size those flaws detected. The flaw sizes are uniformly distributed within the sizing region of interest. The ET system passes if the MSE is less than the critical value, c and fails otherwise. The objective of these calculations is to determine reasonable values for n and c. Table 11 presents sample size requirements for Type I = 5% and Type II = 10%.

It is evident that the number of samples needed to conduct an adequate performance demonstration test for flaw sizing is considerably smaller than for the POD test.

Table 11. Sample Size Requirements for Flaw Sizing Test, Type I = 5%, Type II = 10%

σ_u	Type I Pr(Pass)	σ_a	Type II Pr(Pass)	Number of Samples	Pass Criteria
20	0.05	17.00	0.90	170	18.20
20	0.05	17.25	0.90	200	18.35
20	0.05	17.50	0.90	250	18.52

7.0 Summary and Conclusions

Statistically based performance demonstration qualification requirements have been developed to ensure that field ET inspection systems can reliably detect and size all of the known forms of tube damage that occur in operating steam generators. For this work the goal of steam generator tube ISI was to identify most or all defective tubes which could fail during reactor operation. An extensive data base on the failure pressure of degraded steam generator tubes as a function of flaw type and size was utilized to define a defective tube as one with degradation ≥75% TW (Alzheimer 1979 and Kurtz et al. 1988).

Information from a study on the reliability of ET systems to detect and size service-induced tube degradation, coupled with results from an effort to develop and evaluate sampling plans for ISI was used to select thresholds on POD performance, flaw sizing accuracy, and the false call rate (Kurtz et al. 1990 and Bowen et al. 1989). Thresholds were selected such that an ET system likely to pass the test would have a 90% composite probability of detecting and plugging a defective tube, provided the tube was inspected. Thresholds were also established for degraded but not defective tubes because current and proposed ISI sampling plans rely on detection of lower levels of tube degradation to trigger additional inspection, and to call attenation to conditions which may require corrective actions to mitigate further tube damage.

The POD, flaw sizing, and false call rate thresholds were used in statistical test design calculations to determine the appropriate number and size distribution of flawed steam generator tube samples that would be needed in a steam generator tube bundle mockup to ensure reliable ET inspection system performance. Binomial calculations and Monte Carlo simulations were performed for mockups containing different numbers and variations of a particular distribution of flawed tube samples to determine the probability of an acceptable ET system failing the test and for an unacceptable ET system passing the test. For the POD test a mockup consisting of 360 flawed tube samples would be needed to meet the performance goals selected. A computer program has been developed for grading the POD test. For the flaw sizing test only about 170 flawed tube samples are needed to establish acceptable sizing performance.

The acceptable RMSE level needed depends on the definition of a defective tube and the plugging limit selected. As the plugging limit is increased for the same defective tube definition, the flaw sizing accuracy must improve considerably. For a plugging limit of 40% and defective tube definition of 75%, the required RMSE is 18.2%. For plugging limits of 50%, 60%, and 70%, the required RMSE is 13.0%, 7.8%, and 2.6%, respectively.

8.0 References

Alzheimer, J. M., R. A. Clark, C. J. Morris, and M. Vagins. 1979. *Steam Generator Tube Integrity Program Phase I Report*. NUREG/CR-0718, PNL-2937, Pacific Northwest Laboratory, Richland, Washington.

Bowen, W. M., P. G. Heasler, and R. B. White. 1989. *Evaluation of Sampling Plans for Inservice Inspection of Steam Generator Tubes: Part I*. NUREG/CR-5161, PNL-6462, Pacific Northwest Laboratory, Richland, Washington.

Electric Power Research Institute. 1992. *PWR Steam Generator Inspection Guidelines: Revision 3, 1992*, EPRI NP-6201, Prepared by the Electric Power Research Institute, Palo Alto, California.

Hanlen, R.C. 1990. *Evaluation of Sampling Schemes for In-Service Inspection of Steam Generator Tubing*, EPRI NP-6774, Prepared by Battelle Pacific Northwest Laboratories, Richland, Washington.

Kurtz, R. J., J. M. Alzheimer, R. L. Bickford, R. A. Clark, C. J. Morris, F. A. Simonen, and K. R. Wheeler. 1988. *Steam Generator Tube Integrity Program Phase II Final Report*. NUREG/CR-2336, PNL-4008, Pacific Northwest Laboratory, Richland, Washington.

Kurtz, R. J., R. A. Clark, E. R. Bradley, W. M. Bowen, P. G. Doctor, F. A. Simonen, and R. H. Ferris. 1990. *Steam Generator Tube Integrity Program/Steam Generator Group Project - Final Summary Report*. NUREG/CR-5117, PNL-6446, Pacific Northwest Laboratory, Richland, Washington.

Kurtz, R. J., P. G. Heasler, and D. B. Baird. 1994. *Evaluation of Sampling Plans for In-Service Inspection of Steam Generator Tubes: Comprehensive Analytical and Monte Carlo Simulation Results for Several Sampling Plans*. NUREG/CR-5161, PNL-6462, Vol. 2, Pacific Northwest Laboratory, Richland, Washington.

U.S. Nuclear Regulatory Commission. 1975. *Inservice Inspection of Pressurized Water Reactor Steam Generator Tubes*. Regulatory Guide 1.83, Rev. 1., Washington, D.C.

Appendix A: Functional Form for POD Model

The functional form of the logistic regression curve contains three unknown and independent parameters. The mathematical form of the curve can be expressed as;

$$POD(s,\beta) = \begin{bmatrix} logit(\beta_0 + \beta_1 * s) \; for \; s < 40\% \\ logit(\beta_2 + \beta_3 * s) \; for \; s \geq 40\% \end{bmatrix} \tag{A.1}$$

This formulation is not the most convenient for performing the regression. An equivalent and more useful formulation is;

$$POD(s,\beta) = logit(\beta_0 \psi_0(s) + \beta_1 \psi_1(s) + \beta_2 \psi_2(s)) \tag{A.2}$$

In the above formulas, the term logit(•) represents the logistic function, which is defined by;

$$logit\ (z) = (1 + \exp(-z))^{-1} \tag{A.3}$$

and the functions $\Psi_i(s)$ are hat functions defined so as to produce a piece-wise linear function between 0 and 1 with a "kink" at $s = 40\%$, where 0, 0.4, and 1.0 are called knot points. If the knot points are defined as a vector (k_1, k_2, k_3) where $k_1 = 0$, $k_2 = 0.40$, and $k_3 = 1.0$ for the three knot case, then the specific definitions are;

$$\psi_0(s) = \begin{bmatrix} \dfrac{(k_2 - s)}{(k_2 - k_1)} \; for \; k_1 \leq s < k_2 \\ 0 \; otherwise \end{bmatrix} \tag{A.4}$$

and

$$\psi_1(s) = \begin{bmatrix} \dfrac{(s - k_1)}{(k_2 - k_1)} \; for \; k_1 \leq s < k_2 \\ \dfrac{(k_3 - s)}{(k_3 - k_2)} \; for \; k_2 \leq s < k_3 \\ 0 \; otherwise \end{bmatrix} \tag{A.5}$$

and

$$\psi_2(s) = \begin{bmatrix} 0 \; for \; s < k_2 \\ \dfrac{(s - k_2)}{(k_3 - k_2)} \; for \; k_2 \leq s < k_3 \end{bmatrix} \tag{A.6}$$

This model is fit to data using a logistic regression program, which performs fits of the form:

$$Y = N \; logistic \; (X\beta) + E \tag{A.7}$$

In this equation, Y represents a vector of detections, N a vector describing the number of cases, and X a matrix of independent variables and β the unknown parameter vector. For the model of interest, the matrix X would contain three columns (Ψ_0,Ψ_1,Ψ_2) evaluated at the appropriate sizes s.

Appendix B: A PC-Based Logistic Curve Fitting Program

In order to fit POD curves to inspection data, one can use any one of approximately a dozen different commercial statistical packages (SAS, Splus, Glim, etc.) that include a logistic curve fitting routine. However, for those people who do not want to take the time to learn how to use a statistical package, we have constructed a very simple FORTRAN program that does logistic curve fits. This appendix describes the program, which can be obtained from PNL on disk.

The program IO is very straightforward. The user prepares an ASCII file containing his data and some control parameters. The program reads this file and then produces an ASCII file containing a description of the model fit.

B.1 Operating Instructions

Prepare the ASCII input file containing the data. The first line of the input file contains control parameters, while the following lines contain the actual data. Each line consists of data values separated by commas or blanks (i.e., the lines are free-format). The first line has the format:

Field 1: nn, total number of columns in the data matrix (which follows on line 2).

Field 2: n2, number of independent variables in the problem (i.e., dimension of the X matrix, possibly without the intercept). n2 is always equal to *nn*-2.

Field 3: Column number of the detection variable (variable specifying the number of detections to have occurred).

Field 4: Column number for the number of cases observed.

Field 5: Iteration convergence tolerance (0 sets to default of 1.5e-7). The iteration convergence tolerance is how little the regression parameters must change for iterations to cease.

Field 6: Intercept Switch: 0 - intercept is automatically supplied; 1 - no intercept.

Field 7: Confidence Interval Switch: 0 - compute single confidence intervals for data points; 1 - compute joint confidence intervals.

Field 8: Variability Switch: 0 - standard, assume rms=1; 1 - scale covariance matrix by rms. It is reasonable to scale the covariance matrix by the rms if the data contains non-binomial sources of variability.

Field 9: Output Switch: 0 - do not print input data; 1 - print input data.

Field 10: Output Switch: 0 - do not print output predictions; 1 - print output predictions.

Following this control line, a data matrix consisting of n2 rows and n1 columns should exist in the input file. The program will fit the model:

$$Y = N\ logistic(X\beta) + E$$

where the variable Y represents the number of detections (as specified in Field 3) and the variable N represents the number of observations (as specified in Field 4), while X represents the remaining columns in the data matrix. The remaining columns are considered independent variables in the regression problem.

Some other important features of the program include:

- The program senses EOF after the last data line so there should be no terminating line after the last data line.
- Current capacity limits are 1024 data lines and 16 independent variables, counting the intercept term. Capacity can be changed by altering array limits M1, M2, and M3 in LOGOS.INC.
- The output report is written as a text file identified by the file name and the file extension "OUT." Currently both independent variables and predictions with confidence limits are output to the same file.
- The program calculates confidence limits at α = 80% and 95%.
- Confidence limits are calculated on each input data point.

B.2 Example Input and Output

Here is an example run using this program. The following input file contains 19 observations, with two independent variables in columns 1 and 2.

```
   4    2    3    4   1.5e-7   1   0   0   0   0
   7   1.0   0   10
   7   1.7   0   17
   7   2.2   0    7
   7   2.8   0   12
   7   4.0   0    9
  14   1.0   0   31
  14   1.7   0   43
  14   2.2   2   33
  14   2.8   0   31
  14   4.0   0   19
  27   1.0   1   56
  27   1.7   4   44
  27   2.8   1   22
  27   4.0   1   16
  51   1.0   3   13
  51   1.7   0    1
  51   2.2   0    1
  51   2.8   0    1
```

```
LOGISTIC REGRESSION OUTPUT

  RUN:  LOGI00.DAT

  Control Output
  Number Cases:                 19    Tol:       .1500E-06
  Number Indep. Variables:       3    N-iterations:     7.
  Number rows in "mo":          19    RMS error:         .7773
  Unused:                        0    GOF (-2LL):     13.7092
  Unused:                        0    DOF:            16.
 0Intercept term X1 added by program
  Confidence limits are Single
  (Standard) Covariance matrix called by rms=1
```

OINPUT DATA LISTING

I	X1	X2	X3	Y	N	PHAT
1	1.00	7.00	1.00	.00	10.00	.01
2	1.00	7.00	1.70	.00	17.00	.01
3	1.00	7.00	2.20	.00	7.00	.01
4	1.00	7.00	2.80	.00	12.00	.01
5	1.00	7.00	4.00	.00	9.00	.01
6	1.00	14.00	1.00	.00	31.00	.01
7	1.00	14.00	1.70	.00	43.00	.01
8	1.00	14.00	2.20	2.00	33.00	.01
9	1.00	14.00	2.80	.00	31.00	.01
10	1.00	14.00	4.00	.00	19.00	.02
11	1.00	27.00	1.00	1.00	56.00	.03
12	1.00	27.00	1.70	4.00	44.00	.04
13	1.00	27.00	2.20	.00	21.00	.04
14	1.00	27.00	2.80	1.00	22.00	.04
15	1.00	27.00	4.00	1.00	16.00	.05
16	1.00	51.00	1.00	3.00	13.00	.21
17	1.00	51.00	1.70	.00	1.00	.22
18	1.00	51.00	2.20	.00	1.00	.23
19	1.00	51.00	2.80	.00	1.00	.24

END OF DATA SET

Summary of Iterations

Iter	RMS error	¦¦DBETA¦¦
1	.877538E+00	.507031E+00
2	.367171E+00	.212171E+00
3	.452136E-01	.261312E-01
4	.621610E-03	.359408E-03
5	.326285E-06	.191546E-06
6	.000000E+00	.000000E+00

REGRESSION COEFFICIENTS

Parameter	1	Beta=	-5.659979
Parameter	2	Beta=	.083154
Parameter	3	Beta=	.093852

OCOVARIANCE MATRIX

Param		1	2	3
Param	1	1.37832	-.02348	-.31758
Param	2	-.02348	.00059	.00323
Param	3	-.31758	.00323	.11904

Case	Y	P-HAT	LCL at 95%	LCL at 80%	UCL at 80%	UCL at 95%
1	.000000	.006800	.001484	.002515	.018252	.030583
2	.000000	.007258	.002050	.003177	.016493	.025365
3	.000000	.007604	.002371	.003551	.016208	.024111
4	.000000	.008041	.002495	.003743	.017188	.025601
5	.000000	.008991	.001927	.003287	.024349	.040897
6	.000000	.012105	.003477	.005360	.027104	.041258
7	.000000	.012916	.004845	.006808	.024368	.033975
8	2.000000	.013528	.005505	.007521	.024217	.032858
9	.000000	.014301	.005474	.007639	.026615	.036833
10	.000000	.015978	.003721	.006173	.040720	.065941
11	1.000000	.034859	.014860	.019999	.060083	.079596
12	4.000000	.037138	.020275	.025029	.054776	.067065
13	.000000	.038853	.021117	.026110	.057447	.070413
14	1.000000	.041011	.018141	.024109	.068926	.090066
15	1.000000	.045677	.010355	.017402	.114535	.179618
16	3.000000	.209942	.079392	.112935	.356761	.450185

17	.000000	.221046	.082905	.118412	.374817	.471119
18	.000000	.229232	.077079	.114751	.405595	.514348
19	.000000	.239332	064985	.104920	.457857	.587517

END OF DATA SET

B.3 Program Files

The program disk comes with the following files. LOGI.EXE contains the actual program. The disk also contains the source code, for those that would like to modify the program. Such a regression program requires matrix inversion routines, which have been borrowed from "Linpack." This matrix routines are in the file LINPACK.FOR.

LINPACK	FOR	155741	LOGI00	DAT	335	LOGI10	DAT	335
LOGI01	DAT	335	LOGI11	DAT	335	BLAS	FOR	42906
BLAS	OBJ	13564	LINPACKA	OBJ	61498	LINPACKB	OBJ	70775
LOGOS	R	5556	LOGI	FOR	24498	LOGI	OBJ	15420
LOGI	EXE	374192	LOGISTIC	DOC	4226	LOGOS	INC	1804
LOGOS	FOR	11786	LOGOS	OBJ	6789	LOGI00	OUT	3799
LOGI01	OUT	3801	LOGI10	OUT	3799	LOGI11	OUT	3801
MAKEFILE	BAT	39						

Distribution

No. of
Copies

OFF-SITE

J. Muscara
Nuclear Regulatory Commission
Mail Stop T-10E10
Washington, DC 20555

K. J. Karwoski
Nuclear Regulatory Commission
Mail Stop 7 D4
Washington, DC 20555

E. L. Murphy
Nuclear Regulatory Commission
Mail Stop 7 D4
Washington, DC 20555

J. Strosnider
Nuclear Regulatory Commission
Mail Stop 7 D4
Washington, DC 20555

E. J. Sullivan, Jr.
Nuclear Regulatory Commission
Mail Stop 7 E23
Washington, DC 20555

I. Barnes
Nuclear Regulatory Commission
Region IV
611 Ryan Plaza Drive
Suite 1000
Arlington, TX 76011

J. Blake
Nuclear Regulatory Commission
Region II
101 Marietta Street
Suite 2900
Atlanta, GA 30323

No. of
Copies

J. P. Durr
Nuclear Regulatory Commission
Region I
475 Allendale Road
King of Prussia, PA 19460

N. Economos
Nuclear Regulatory Commission
Region II
Division of Reactor Safety
101 Marietta St., NW
Suite 2900
Altanta, GA 30303

E. H. Gray
Nuclear Regulatory Commission
Region I
475 Allendale Road
King of Prussia, PA 19460

J. Jacobson
Nuclear Regulatory Commission
Region III
801 Warrenville Road
Lisle, IL 60532-4351

M. C. Modes
Nuclear Regulatory Commission
Region I
475 Allendale Road
King of Prussia, PA 19460

D. C. Adamonis
WesDyne International
Murry Corporate Park
1002 Corporate Drive
Export, PA 15632

F. L. Becker
EPRI NDE Center
PO Box 217097
Charlotte, NC 28212

M. Behravesh
Electric Power Research Institute
P. O. Box 10412
Palo Alto, CA 94303

N. R. Bentley
Tennessee Valley Authority
STC-1I-SQ
PO Box 2000
Soddy Daisy, TN 37379

R. L. Beverly
Houston Lighting & Power Company
S. Texas Project Electric Gen. St.
PO Box 289
Wadsworth, TX 77483

A. L. Bradshaw
Westinghouse Savannah River Co.
Mail Zone 233-6H
Aiken, SC 29808

S. Brown
Consultant
5345 W. Pike
Zanesville, OH 43701

C. B. Cheezem
Mail Code MG02ND
Duke Power
13225 Hagers Ferry Road
Huntersville, NC 28078

W. T. Clayton
Southwest Research Institute
6220 Culebra Road
San Antonio, TX 78228

C. D. Cowfer
Westinghouse Savannah River Oo
MS 703-13C
PO Box 616
Aiken, SC 29802

B. Curtis
Allen Nuclear Associates, Inc.
930F Calle Negacio
San Clemente, CA 92673

F. J. Dodd
WesDyne International
5052 Commercial Circle
Concord, CA 94520

T. N. Epps
Contractor
2628 Buttewoods Dr.
Birmingham, AL 35242

P. Fisher
Hartford Steam Boiler Insp. Co.
2443 Warrensville Road
Lisle, IL 60532

H. B. Garland
Entergy Operations
PO Box B
Killona, LA 70066

G. Gladney
Duke Power Company
MG03TD
13339 Hagers Ferry Road
Huntersville, NC 28708

D. O. Henry
General Electric Co.
PO Box 5220
Pleasanton, CA 94566

H. Houserman
Zetec, Inc.
PO Box 140
Issaquah, WA 98027

M. R. Hum
Consultant
15520 Quince Ridge Lane
North Potomac, MD 20878

J. J. McArdle, III
Duke Power Co.
13325 Hagers Ferry Rd. MG 2nd
Huntersville, NC 28078

V. H. Hight, Jr.
V. H. Hight Consulting
2210 Devon
Montrose, CO 81401

S. N. Liu
SAIC
94 Esparito Ave.
Fremont, CA 94539

D. Malinowski
Westinghouse NSD Waltzmill Site
PO Box 158
Madison, PA 15230-0355

R. D. McGuire
The National Board
1055 Crupper Ave.
Columbus, OH 43229-1183

T. Olson
Wisconsin Public Service Corp.
Kewaune Nuclear Power Plant
N 490 Hwy 42
Kewaunee, WI 54216-9510

C. R. Osman
Carolina Power & Light
PO Box 327
HE&EC-MQS
New Hill, NC 27562

S. A. Redner
Northern States Power
414 Nicolett Mall IN03
Minneapolis, MN 55401

M. F. Sherwin
MQS Inspection, Inc.
2301 Arthur Ave.
Elk Grove Village, IL 60007

S. Slosnerick
Toledo Edison
MS 1056
300 Madison Ave.
Toledo, OH 43652

A. L. Smith
WPPSS, MD 1029
3000 George Wash. Way
Richland, WA 99352

D. Spake
EPRI NDE Center
1300 Harris Blvd.
Charlotte, NC 28262

J. C. Spanner, Jr.
EPRI NDE Center
1300 Harris Blvd.
Charlotte, NC 28262

E. Throckmorton
Virginia Power
5000 Dominion Blvd.
Glen Allen, VA 23060

C. Welty
Electric Power Research Institute
P. O. Box 10412
Palo Alto, CA 94303

C. J. Wirtz
Cleveland Electric Illuminating Co.
730 Meadowlark Rd.
Painesville Twp, OH 44077

FOREIGN

L. von Bernus
Siemens AG
Postfach 3220 (UB KWU)
D-91050 Erlangen
Germany

C. Birac
EDF GDL
Allee Privee 21 - Carrefour Pleyel
F-93206 St. Denis Cedex
France

K. Bowker
Nuclear Electric
Berkeley Technology Centre
UK-Gloucestershire GL 13 9PB
England

A. Garcia Bueno
Tecnatom S. A.
Avda Monles de Oca 1
E-28709 San Sebastian de Los Reyes/Madrid
Spain

R. Comby
EDF GDL
Allee Privee 21 - Carrefour Pleyel
F-93206 St. Denis Cedex
France

S. Crutzen
Commission of the European Communities
Joint Research Centre
I-21020 Ispra (Varese)
Italy

A. Lipponen
VTT - Technical Research Centre of Finland
P. B. 1704 (Metals Laboratory)
FIN-02044 VTT
Finland

G. Macgia
ENEL DCO
Via Nino Bixio 39
I-29100 Piacenza
Italy

Y. Miyake
Mitsubishi Heavy Industries Ltd.
Kobe S. & M. Works 1-1
1-Chome Wadasaki-Cho Hyogo-Ku
J-Kobe Zpcode 652
Japan

R. W. Nichols
Squirrels
35 Cockney Hill
Tilehurst, Reading
Berkshire
Great Britian RG3 4HF

J. Perez Prat
Joint Research Center of CEC
P. B. 2
NL-1755 ZG Petten
The Netherlands

O. L. Zanella
ENEL DCO
Via Nino Bixio 39
I-29100 Piacenza
Italy

ON-SITE

21 Pacific Northwest Laboratory

CM Anderson, K7-34
SR Doctor, K5-26
RH Ferris, K5-26
PG Heasler, K7-34
RJ Kurtz (10), K5-26
JC Spanner, K5-26
Publishing Coordination
Technical Report Files (5)

NRC FORM 335 (2-89) NRCM 1102, 3201, 3202

U.S. NUCLEAR REGULATORY COMMISSION

BIBLIOGRAPHIC DATA SHEET

(See instructions on the reverse)

1. REPORT NUMBER (Assigned by NRC, Add Vol., Supp., Rev., and Addendum Numbers, if any.)

NUREG/CR-6227
PNNL-9433

2. TITLE AND SUBTITLE

Performance Demonstration Tests for Eddy Current Inspection of Steam Generator Tubing

3. DATE REPORT PUBLISHED

MONTH	YEAR
MAY	1996

4. FIN OR GRANT NUMBER

B2097

5. AUTHOR(S)

R.J. Kurtz, P.G. Heasler, C.M. Anderson

6. TYPE OF REPORT

7. PERIOD COVERED *(Inclusive Dates)*

10/89-10/92

8. PERFORMING ORGANIZATION — NAME AND ADDRESS *(If NRC, provide Division, Office or Region, U.S. Nuclear Regulatory Commission, and mailing address; if contractor, provide name and mailing address.)*

Pacific Northwest National Laboratory
Richland, WA 99352

9. SPONSORING ORGANIZATION — NAME AND ADDRESS *(If NRC, type "Same as above"; if contractor, provide NRC Division, Office or Region, U.S. Nuclear Regulatory Commission, and mailing address.)*

Division of Engineering Technology
Office of Nuclear Regulatory Research
U.S. Nuclear Regulatory Commission
Washington, DC 20555-0001

10. SUPPLEMENTARY NOTES

J. Muscara, NRC Project Manager

11. ABSTRACT *(200 words or less)*

This report describes the methodology and results for development of performance demonstration tests for eddy current (ET) inspection of steam generator tubes. Statistical test design principles were used to develop the performance demonstration tests. Thresholds on ET system inspection performance were selected to ensure that field inspection systems have a high probability of detecting and correctly sizing tube degradation. The technical basis for the ET system performance thresholds is presented in detail. Statistical test design calculations for probability of detection and flaw sizing tests are described. A recommended performance demonstration test based on the design calculations is presented. A computer program for grading the probability of detection portion of the performance demonstration test is given.

12. KEY WORDS/DESCRIPTORS *(List words or phrases that will assist researchers in locating the report.)*

eddy current, steam generators, tube inspection, performance demonstration

13. AVAILABILITY STATEMENT

Unlimited

14. SECURITY CLASSIFICATION

(This Page)

Unclassified

(This Report)

Unclassified

15. NUMBER OF PAGES

16. PRICE

NRC FORM 335 (2-89)

Printed on recycled paper

Federal Recycling Program