Westinghouse Non-Proprietary Class 3

WCAP-14277
Revision 1

◆ ◆ ◆ ◆ ◆ ◆ ◆ ◆

SLB Leak Rate and Tube Burst Probability Analysis Methods for ODSCC at TSP Intersections

Westinghouse Energy Systems

SLB Leak Rate and Tube Burst Probability Analysis Methods for ODSCC at TSP Intersections

TABLE OF CONTENTS

SECTION	PAGE
1.0 Introduction	1-1
2.0 Summary and Conclusions	2-1
3.0 Methods for Projection of EOC Voltage Distributions	3-1
3.1 General Description of Methods	3-1
3.2 BOC Voltage Distributions	3-1
3.3 Voltage Growth Rates	3-3
3.4 NDE Uncertainties	3-4
3.5 Monte Carlo Methods	3-5
3.6 Projected EOC Voltage Distributions	3-6
3.7 Supplementary Considerations Relative to Growth Rates	3-7
4.0 Burst Pressure and SLB Leakage Correlations	4-1
4.1 General Correlation Considerations	4-1
4.2 Database Used for the Correlations	4-2
4.3 Material Properties Considerations	4-2
4.4 Burst Pressure versus Bobbin Voltage Correlation	4-3
4.5 Probability of Leakage Correlation	4-5
4.6 SLB Leak Rate versus Voltage Correlation	4-7
4.7 Inclusion of Future Data in the Correlations	4-9
4.8 References	4-11
5.0 SLB Leak Rate and Tube Burst Probability Analysis Methodology	5-1
5.1 General Methods Considerations	5-1
5.2 Deterministic Methods for Sensitivity Analyses	5-1
5.2.1 Deterministic Estimation of the Total Leak Rate	5-1
5.2.2 Deterministic Estimation of the Probability of Burst	5-4
5.3 Deterministic Analysis Results	5-5
5.4 Monte Carlo Analysis Methodology	5-5
5.4.1 Probability of Leak Simulation	5-7
5.4.2 Leak Rate Versus Bobbin Amplitude Simulation	5-8
5.4.3 SLB Tube Burst Pressure Simulation	5-10
5.5 SLB Leak Rate Analysis Methodology	5-11
5.6 SLB Tube Burst Probability Analysis Methodology	5-12

TABLE OF CONTENTS (Cont.)

SECTION	PAGE
6.0 Example Analysis Results	6-1
6.1 Simulation Code Description	6-1
6.2 Example Correlations and Distributions	6-2
6.3 Deterministic Analysis Results	6-2
6.4 SLB Leak Rate Analysis Results	6-3
6.5 SLB Tube Burst Probability Analysis Results	6-4
Appendix A: Regression Analysis	
A.1 Introduction	A.1
A.2 The Linear Regression Model	A.3
A.3 Consideration of Variable Error	A.5
A.4 Detection of Outliers	A.7
A.5 Selection of a Regression Coordinate System	A.8
A.6 Selection of a Regression Direction	A.9
A.7 Significance of the Regression	A.11
A.8 Analysis of Regression Residuals	A.13
A.9 References	A.14

SLB Leak Rate and Tube Burst Probability Analysis Methods for ODSCC at TSP Intersections

1.0 Introduction

The purpose of this report is to document the methods[1] used for the analyses supporting application of alternate plugging criteria (APC), also known as alternate repair criteria (ARC), for the disposition of outside diameter stress corrosion cracking (ODSCC) indications detected in steam generator (SG) tubes at locations corresponding to the elevations of the tube support plates (TSPs). Using this report as a reference for the analysis of the indications in a specific plant's SGs obviates the need to document the analysis methods in the plant specific report. These methods are intended to be in accord with the Nuclear Regulatory Commission's (NRC) generic letter (GL) 95-05, "Voltage-Based Repair Criteria for Westinghouse Steam Generator Tubes Affected by Outside Diameter Stress Corrosion Cracking," which was issued on August 3, 1995.

The eddy current inspection (ECT) of the SG tubes may identify a significant number of bobbin coil indications at the intersections of the tubes with the TSPs, of which, several may be confirmed as being axial crack-like ODSCC indications using rotating pancake coil (RPC) inspection techniques. The use of traditional repair or plugging criteria, e.g., 40% depth, is overly conservative for this mode of degradation, and can result in significant tube repairs or plugging that would not be required to be performed if alternate plugging or repair criteria, e.g., the repair guidelines in the NRC's draft Regulatory Guide (RG) 1.121, were used instead. Specific plants may request a change to their Technical Specification to implement an alternate plugging criteria (APC)[2] for the disposition of those indications. This alternate criteria consists of a bobbin amplitude, i.e., voltage, based repair limit in lieu of a depth based repair limit.

The methodology to support the implementation of APC consists of establishing correlations between the expected burst pressure, the probability of leak, and the expected leak rate to the bobbin voltage of the indication. The correlations are then used in conjunction with a measured or calculated end-of-cycle (EOC) distribution of indications to estimate the likelihood of a tube burst and the primary-to-second-

---

[1] The simulation methods described in this document were previously described in WCAP-14046 (Proprietary), Revision 1, "Braidwood Unit 1 Technical Support for Cycle 5 Steam Generator Interim Plugging Criteria," Westinghouse Electric Corporation, August 1994.

[2] This is also known as an interim plugging criteria (IPC) when submitted for implementation for a limited time period, e.g., one fuel cycle.

ary total leak rate for the SG during a postulated steam line break (SLB) event. If the probability of burst is sufficiently small, and if the total estimated leak rate, at a specified confidence level, is less than acceptable limits the voltage criterion may be implemented. If either of the requirements is not met, additional tubes would be repaired until both of the requirements would be projected to be met at the EOC.

The data used in the correlations are to be based on the latest available Electric Power Research Institute (EPRI) database (including data obtained within approximately six months prior to the inspection outage at which the APC will be applied) and which have been approved for use by the NRC. The actual database used should be referenced or documented in the plant specific APC report. Any departures from the NRC approved EPRI database should be documented, and approval for use of the modified database should be obtained from the NRC. An example of a modification to the database is direction from the NRC staff to include data which might otherwise be excluded, or newly developed data which might not be formally incorporated into the EPRI database.

The evaluations supporting the application of plant specific APC are based upon the bobbin coil voltage amplitude, which is correlated with tube burst capability and leakage potential. For SLB leakage and burst analyses, the tube support plate crevices are assumed to be free span or open crevices. This assumption leads to more conservative leak rates and burst probabilities when compared to rates and probabilities associated with expected packed crevices under normal and accident conditions. If APC based on limited TSP displacement are implemented, which may only follow NRC approval to do so, future revisions of this report will include delineation of the analyses methods used to support those criteria.

Revison 1 of this report was prepared to incorporate NRC staff comments transmitted to the Indiana Michigan Power Company in October of 1996.[3] In addition, some typographical errors were corrected and minor changes were made to facilitate understanding. The latter changes also included recommendations informally obtained from Steve Long during telephone discussion with members of the NRC's staff shortly after the initial issue of the document.

---

[3] "Request for Additional Information Related to D.C. Cook's Amendment Request to Incorporate 2 Volt Steam Generator Tube Support Plate Repair Criterion (TAC No. M95894)," J. Hickman (October 28, 1996).

2.0 Summary and Conclusions

This report presents methods used for the evaluation of data gathered to support the application of APC to the tubes in the SGs at nuclear power plants. The methods documented for the performance of correlation analyses are based on standard methods described in references on statistics and regression analysis. The correlations used to support APC are based on estimating the parameters of a correlating equation based on the principles of maximum likelihood.

Methods presented herein for the performance of Monte Carlo simulations reflect the conclusions reached from discussion with the NRC and its advisors on the appropriate techniques to be employed to properly account for the variances and covariances of the parameters of the correlations used. The simulations thus performed are expected to lead to conservative and reliable estimates of the total EOC SG leak rate and of the probability of burst of EOC indications during a postulated SLB.

Per the NRC generic letter, estimatied voltage growth rates for the next cycle of operation are to be based on the voltage growth rates for previous cycles of operation. If only one cycle of previous operating data is available, it may be used to estimate the growth during the next cycle of operation. If two cycles of previous operating data are available, the data for the cycle with the higher growth rates should be used. There may be technical justification for using the most recent growth rate if it can be shown to be lower for cause, however, this would not be consistent with currently approved practice and use would require specific NRC staff approval. Growth rates may be SG specific. Growth rates observed at prior cycles are used to create a cumulative percentage distribution of growth rates. The distribution is linearly scaled to the length of the cycle to be projected.

EOC voltage distributions are obtained from the BOC distributions by Monte Carlo simulation of the NDE uncertainties and the voltage growth rates. The simulation of many thousands of distributions for a single SG are combined to provide a single predicted distribution to be expected at the EOC. Comparisons of predicted distributions with actual distributions after a cycle of operation has shown the ability of the simulation technique to result in conservative estimates of the EOC indications.

3.0 Methods for Projection of EOC Voltage Distributions

EOC voltage distributions are projected from an estimated BOC voltage distribution using a voltage growth model developed from tracking the growth of indications during previous cycles. This section is to describes the methods used to estimate the BOC distribution, the techniques for specifying the voltage growth distribution, the estimation of NDE uncertainties relative to the BOC distribution of indications, the methods employed to simulate growth, and the development of a final EOC distribution from the simulation results.

3.1 General Description of Methods

The progression of ODSCC indications at the TSPs is determined by reevaluation of prior inspection ECT records at the locations identified with indications in the prior inspection. In most cases, some element of a precursor is identified as corresponding to the flaw signal reported in the prior inspection. However, it should be noted that rather conservative analysis criteria are invoked to accomplish this task. In this process, analysts are required to forego the behavior criteria they may have employed to screen out low signal-to-noise indications, and to report possible flaw-like behavior in the TSP mix residual regardless of clarity. Review of the growth data identifies any anomalous growth data, and these are subjected to further scrutiny to eliminate spurious data.

3.2 BOC Voltage Distributions

The bobbin voltage distribution for the beginning of the next cycle (BOC) following the current outage is developed by applying a probability of detection (POD) to all indications found at the end of the previous cycle (EOC), relative to the current outage. This method of accounting for probability of detection is per the direction of the NRC draft NUREG-1477, the draft generic letter, and the January 18, 1995, NRC/industry meeting on the resolution of comments to the draft generic letter. This methodology divides the EOC voltage distribution by the POD, and then subtracts the repaired indications to define the BOC distribution. The number of indications that are to be considered as being returned to service, N, is,

$$N = N_d + N_{nd} - N_r = N_d + \frac{1 - \mathrm{POD}}{\mathrm{POD}} N_d - N_r = \frac{N_d}{\mathrm{POD}} - N_r , \tag{3.1}$$

where,

$$N_d = N_{rc} + N_{rni} + F_{rci} N_{rndd} , \tag{3.2}$$

and, N_r = number of repaired indications,

N_{nd} = number of indications not detected by the bobbin inspection,

POD = probability of detection,

N_{rc} = number of detected bobbin indications that were confirmed by RPC,

N_{rni} = number of detected bobbin indications that were not inspected by RPC,

F_{rci} = fraction of RPC NDD indications called NDD at one inspection and found to be RPC confirmed indications at the subsequent inspection,

N_{rndd} = number of detected bobbin indications not confirmed as flaw indications by RPC inspection.

The above adjustments for POD are incorporated in the BOC and EOC voltage distributions so that no further adjustments are required for the leakage calculation.

The value used for the POD in equation (3.1) is 0.6 unless an alternate value or a voltage dependent POD is approved by the NRC for APC applications. This is consistent with the requirements given in the NRC generic letter. It is noted that when voltage indications above a few volts are found from the inspection, for which the POD would be expected to be significantly greater than 0.6, this methodology becomes very conservative for determining the BOC distribution. This is because it leaves 0.7 of an indication to be simulated as in service for each indication found from the inspection, independent of the voltage level of the indication.

The value for F_{rci} is currently required to be 1.0 until an alternate value is approved for used by the NRC staff. It has been recommended by Westinghouse, but not approved by the NRC staff, that the value of F_{rci} should be based on plant specific inspection results, except as noted below, when data to evaluate F_{rci} are available for at least two cycles of operation with inspections performed using ECT analysis guidelines consistent with those used for APC inspections. The value used for F_{rci} on a plant specific basis must be obtained with *reasonable conservatism*. An example of reasonable conservatism is to apply the largest value obtained for any SG over the last two operating cycles. A population of ≥50 RPC NDD indications is judged to be necessary to reasonably estimate F_{rci}. If adequate plant specific data are not available, the value used for F_{rci} should be selected to bound results obtained from assessments of APC/IPC inspections at other domestic plants with the same tube size. Based on four APC/IPC assessments conducted through 1994, a minimum, plant non-specific value for F_{rci} would be 0.25. Plant specific reports should clearly identify the value and basis for the value of F_{rci} used in the analyses. It is reiterated that this value is currently 1.0 unless another value is subsequently approved for use by the NRC staff.

3.3 Voltage Growth Rates

The plant specific report should identify the operational periods for which growth values were determined. The distribution of the growth rate data, expressed as volts difference in the amplitude readings for two inspections, are usually tabulated in 0.1 volt bins. The width of a bin may be increased to 0.2 volts at some voltage level, and may be further increased to 0.5 volts for voltages above a certain threshold, up to the maximum observed change. For each bin the number of indications is entered along with the corresponding cumulative probability of occurrence value. The bin identification value represents the highest voltage level in the bin. For example, an amplitude of 1.35 volts would be included in the next highest 0.1 volt bin, i.e., 1.4 volts. The cumulative probability is calculated as the sum of indications up to and including the bin value divided by the total number of indications. While the raw data used in developing the voltage growth distribution may contain indications for which negative growth would be calculated, the developed growth distribution is not permitted to exhibit any negative growth characteristics. This is achieved by considering all indications with calculated negative growth as zero growth. For use in the Monte Carlo analyses, the voltage growth distributions may be normalized to a growth per EFPY basis by assuming growth is linear with time. In general, the cycle time to be projected is similar to that for which the growth rates were calculated and normalization is not considered significant.

The voltage growth histograms for each of the operational periods evaluated should be presented in the plant specific APC report. These may appear as composite, i.e., all SGs, for the prior cycles, but should be presented for each SG for the most recent cycle. For indications with appreciable BOC amplitude readings, i.e., greater than or equal to 0.75 volt, average growth rates may be calculated and compared to the overall growth rates for information purposes: In addition, information may be presented relative to growth as a function of position in the SG, e.g., the dominance of the incidence of ODSSC indications at the elevations of the lower TSPs may also be reflected in the growth rates.

In developing the voltage growth distribution, only NDE indications with flaw indication characteristics in both cycles are to be included in the analysis, except if an indication changes fron non-detectable to a relatively high voltage, e.g., 2.0 volts. A minimum of 200 indications are required to define a voltage distribution. SG specific distributions also require a minimum of 200 indications for application to projecting EOC voltages. For projections of a specific SG, the more conservative (relative to projecting leakage and burst) growth between the distributions for the specific SG, and for all SGs collectively, should be used. If two cycles of growth distributions obtained with APC/IPC ECT guidelines are available, the larger distribution of the last two cycles should be used for the projections. If two cycles of data are not available, the prior cycle growth distribution may be used. If the last cycle of growth is significantly lower than the prior cycle, and can be attributed

to enhancements in secondary chemistry affecting only the last cycle, the last cycle growth rates may be used for the APC/IPC analyses if specific approval from the NRC staff is obtained to do so. The plant specific APC report should include a discussion of the specific justification the basis for the lower growth rates. When an APC is being applied without 200 indications to define the growth distributions, a bounding growth distribution from other domestic plants may be applied.

3.4 NDE Uncertainties

For APC applications, NDE uncertainties must be accounted for in projecting the distribution of the BOC indications to the EOC. This is accounted for by using Monte Carlo techniques. The database supporting NDE uncertainties is described in Reference 5-1, and NDE uncertainties for APC applications are given in the EPRI repair criteria report, Reference 5-2. From Reference 5-2, the NDE uncertainties are comprised of uncertainties

1) due to the data acquisition technique, which is based on use of the probe wear standard, and

2) due to analyst interpretation, which is sometimes called the analyst variability uncertainty.

If a transfer standard is not employed, manufacturing tolerances in the probe calibration standards would be expected to constitute an additional source of uncertainty in the NDE results, and should be accounted for in the plant specific computations.

The data acquisition, or probe wear, uncertainty has a standard deviation of 7.0% about a mean of zero. Variation due to probe wear is restricted to ≤15% of the bobbin amplitude, contingent on the implementation of the probe wear standard requiring probe replacement at 15% differences between new and worn probes. ASME standards cross-calibrated against the reference laboratory standard and the probe wear standard should be implemented during the inspection to avoid the additional consideration of manufacturing tolerances.

The analyst interpretation (analyst variability) uncertainty has a standard deviation of 10.3% about a mean of zero. Typically, this uncertainty would have a computational cutoff at 20% based on requiring resolution of analyst voltage calls differing by more than 20%, however, the NRC has not accepted the 20% limitation on the analyst interpretation uncertainty. Pending a further resolution of this issue with the NRC, the analyst interpretation uncertainty is applied without a cutoff. Thus, for EOC voltage projections, separate distributions are applied for probe wear with a cutoff at 15%, and for analyst interpretation with no cutoff.

3.5 Monte Carlo Methods

The Monte Carlo simulations to estimate the EOC voltage distribution begin with the BOC distribution as described previously. The EOC distribution of indications is calculated several thousand times, accounting for the uncertainties in the NDE and in the voltage growth. The cumulative distribution of indications from all of the simulations is calculated, then adjusted to reflect the total number of indications in the BOC distribution. The methods used to account for the uncertainties are discussed in the following subsections. The calculation of a single representative EOC distribution is discussed in the following section of this report.

3.5.1. NDE Uncertainties

The method of accounting for NDE uncertainties in the Monte Carlo analyses is to adjust the field measured voltage by multiplying the standard deviation of the uncertainty under consideration by a standardized normal distribution deviate. The uncertainty associated with probe wear is assumed to be characterized by a normal distribution with a mean of zero and standard deviation expressed as a percentage of the *true*, but unknown, voltage of the indication. The uncertainty of the analyst is assumed to be characterized by a normal distribution with a mean of zero and a standard deviation expressed as a percentage of the *true probe* voltage of the indication.

Assuming no analyst variability, the distribution of voltages read by a probe, V_p, corresponding to a true indication voltage of V would be simulated as,

$$V_p = V(1 + Z_r \xi_p), \tag{3.3}$$

where $V\xi_p$ is the standard deviation of the probe error about the true voltage of the indication and Z_r is the distribution of standard normal deviates. The measured voltages, V_m, accounting for analyst variability, would then be distributed as a function of the probe voltage as,

$$V_m = V_p(1 + Z_s \xi_a), \tag{3.4}$$

where $V_p \xi_a$ is the standard deviation of the analyst error about the probe voltage of the indication, and Z_s is the distribution of standard normal deviates. Combining equations (3.3) and (3.4), the distribution of true voltage of an indication about the measured voltage of the indication is then,

$$V = \frac{V_m}{(1 + Z_r \xi_p)(1 + Z_s \xi_a)}. \tag{3.5}$$

Thus, the simulation of the true voltage of an indication at the BOC is based on two independent draws from a standard normal distribution. To account for the limit on probe wear, the values of Z_r are limited to an absolute value no larger than 15/7, i.e., values of Z_r are repeatedly drawn until $|Z_r| \leq 15/7$.

Alternatively, if the analyst variability was assumed to be distributed about the *true* voltage of the indication, equation (3.5) would be replaced by,

$$V = \frac{V_m}{(1 + Z_r\xi_p + Z_s\xi_a)} . \tag{3.6}$$

The difference between the two expressions is the addition of a term, $Z_r\xi_p Z_s\xi_a$, in the denominator of equation (3.5) relative to equation (3.6). Since the product of ξ_p with ξ_a is on the order of 0.007, this would be expected to be a second order effect. However, to avoid assessing the impact of the additional term, equation (3.5) is used in the Monte Carlo simulations.

3.5.2 Voltage Growth

To account for voltage growth during an operating cycle, the cumulative distribution of voltage growth is entered with a random number, U_i, drawn from a uniform distribution, i.e., $0 < U_i \leq 1$. The growth is then obtained as a linear interpolation between the discrete values used for the cumulative growth distribution. For example, if growth values were specified only at cumulative probability values of 0.0, 0.5, and 1.0, the growth corresponding to a value of U_i of 0.75 would be midway between the growth values corresponding to 0.5 and 1.0 respectively.

3.6 Projected EOC Voltage Distributions

Monte Carlo simulations are then performed to develop the EOC voltage distributions from the BOC distributions. The BOC voltages are increased by allowances for NDE uncertainties and voltage growth to obtain the EOC values. In the Monte Carlo analyses, each voltage bin of the BOC distributions is increased by a random sample of the NDE uncertainty and growth distributions to obtain an EOC voltage sample. Each sample is weighted by the number of indications in the voltage bin. The sampling process is repeated at least 100,000 times for each BOC voltage bin and then repeated for each voltage bin of the voltage distribution. Since the Monte Carlo analyses yield a cumulative probability distribution of EOC voltages, a method must be defined to obtain a discrete maximum EOC voltage value. The method adopted in this report is to integrate the tail of the Monte Carlo distribution over the largest 1/3 of an indication to define a discrete value with an occurrence of 0.33 indication. For N indications in the distribution, this is equivalent to evaluating the cumulative probability of voltages at a probability of $(N\text{-}0.33)/N$. The largest voltages for all distributions developed by Monte Carlo in this report

have been obtained with this definition for the maximum EOC discrete voltage. The next largest discrete EOC voltage indication is obtained by integrating the tail of the Monte Carlo distribution to one indication and assigning the occurrence of 0.67 indication. This process for developing the largest EOC voltage indications provides appropriate emphasis to the high voltage tail of the distribution and permits discrete EOC voltages for deterministic tube integrity analyses.

3.7 Supplementary Considerations Relative to Growth Rates

It is recognized that specific actions may be taken at an operating plant aimed at slowing the progression of Alloy 600 ODSCC, i.e., there is a relationship between operating chemistry and ODSCC growth. For example, a plant could initiate molar ratio chemistry control and boric acid addition. Operating with elevated sodium to chloride molar ratios enhances the possibility of developing caustic crevice conditions conducive to initiation and propagation of Alloy 600 alkaline stress corrosion cracking. Hideout return chemistry data can be used to measure the success of the molar ratio control program in modifying the steam generator environment. Laboratory and operating PWR plant data indicate the usefulness of boric acid as a contributor to the overall corrosion control program.

The success of such efforts has not been quantified to the extent that adjustment of prior cycle growth data for chemistry enhancements is justified for projecting growth rates for the next cycle. Should such data become available, their use would have to be documented and justified in a plant specific report, or in a future revision to this report and specifically by approved for use by the NRC staff. Current analyses do not include consideration of retarding future ODSCC growth via chemistry control.

4.0 Burst Pressure and SLB Leak Rate Correlations

The purpose of this section is to provide information and justification for all of the correlations developed in support of the application of alternate plugging criteria (APC) for the disposition of ODSCC indications in the SG tubes at the elevations of the TSPs at nuclear power plant.

4.1 General Correlation Considerations

To support the implementation of APC at nuclear power plants, correlations have been developed for tubes containing ODSCC indications at TSP locations between the bobbin amplitude, expressed in volts, of those indications and the free-span burst pressure, the probability of leak, and the free-span leak rate for indications that leak, References 4.1 and 4.2. In 1993, the NRC issued draft NUREG-1477, Reference 4.3, for public comment. The draft NUREG delineated a set of guidelines for criteria to be met for the application of Interim Plugging Criteria (IPC) for ODSCC indications. The criteria guidelines permitted the use of, with adequate justification, a burst pressure to bobbin amplitude correlation and a probability of leak to bobbin amplitude correlation. The criteria guidelines did not permit the use of a leak rate to bobbin amplitude correlation for the estimation of end of cycle (EOC) total leak rates. In essence, References 4.1 and 4.2 provided comments on the Reference 4.3 guidelines. Reference 4.4 provided an NRC response and position relative to resolving the differences between References 4.1 and 4.2, and Reference 4.3, along with responses to other public comments. Of significance to this report, is that Reference 4.4 indicated that a correlation between leak rate and bobbin amplitude could be employed if the correlation could be statistically justified at a 95% confidence level, and provided direction for the development of guidelines, e.g., Reference 4.5, that could then be employed for the identification and exclusion of outlying experimental data. Further delineation of the NRC's position was published as a draft generic letter (GL), Reference 4.6. NRC resolutions of public comments on the draft GL are given in Reference 4.7, with the final GL letter being issued as Reference 4.8. The methods of this report are intended to be consistent with the methodology of References 4.1 to 4.8 with emphasis on the GL itself, i.e., Reference 4.8, and with prior use of these methods as described in Reference 4.11.

Discussions with NRC personnel reviewed potential issues associated with the manner in which the leak rate to bobbin amplitude correlation was being used, and questioned the ability of a deterministic model of the total leak rate to accurately account for the variability due to the uncertainties in the parameters of the correlation model. Thus, the potential leak rate during a postulated steam line break (SLB) is estimated by utilizing both deterministic and Monte Carlo methods. The deterministic method is used tp screen potential leak rates, and the Monte Carlo method is used for the final determination of potential EOC leak rates.

Information is presented in the following sections on the database for the correlations, material properties as related to burst pressure considerations, the correlation of burst pressure to bobbin amplitude, i.e., indication voltage, the correlation between the probability of leak of an indication and the indication voltage, and lastly a discussion of the correlation of leak rate to volts. The use of each of the correlations is also discussed. A general discussion of the linear regression techniques employed is contained in Appendix A to this report.

All of the techniques described for the support of APC implementation are intended to be in accord with the requirements of NRC Generic Letter 95-05.

4.2 Database Used for the Correlations

The baseline database used for the development of the correlations should be presented or properly referenced in the plant specific report(s) supporting the implementation of APC. One such database is presented and discussed in Reference 4.2. Future development of APC criteria for specific plants may include additional data as it becomes available, hence, no referenceable database is included in this report. It is noted that not all additional data may be included in future correlations. The database must comply with the current regulatory requirements regarding approval of the data used for the correlation analyses, see Section 1.0 of this report.

The priorities for identifying the appropriate database are a plant specific NRC Safety Evaluation Report (SER) if used for the cycle of implementation of an APC/IPC, or the guidance of the NRC generic letter if applicable to the plant specific APC/IPC. When new data, such as testing results from recently pulled tube sections, are to be added to the database, the data shall be evaluated against the EPRI data exclusion criteria, Reference 4.5, as approved by the NRC at the time of the APC/IPC assessment. Reference 4.7, includes the status of NRC concurrence with Reference 4.5 at the time of this report.

Any other special circumstances related to data used for the correlations should be discussed or referenced in the plant specific APC report.

CAUTION: The database used in the regression analyses performed for this report was randomly generated and analyzed specifically for this report. Sample correlation results presented herein are for illustration purposes only and are not intended for plant specific APC evaluations.

4.3 Material Properties Considerations

The variation of material properties between tubes is a significant factor in determining the burst pressure. The rupture of tubes fabricated from Alloy 600 material is an elastic-plastic fracture process. A rigorous analysis of the process

would require knowledge of the strain hardening characteristics of the material. However, reasonably accurate predictions may be obtained by empirically correlating the burst pressures obtained from tubes with different material properties by normalizing the results to the flow strength, S_f, of the material. The concept of a flow stress allows the material to be approximated by elastic-perfectly plastic behavior, i.e., at some critical pressure the flanks of the crack deform without bound and the tube ruptures. For a material with no strain hardening capability the flow stress would correspond to the yield strength, S_Y, of the material. In practice the flow strength is taken as some value between the yield stress and the ultimate tensile strength, S_U. For Alloy 600 SG tube material, a flow stress of one-half of the sum of the yield and ultimate strengths has been widely used, thus,

$$S_f = \frac{1}{2}\left(S_Y + S_U\right). \tag{4.1}$$

Alloy 600 material typically exhibits a flow stress on the order of 75 ksi at ambient conditions, hence, test results are usually adjusted to this value for the presentation of the data and the development of the regression equation. Once the correlation has been obtained, it can be scaled by the flow stress to estimate the burst pressure at other temperatures, e.g.,

$$P_B\Big|_{600^\circ F} = P_B\Big|_{70^\circ F} \frac{S_f\Big|_{600^\circ F}}{S_f\Big|_{70^\circ F}}. \tag{4.2}$$

Tube material properties for APC applications are summarized in Table 4-1. While the values presented are not from on a randomized database, they are representative of Westinghouse mill annealed tubing only.

4.4 Burst Pressure Versus Bobbin Voltage Correlation

The bobbin coil voltage amplitude and burst pressure data presented in the EPRI database, Reference 4.2, have been used to estimate the degree of correlation between the burst pressure and bobbin voltage amplitude. The details of performing the regression analysis to determine the degree of correlation and to estimate the parameters of a log-linear relationship between the burst pressure and the bobbin amplitude, are provided in the EPRI database report. General techniques for the performance of regression analysis are described in Appendix A. to this report. The evaluations examined the scale factors for the coordinate system to be employed, e.g., linear versus logarithmic, the detection and treatment of outliers, the order of the regression equation, the potential influence of measurement errors in the variables, and the evaluation of the residuals following the development of a relation by least squares regression analysis. The results of the analyses indicated that an optimum linear, first order relation could be obtained from the regression of

the burst pressure on the common logarithm (base 10) of the bobbin voltage amplitude.

The equation form relating the burst pressure, P_B, of indication i to the logarithm of the bobbin amplitude, V_i, is given by,

$$P_{B_i} = a_1 + a_2 \log(V_i), \tag{4.3}$$

where a_1 and a_2 are least squares estimates of coefficients α_1 and α_2 that would be obtained if the entire population of tubes with indications were tested. Here, the burst pressure is usually measured in *ksi* and the bobbin amplitude is in *volts*.

A typical value for the *index of determination* of the regression of the burst pressure on the bobbin amplitude is 80%. The corresponding correlation coefficient would be 0.90, which is significant at a >99.999% level. This means that the typical *p-value* for the slope of the line is < 0.001%. Hence, equation (4.3) provides an excellent functional form for the prediction of the burst pressure from the bobbin amplitude.

The estimated standard deviation of the residuals, i.e., the error of the estimate, s_p, of the burst pressure is typically on the order of ~0.95 ksi. Examination of the residuals from the regression analyses for 3/4" and 7/8" diameter tubes indicated that they are normally distributed, thus verifying the assumption of normality inherent in the use of least squares regression.

A typical format for reporting the results of the regression analysis is illustrated in Table 4-2. The database used for the analysis and the regression results are shown on Figure 4-1.

Using the regression relationship, a lower 95% prediction bound for the burst pressure as a function of bobbin amplitude is then developed. These values are further reduced to account for the lower 95%/95% tolerance bound for the Westinghouse database of tubing material properties at 650°F. Using the reduced lower prediction bound curve, the bobbin amplitude corresponding to a *free-span* burst pressure of 3657 psi is found.[1] This is the *structural limit* as reported in plant specific APC supporting analyses. A typical value for 3/4" diameter tubes is on the order of 5V. An additional limit corresponding to the actual SLB differential pressure is also calculated and reported.

[1] The value of 3657 psi results from considering a SLB differential pressure of 2560 psi divided by 0.7 in accord with the guidelines of RG 1.121, Reference 4.9.

4.5 Probability of Leakage Correlations

Historically, the probability of leakage has been evaluated by segregating the model boiler and field data into two categories, i.e., specimens that would not leak during a SLB and those that would leak during a SLB. These data were analyzed to fit a sigmoid type equation to establish an algebraic relationship between the bobbin amplitude and the probability of leak. The specific algebraic form used to date has been the logistic function with the common logarithm of the bobbin amplitude employed as the regressor variable, i.e., letting P be the probability of leak, and considering a logarithmic scale for volts, V, the logistic expression is:

$$P(leak \mid V) = \frac{1}{1 + e^{-[\beta_1 + \beta_2 \log(V)]}}. \tag{4.4}$$

This is then rearranged as:

$$\ln\left(\frac{P}{1 - P}\right) = \beta_1 + \beta_2 \log(V), \tag{4.5}$$

to permit an iterative, linear, least squares regression to be performed to find the maximum likelihood estimators, b_1 and b_2, of the coefficients, β_1 and β_2.

The use of the logistic function for the analysis of dichotomous data is standard in many fields. The differential form assumes that the rate of change of the probability of leak is proportional to the product of the probability of leak and the probability of no leak. As noted, the function is sigmoidal in shape, and is similar to the cumulative normal function, and likewise similar to using a probit model (which is a normal function with the deviate axis shifted to avoid dealing with negative values). In principle, any distribution function that has a cumulative area of unity could be fit as the distribution function, a limitless number of possibilities. Trying to identify a latent, or physically based, distribution for the probability of leak would be considered to be unrealistic and unnecessary. For most purposes the logistic and normal functions will agree closely over the mid-range of the data being fitted. The tails of the distributions do not agree as well, with the normal function approaching the limiting probabilities of 0 and 1 more rapidly than the logistic function. Thus, relative to the use of the normal distribution, the use of the logistic function is conservative.

In addition, consideration was given as to whether the bobbin amplitude or the logarithm of the bobbin amplitude should be used. Since the logistic, normal and Cauchy distribution functions are unbounded, the use of volts would result in a finite probability of leak from non-degraded tubes, and would be zero only for V=-∞. By contrast, the use of the logarithm of the voltage results in a probability of leak for non-degraded tubes of zero. Clearly, the second situation is more realistic than

the first, especially in light of the fact that a voltage threshold is a likely possibility.

The log-logistic function falls into a category of models referred to as *Generalized Linear Models* (GLMs). This simply means that the model can be transformed into a linear form, e.g., equation (4.5). The left side of equation (4.5) is referred to as the link function for the model. The parameters of the equation are estimated by fitting the data using an iterative least squares technique, resulting in the maximum likelihood estimate of the parameters.

The results of a typical regression analysis are summarized in Table 4-3. The coefficients of the equation are provided along with the elements of the variance-covariance matrix for the coefficients. In addition, the deviance for the solution is also given. One accepted measure of the goodness of the solution or fit for GLMs is the deviance, given by,

$$D = 2 \sum_{i=1}^{n} \left\{ P_i \ln\left[\frac{P_i}{P(V_i)}\right] + (1 - P_i) \ln\left[\frac{1 - P_i}{1 - P(V_i)}\right] \right\}. \tag{4.6}$$

where P_i is the probability associated with data pair i and $P(V_i)$ is the calculated probability from V_i. The deviance is used similar to the residual sum of squares in linear regression analysis and is equal to the error, or residual, sum of squares (SSE) for linear regression. For the probability of leak evaluation P_i is either zero or one, so Equation (4.6) may be written

$$D = -2 \sum_{i=1}^{n} \left\{ P_i \ln[P(v_i)] + (1 - P_i) \ln[1 - P(v_i)] \right\}. \tag{4.7}$$

Since the deviance is similar to the SSE, lower values indicate a better fit, i.e., for a lower the residual sum of squares, more variation of the data is considered to be explained by the regression equation. Prior to the preparation of this report, analyses were performed to investigate the NUREG-1477 recommended forms for the POL function. The differences in the deviances from the analyses performed were judged to be not numerically significant relative to selecting the best form of a fitting function.

A significant outcome of performing the additional regression analyses was the finding that, taken in conjunction with the leak rate versus voltage correlation, the choice of a probability of leak function is relatively unimportant. For typical APC/IPC voltage distributions, the final total leak rate values obtained using all of the functions tend to differ by only a few percent across the spectrum of POL functions.

An example of the format of reporting the results from the POL regression analysis is provided as Table 4-3. The Pearson standard deviation, σ_{error}, in the table is discussed in Section 5.0 of this report. A plot of the POL database used for the example regression is shown on Figure 4-2, along with the regression curve obtained from the GLM analysis of the data.

4.6 SLB Leak Rate Versus Voltage Correlation

The bobbin coil and leakage data previously reported were used to determine a correlation function between the SLB leak rate and the bobbin amplitude voltage. Since the bobbin amplitude and the leak rate would be expected to be functions of the crack morphology, it is to be expected that a correlation between these variables would exist. Previous plots of the data on linear and logarithmic scales indicated that a linear relationship between the logarithm of the leak rate and the logarithm of the bobbin amplitude would be an appropriate choice for establishing a correlating function via least squares regression analysis. Thus, the functional form of the correlation is

$$\log(Q) = b_3 + b_4 \log(V), \tag{4.8}$$

where Q is the leak rate, V is the bobbin voltage, and b_3 and b_4 are estimates obtained from the data of some coefficients, β_3 and β_4. The final selection of the form of the variable scales, i.e., log-log, was based on performing least squares regression analysis on each possible combination and examining the square of the correlation coefficient for each case. The results of the analyses, using the EPRI database, indicated the appropriate choice of scales to be log-log.

A format for reporting the results from the regression analysis is provided as Table 4-4. The data used for the analysis and regression curve obtained from the analysis are illustrated on Figure 4-3. The example value of r^2 of 59.4% is significant at a level of >99.99% based on an F distribution test of the ratio of the mean square of the regression to the mean square of the error. This can also be interpreted as the probability that the log of the leak rate is correlated to the log of the bobbin amplitude. The p value for the illustrated slope parameter is $1.5 \cdot 10^{-9}$. The conclusion to be drawn from these results is that it is very likely that the variables are correlated. Per GL 95-05, the validity of the regression is judged by the p value associated with the slope. Since this is significantly less than the 0.05 value stipulated in GL 95-05, the regression would be concluded to be valid, and the use of the linear regression results would be acceptable.

The expected, or arithmetic average (AA), leak rate, Q, corresponding to a voltage level, V, was also determined from the above expressions. Since the regression was performed as log(Q) on log(V) the regression line represents the mean of log(Q) as a function of bobbin amplitude. This is not the mean of Q as a function of V. The

residuals of log(Q) are expected to be normally distributed about the regression line. Thus, the median and mode of the log(Q) residuals are also estimated by the regression line. However, Q is then expected to be distributed about the regression line as a log-normal distribution. The regression line still estimates the median of Q, but the mode and mean are displaced. The corresponding adjustment to the normal distribution to obtain the AA of Q for a log-normal distribution is

$$Q = E\{Q \mid V\} = 10^{b_3 + b_4 \log(V) + \frac{\ln(10)}{2}\sigma^2}, \tag{4.9}$$

for a given V, where σ^2 is the estimated variance of log(Q) about the regression line. The variance of the expected leak rate about the regression mean is then obtained from

$$\mathrm{Var}(Q) = Q^2\left[10^{\ln(10)\sigma^2} - 1\right]. \tag{4.10}$$

To complete the analysis for the leak rate, the expected leak rate as a function of log(V) was determined by multiplying the AA leak rate by the probability of leak as a function of log(V). The results of this calculation for an example database are also depicted on Figure 4-3 for a steam line break differential pressure of 2560 psi.

Leak Rate Model when the p-Value is Greater than 5%

The NRC generic letter requires that the dependence of the leak rate to the bobbin amplitude be demonstrated by a rigorous statistical analysis. If the p value obtained from the regression for the slope parameter is less than or equal to 5%, the dependence of the leak rate on the bobbin amplitude is considered to have been demonstrated. There is the potential for the 7/8" tube data to exhibit a p value in excess of 5% depending on the level of application of the data exclusion criteria permitted for a plant specific APC implementation. The p value can be thought of as the probability that the true slope of the correlation is zero even though the value obtained from the regression analysis is other than zero. If the p value is greater than 5%, it must be assumed that a correlation does not exist between the leak rate and the bobbin amplitude.

If the leak rate is considered to be independent of the bobbin amplitude, the leak rate model is,

$$Q_i = \beta_3 + \varepsilon, \tag{4.11}$$

where Q_i again represents the common logarithm of the leak rate. The mean of the data in the database, b_3, is used to estimate β_3, and ε is the estimated error about the mean of the logarithm of the leak rates in the database. Again, ε is assumed to

be distributed such that it has a mean of zero. The standard deviation of the errors is estimated by the standard deviation of the data. For both databases, statistical analyses have been performed that demonstrate, at a greater than 95% confidence level, that the leak rate data are log-normally distributed, independent of correlation considerations. Hence, the errors about the mean of the log leak rate are assumed to be log-normally distributed.

Leak Rate versus Bobbin Amplitude when the p-Value is greater than 5%

If the *p* value from the regression analysis is greater than 5%, it is still possible to rigorously demonstrate a dependence of the leak rate on the bobbin amplitude. In this case, it is assumed that the model is either invalid or may not apply over the entire range of the data. Analyses have been performed by segregating the data at the median value of the voltage range. Using the model of equation (4.11), it has been shown at a greater than 95% confidence level that the leak rates for the lower half of the voltages are not from the same population as the leak rates for the upper half of the voltages. This has been demonstrated for the 3/4" and 7/8" tube data. Thus, it would be possible to select an upper bound voltage to be expected at the EOC, and to use an equation (4.11) model based only on the leak rate data from indications with bobbin amplitudes less than the upper bound. While not currently approved by the NRC for use if the *p* value is greater than 5%, this approach may be pursued in the future. NRC review and approval would be required prior to implementation, and the evaluation method would be documented in a revision to this report.

4.7 Inclusion of Future Data in the Correlations

The initial analyses performed for the implementation of APC verified the validity of the regressions for the burst pressure, the POL, and the leak rate. For each of the correlations, additional verification of the appropriateness of the regression was obtained by analyzing the regression residuals, i.e., the actual variable value minus the predicted variable value from the regression equation. Plots of the residuals as a function of the predicted values was found to be nondescript, indicating no apparent correlation between the residuals and the predicted values. Cumulative probability plots of the residuals on normal probability paper approximated a straight line, thus verifying the assumption inherent in the regression analysis that the residuals are normally distributed. Based on the results of the residuals scatter plots and the normal probability plots, it was concluded that the regression curves and statistics could be used for the prediction of the burst pressure, the POL, and leak rate as functions of the bobbin amplitude of the indications, and for the establishment of statistical inference bounds.

As additional data become available they should be incorporated into the reference database based on passing NRC approved outlier criteria. Verification of continued use of the specified equations may be based on a visual inspection of the data relative to the database and the correlation equations. Analyses of additional data since the original determination of the regression equations has revealed no circumstances where the form of the equations should be questioned. In these cases, the analyses of the residuals does not have to be repeated to justify the use of the equation forms specified in this report. As new data are added to the database and new correlation parameters calculated for the implementation of APC at specific plants, the appropriate tests on the data relative to incorporation should be performed and documented in the report for that plant.

Sample new burst data and the regression curve obtained by including the data in the database are shown on Figure 4-1. The results of the regression analysis performed with the new data included in the database are summarized in Table 4-5, and the results of the regression analysis of the new data only are summarized in Table 4-6. The new data may be judged by inspection to fit with the reference database. It is also noted that the slope of the reference database regression curve is only about one standard deviation different from the slope of the line using the new data only. Hence, the new data could be statistically demonstrated at a high level of confidence to be from the same population as the data in the reference database.

Sample new POL data and the regression curve obtained by including the data in the database are shown on Figure 4-2. The results of the regression analysis performed with the new data included in the database are summarized in Table 4-7. By inspection, the data are similar to that in the reference database. The regression curve obtained by including the new data in the analysis is not significantly different from that obtained using the reference data only.

Sample new leak rate data and the regression curve obtained by including the data in the database are shown on Figure 4-3. The results of the regression analysis performed with the new data included in the database are summarized in Table 4-8. The new data would be judged by inspection to belong to the same population as the data from the reference database. The inclusion of the data in the regression analysis has an insignificant effect on the regression line.

It should be noted that for the examples analyzed herein, demonstrating that the new data should be included in the reference database was straightforward. This may not always be the case with real data. Thus, additional analyses may have to be performed for plant specific reports if the reference database is to be expanded. If the new data do not appear to belong to the same database, an investigation and identification of cause must be performed. There are several possibilities that could result from such a finding, including significant changes to the results of the correlation analyses.

4.8 References

The references used in the preparation of this report section were:

4.1 TR-100407, Revision 2A (draft), "PWR Steam Generator Tube Repair Limits - Technical Support Document for Outside Diameter Stress Corrosion Crack at Tube Support Plates," Electric Power Research Institute, January 1995.

4.2 NP-7480-L, Volume 2, "Steam Generator Tubing Outside Diameter Stress Corrosion Cracking at Tube Support Plates - Database for Alternate Repair Limits, Volume 2: 3/4 Inch Diameter Tubing," Electric Power Research Institute, October, 1993.

4.3 NUREG-1477 (draft), "Voltage-Based Interim Plugging Criteria for Steam Generator Tubes - Task Group Report," United States Nuclear Regulatory Commission (NRC), June 1, 1993.

4.4 [United States Nuclear Regulatory Commission] Meeting with EPRI, NUMARC, "Resolution of Public Comments on Draft NUREG-1477," United States Nuclear Regulatory Commission, February 8, 1994.

4.5 EPRI Letter, "Exclusion of Data from Alternate Repair Criteria (ARC) Databases Associated with 7/8 inch Tubing Exhibiting ODSCC," D. A. Steininger (EPRI) to J. Strosnider (USNRC), April 22, 1994 [to become Appendix E of Reference 4.2].

4.6 Draft Generic Letter 94-XX, "Voltage-Based Repair Criteria for the Repair of Westinghouse Steam Generator Tubes Affected by Outside Diameter Stress Corrosion Cracking," United States Nuclear Regulatory Commission, *Federal Register*, Vol. 59, No. 155, August 12, 1994, pp. 41520-41529.

4.7 USNRC/Industry Meeting, "Resolution of Public Comments, NRC Draft Generic Letter 94-XX," January 18, 1995.

4.8 NRC Generic Letter 95-05, "Voltage Based Repair Criteria for Westinghouse Steam Generator Tubes Affected by Outside Diameter Stress Corrosion Cracking," United States Nuclear Regulatory Commission, August 3, 1995.

4.9 Regulatory Guide 1.121 (draft), "Bases for Plugging Degraded PWR Steam Generator Tubes," United States Nuclear Regulatory Commission, issued for comment in August, 1976.

4.10 Docket STM-50-456, "Safety Evaluation by the Office of Nuclear Reactor Regulation Related to Amendment No.   to Facility Operating License No. NPF-72 Commonwealth Edison Company Braidwood Station, Unit No. 1," United States Nuclear Regulatory Commission, May, 1994.

4.11 WCAP-14046 (Proprietary), Revision 1, "Braidwood Unit 1 Technical Support for Cycle 5 Steam Generator Interim Plugging Criteria," Westinghouse Electric Corporation, August 1994.

Table 4-1: Tube Materal Properties for APC Applications (Westinghouse)		
Property	**Value at RT**	**Value at 650°F**
Alloy 600 Mill Annealed 3/4" x 0.043" SG Tubes		
Sample Size	635	627
Yield Strength Mean	53.05	45.78
Yield Strength St. Dev.	4.8602	3.9081
Tensile Strength Mean	101.29	97.35
Tensile Strength St. Dev.	4.2173	3.9676
Flow Stress Mean	77.17	71.57
Flow Stress St. Dev.	4.1422	3.5668
95%/95% LTL Flow	69.925	65.325
Alloy 600 Mill Annealed 7/8" x 0.050" SG Tubes		
Sample Size	361	360
Yield Strength Mean	50.98	41.89
Yield Strength St. Dev.	4.2068	3.5856
Tensile Strength Mean	99.96	95.67
Tensile Strength St. Dev.	3.6123	3.4196
Flow Stress Mean	75.47	68.78
Flow Stress St. Dev.	3.5002	3.1725
95%/95% LTL Flow	69.225	63.115

Table 4-2: Regression Analysis Results - Burst Pressure vs. log(Bobbin Amplitude) Alloy 600 MA SG Tubes (Reference σ_f = 75 ksi)

CAUTION: Random database used, for illustration only.

Parameter	Value	Value	Parameter
b_1	-3.126	7.832	b_0
SE b_1	0.168	0.129	SE b_0
r^2	81.2%	0.946	SE P_B
F	346.5	80	DoF
SS_{reg}	310.04	71.58	SS_{res}
Pr(F)	8.3E-31	31.72	$SS_{log(V)}$
p_1-value	8.3E-31	8.5E-69	p_0-value

Table 4-3: Results of POL Regression Analysis @ 620°F and ΔP = 2560 psi **CAUTION: Random database, for illustration only.**	
Parameter	Values
b_1	-4.947
b_2	8.337
V_{11}	1.369
V_{12}	-1.932
V_{22}	3.106
Deviance	40.37
Pearson σ_{error}	0.77

Table 4-4: Regression Analysis Results:
log(Leak Rate) vs iog(Volts)
for Alloy 600 SG Tubes
@ 620°F and ΔP = 2560 psi

CAUTION: Random database, for illustration only.

Parameter	Value	Value	Parameter
b_4	3.259	-2.000	b_3
SE b_4	0.421	0.410	SE b_3
r^2	59.4%	0.707	SE log(Q)
F	60.1	41	DoF
SS_{reg}	30.01	20.48	SS_{res}
Pr(F)	1.5E-09	2.825	$SS_{log(V)}$
p_1-value	1.5E-09	1.67E-05	p_0-value

Table 4-5: Regression Analysis Results - Burst Pressure vs. log(Bobbin Amplitude) NEW DATA ADDED Alloy 600 MA SG Tubes (Reference σ_f = 75 ksi) **CAUTION: Random database used, for illustration only.**			
Parameter	Value	Value	Parameter
b_1	-3.145	7.832	b_0
SE b_1	0.153	0.116	SE b_0
r^2	83.0%	0.915	SE P_B
F	423.3	87	DoF
SS_{reg}	354.51	72.86	SS_{res}
Pr(F)	3.5E-35	35.83	$SS_{log(V)}$
p_1-value	3.5E-35	6.2E-77	p_0-value

Table 4-6: Regression Analysis Results - Burst Pressure vs. log(Bobbin Amplitude)
NEW DATA ONLY
Alloy 600 MA SG Tubes
(Reference σ_f = 75 ksi)

CAUTION: Random database used, for illustration only.

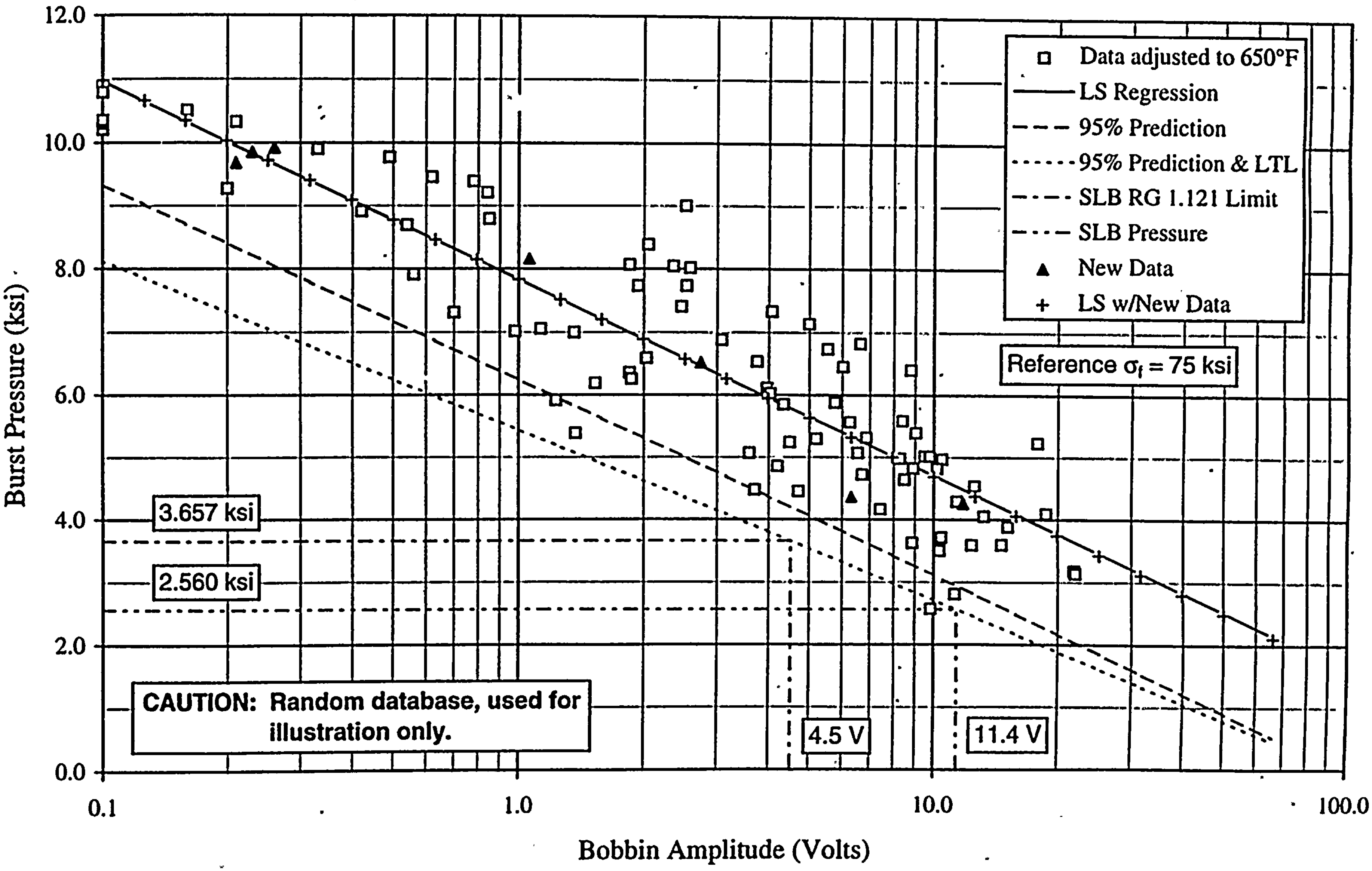
Parameter	Value	Value	Parameter
b_1	-3.420	7.749	b_0
SE b_1	0.246	0.166	SE b_0
r^2	97.5%	0.437	SE P_B
F	194.03	5	DoF
SS_{reg}	36.98	0.953	SS_{res}
Pr(F)	3.4E-07	3.16	$SS_{log(V)}$
p_1-value	3.4E-07	8.4E-08	p_0-value

Table 4-7: Results of POL Regression Analysis with New Data Added @ 620°F and ΔP = 2560 psi **CAUTION: Random database, for illustration only.**	
Parameter	Values
b_1	-5.091
b_2	8.582
V_{11}	1.399
V_{12}	-1.987
V_{22}	3.197
Deviance	41.14
Pearson σ_{error}	0.80

Table 4-8: Regression Analysis Results: log(Leak Rate) vs log(Volts) with New Data Added for Alloy 600 SG Tubes @ 620°F and ΔP = 2560 psi **CAUTION: Random database, for illustration only.**			
Parameter	Value	Value	Parameter
b_4	3.244	-1.981	b_3
SE b_4	0.408	0.397	SE b_3
r^2	59.6%	0.691	SE log(Q)
F	63.2	43	DoF
SS_{reg}	30.19	20.53	SS_{res}
Pr(F)	5.5E-10	2.868	$SS_{log(V)}$
p_1-value	5.5E-10	1.03E-06	p_0-value

Figure 4-1: Burst Pressure vs Bobbin Amplitude

3/4" x 0.043" Alloy 600 MA SG Tubes (Random Database @ 650°F)

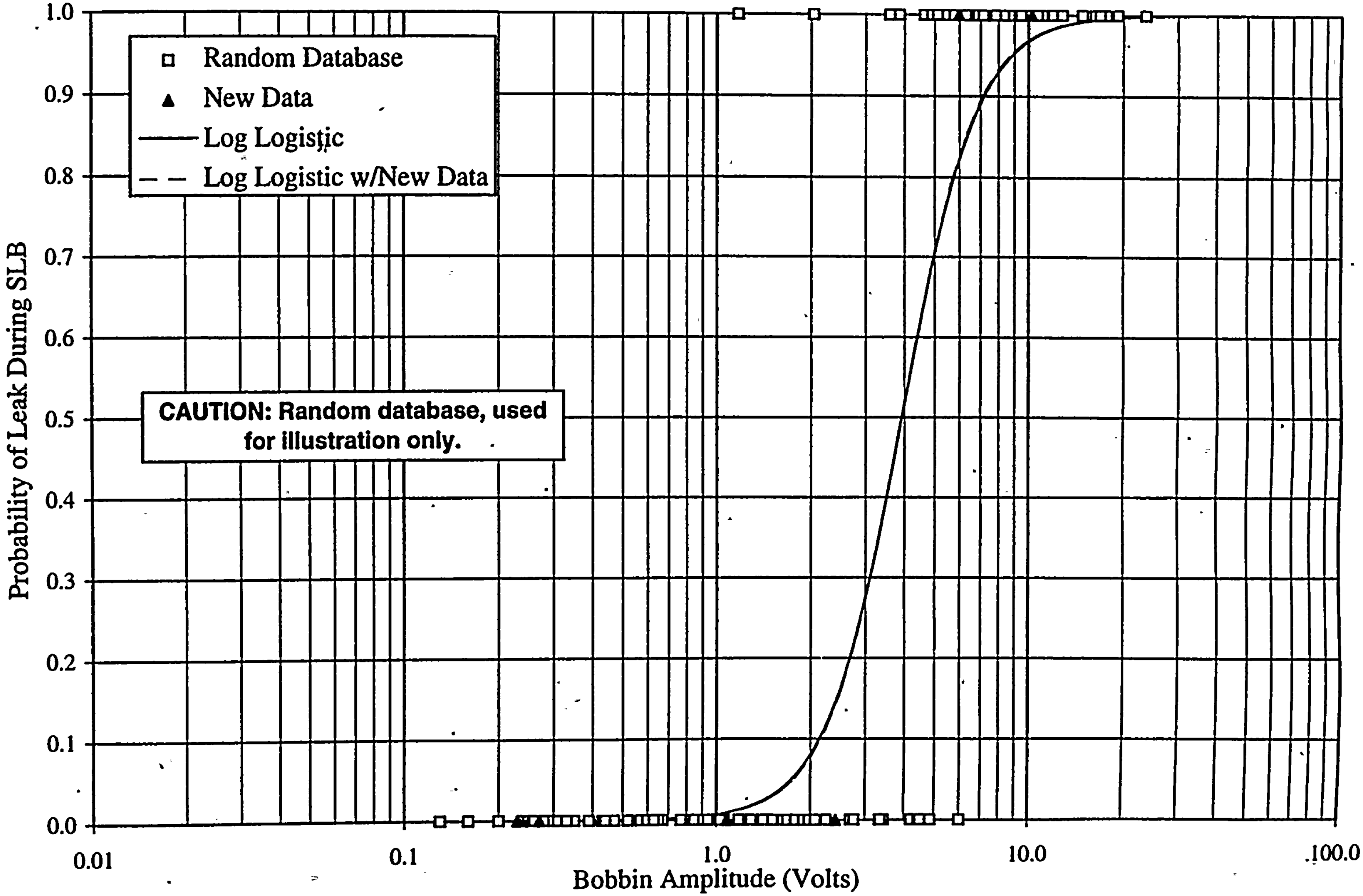


[34bw_gr1.xls] 34PV Gen

RFK: 12/8/96, 9:28 AM

Figure 4-2: Probability of Leak for 3/4" SG Tubes @ 650°F, ΔP = 2560 psi

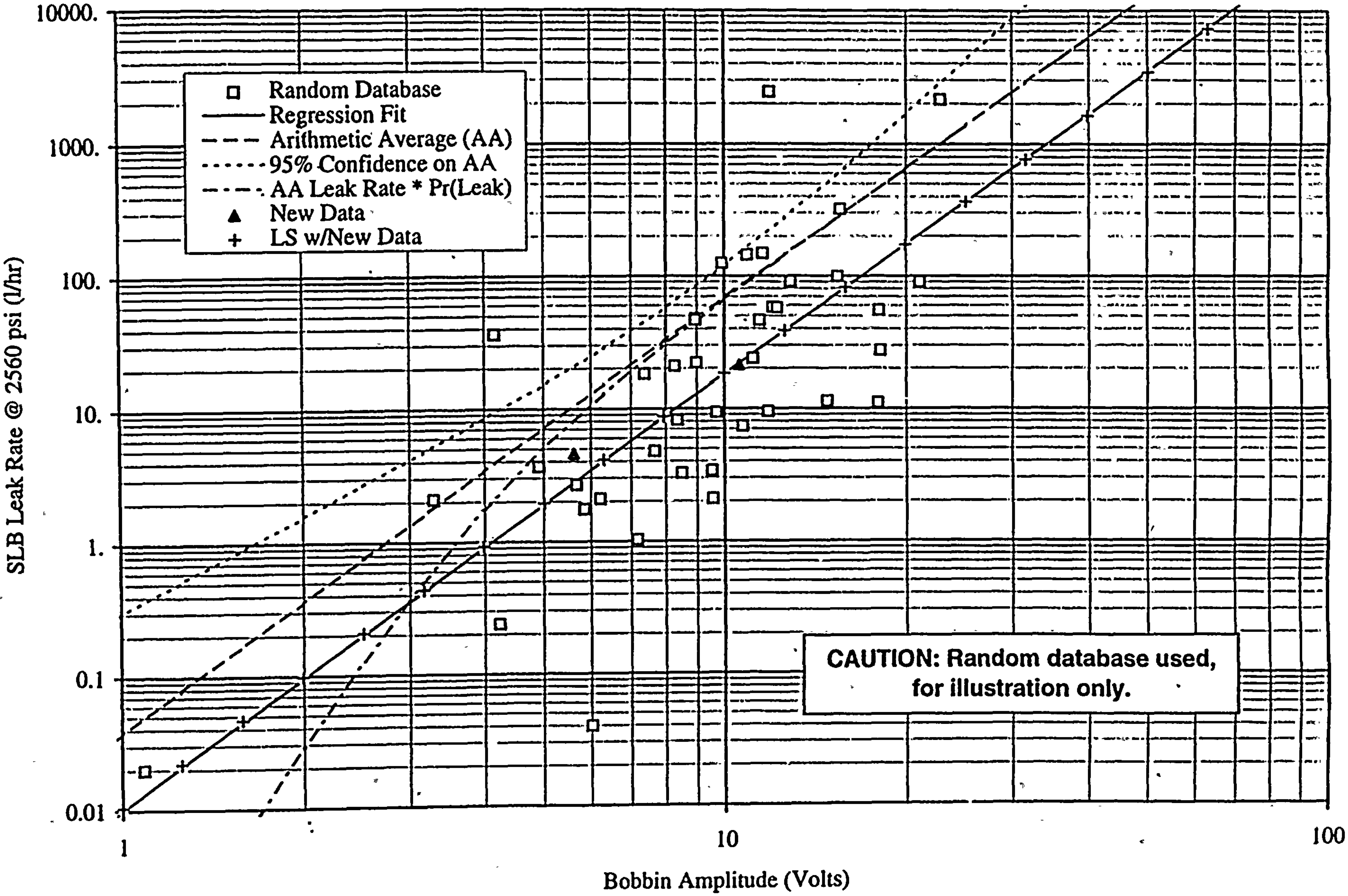
Comparison of New Data with Random Database



[34pv_gr1.xls] PoL.Gen

RFK: 12/8/96, 9:32 AM

Figure 4-3: SLB Leak Rate vs. Bobbin Amplitude, Random Database
3/4" x 0.043" Alloy 600 SG Tubes @ 650°F, ΔP = 2560 psi



(34qv qr1 xls) QvsV Gen

RFK: 12/8/96, 9:33 AM

5.0 SLB Leak Rate & Tube Burst Probability Analysis Methodology

5.1 General Methods Considerations

The purpose of this section is to provide information on the use of the correlations described in Section 4.0 of this report in support of the application of APC to specific indications in tubes in SGs at nuclear power plants. Information is presented on the use of deterministic models used for sensitivity analyses, and on the use of statistical simulation methods, i.e., Monte Carlo analyses, to estimate the total leak rate of all indications in the SG and the probability of burst of one or more of the indications in the SG.

The NRC generic letter requires calculations of SLB SG total leak rates and tube burst probabilities from the projected next EOC voltage distribution. It also permits licensees to calculate the tube leakage and burst probabilities from the measured EOC voltage distribution if it is not practical to complete the calculations using the projected EOC voltage distribution prior to returning the steam generators to service. The methods of this section may be applied to either voltage distribution.

5.2 Deterministic Methods for Sensitivity Analyses

The leak rate versus voltage correlation can be simulated in conjunction with the EOC voltage distributions obtained by Monte Carlo methods, or by applying the POL and leak rate correlations to the EOC voltage distribution obtained by Monte Carlo methods as applied for the draft NUREG methodology. This second approach is a hybrid that joins Monte Carlo and deterministic calculations. Parallel analyses verified that the full Monte Carlo leak rates and the direct application of the correlations to the EOC voltage distribution yield essentially the same results. Thus, it is adequate to apply the correlations to the EOC voltage distributions.

5.2.1 Deterministic Estimation of the Total Leak Rate

The determination of the end of cycle leak rate estimate proceeds as follows. The beginning of cycle voltages are estimated using the methodology described in Section 3.0 of this report. The distribution of indications is binned in 0.1V increments. The number of indications in each bin is divided by the POD. The resulting number of indications in each bin is reduced by the number of indications plugged in each bin. The final result is the beginning of cycle distribution used for the Monte Carlo simulations. The NDE uncertainty and growth rate distributions are then independently sampled to estimate an end of cycle distribution, also reported in bins of 0.1V increment. Given the EOC voltage distribution the calculational steps to obtain an estimate of the total leak rate are as follows:

(1) For each voltage bin, the leak rate versus bobbin amplitude correlation is used to estimate an expected, or average, leak rate for indications in that bin.

(2) The probability of leakage correlation is then used to estimate the mean probability of leak for the indications in each bin.

(3) The relationships derived in Appendix C of draft NUREG-1477 for the variance of the product of the probability of leak with the leak rate and for the total leak rate are then used to estimate the expected total leakage and variance for the sum of the indications in each bin as a function of the correlation means and estimated variances for the leak rate and probability of leak.

Recall from Section 4.0 that the expression used for the probability of leak of an indication, i, with a bobbin amplitude V_i is,

$$P(leak \mid V_i) = \frac{1}{1 + e^{-[\beta_1 + \beta_2 \log(V_i)]}}, \tag{5.1}$$

and the expression for the expected, or AA, leak rate, Q_i, from indication i as a function of volts, V_i, is given by,

$$Q_i = E\{Q_i \mid V_i\} = 10^{b_3 + b_4 \log(V_i) + \frac{\ln(10)}{2}\sigma_i^2}, \tag{5.2}$$

with a variance of the expected leak rate about the regression mean of,

$$\mathrm{Var}(Q_i) = Q_i^2\left[10^{\ln(10)\sigma_i^2} - 1\right].. \tag{5.3}$$

To account for the variances of the coefficients of the regression equation for the leak rate, the σ_i used in equations (5.2) and (5.3) is that from the predictive distribution for the logarithm of the leak rate as a function of bobbin amplitude, i.e., for each voltage, V_i, an effective standard deviation of the regression error, σ_i, is calculated as

$$\sigma_i = \sigma_e \sqrt{1 + \frac{1}{N} + \frac{\left[\log(V_i) - \overline{\log(V)}\right]^2}{\sum_j \left[\log(V_j) - \overline{\log(V)}\right]^2}}, \tag{5.4}$$

where N is the number of data pairs in the regression analysis, and σ_e is an unbiased estimate of σ for the population.

The expected total leak rate from all of the indications in all of the bins is

$$T = \sum_{i=1}^{N_B} \frac{n_i}{1 + e^{-[b_1 + b_2 \log(V_i)]}} 10^{b_3 + b_4 \log(V_i) + \frac{\ln(10)}{2}\sigma_i^2}, \qquad (5.5)$$

where N_B is the number of bins, and n_i is the number of indications in the bin, which is not necessarily an integer number, with bobbin amplitude V_i. Thus, the expected total leakage for the entire distribution is obtained as the sum of the expected leak rates for each bin.

In order to estimate an upper confidence bound for the total leak rate an expression is needed for the variance of the total leak rate. There are two sources of variance to be considered, the variance *about* the predicted expected value and the variance *of* the predicted expected value; the estimated total variance about the predicted expected value being the sum of the two. Moreover, the variance of the probability of leak must be considered in the variance about the predicted value. The variance of the total leak rate about the predicted expected value, including consideration of the variance of the probability of leak, is

$$\mathrm{V}(T) = \sum_{i=1}^{N_B} n_i \left\{ P_i Q_i^2 \left[e^{\ln(10)\sigma_i^2} - 1 \right] + Q_i^2 P_i (1 - P_i) \right\}, \qquad (5.6)$$

where P_i is the probability of leak from equation (5.1). Equation (5.6) is based on an application of the standard expression for the biased estimate of the variance of a product. As noted, an additional variance term is added in order to estimate the contribution to the variance from the correlation between the individual leak rates, i.e., from the covariance, which arises as a consequence of using the regression equations. Thus, the second term accounts for the variances of the positions of the regression equations. A linearized approximation (via Taylor's Theorem) of the variance of the mean of the regression prediction, T_μ, is given by

$$\mathrm{V}(T_\mu) = \sum_{i=1}^{N_B} n_i \left\{ \frac{dT}{d\beta_j} \right\}_i^T \begin{bmatrix} [Cov(\beta_1,\beta_2)] & 0 & 0 \\ 0 & [Cov(\beta_3,\beta_4)] & 0 \\ 0 & 0 & \mathrm{V}(\sigma_i^2) \end{bmatrix} \left\{ \frac{dT}{d\beta_j} \right\}_i, \qquad (5.7)$$

where the derivative of the total leak rate vector contains five elements for j=1,...,5, and the *Covariance Matrix* is a square 5x5 matrix consisting of the estimated variances and covariances of the estimated individual regression coefficients and σ_i. Note that here $[Cov(\beta_1, \beta_2)]$ and $[Cov(\beta_3, \beta_4)]$ are each 2x2 matrices, where the β's are estimated by b_1 through b_4, and recall that σ_i is an estimate of β_5. The variance of the variance is estimated as

$$V(\sigma_i^2) = \frac{2\sigma_i^4}{n-2}, \tag{5.8}$$

where n is the number of data pairs used in the leak rate regression analysis. The standard deviation of the total leak rate is then taken as the square root of the variance of the total leak rate. The upper bound 95% confidence limit on the total leak rate is then obtained as the expected total leak rate plus 1.645 times the standard deviation of the total leak rate. The results obtained with this approach have been compared to results obtained from the Monte Carlo simulations without significant differences being observed for total leak rates at a 95% confidence level when leak rates in excess of 1 GPM are predicted. At higher confidence levels, e.g., 99%, the differences could be significant. Because of this uncertainty the deterministic analyses are limited to sensitivity studies.

5.2.2 Deterministic Estimation of the Probability of Burst as a Function of Volts

Using BOC or EOC distributions and the regression results for the correlation of burst pressure to bobbin amplitude, an estimate of the probability burst of one or more tubes in the SG can be estimated. The regression curve was given in Section 4.0 as,

$$P_{B_i} = a_1 + a_2 \log(V_i) \tag{5.9}$$

where the burst pressure is measured in *ksi* and the bobbin amplitude is in *volts*. Here, a_1 and a_2 are estimates of unknown parameters α_1 and α_2 of the relation. The value obtained from equation (5.9) applies to tubes with a flow stress equal to the reference flow stress, S_{ref}, used in estimating the equation parameters. A normalized value of the burst pressure can then be found as,

$$P_{N_i} = \frac{P_{B_i}}{S_{ref}}. \tag{5.10}$$

The burst pressure for any single indication is then given by,

$$P_{A_i} = P_{N_i} S_f. \tag{5.11}$$

and the variance of the burst pressure accounting for the variance of the residuals about the regression curve and the variation in S_f can be calculated as

$$V(P_B) = \left(\frac{2t}{R_m}\right)\left[P_{bar}^2 V(S_f) + \overline{S}_f^2 V(P_{bar}) - V(P_{bar})V(S_f)\right], \tag{5.12}$$

where V is used to represent an unbiased estimate of the variance of the respective variable in parentheses. The standard deviation of the burst pressure is then taken as the square root of the variance of the burst pressure. For any voltage level, the number of standard deviations difference between the predicted burst pressure and the SLB differential pressure can be calculated. The probability of burst, Pr_i, is then obtained from a Student's t distribution. The probability of burst as a function of indication volts for 3/4" diameter tubes is illustrated on Figure 5-1. For n indications in a voltage bin, the probability that none of the indications in the bin burst is then,

$$Pr = (1 - Pr_i)^n. \tag{5.13}$$

For all of the indications in all of the bins, the probability of burst of one or more indications is then,

$$Pr = 1 - \sum_{k=1}^{N} (1 - Pr_k)^{n_k}, \tag{5.14}$$

where n_k is the number of indications in bin k, not necessarily an integer number, and Pr_k is the probability of burst of a single indication in bin k. As for the deterministic estimates of the leak rate, the deterministic estimates of the probability of burst are limited to sensitivity studies.

5.3 Deterministic Analysis Results

An example of the results from a deterministic analysis of a sample data set is presented in Section 6.0 of this report.

5.4 Monte Carlo Analysis Methodology

The estimated, total end of cycle leak rate can also be calculated using Monte Carlo techniques which simulate the variation in the parameters, i.e., coefficients, and the variation of the dependent variable about the regression line. A 95% confidence bound on the total leak rate from the SG is calculated using a Monte Carlo simulation. The results from the deterministic analyses are used as an order of magni-

tude verification of the Monte Carlo results. The method described herein simulate the variation of each parameter of the correlation equations and the effect of the covariance of the individual indication leak rates.

In order to simplify the discussion of the Monte Carlo techniques, different nomenclature is used from that of the previous section, i.e., Q_i is used to represent the common logarithm of the leak rate, and V_i is used to represent the common logarithm of the bobbin amplitude. Thus, the following model is used to describe a working relationship between the logarithm of the leak rate and the logarithm of the bobbin amplitude,

$$Q_i = b_3 + b_4 V_i + \varepsilon , \tag{5.15}$$

where ε is the estimated error of the residuals, assumed to be from a population that has a zero mean, and a variance that is not dependent on the magnitude of V_i. The coefficients, b_3 and b_4 are the estimates from the regression analysis of some true coefficients, β_3 and β_4, representing the intercept and slope of the equation, respectively.

The method used by Westinghouse for simulating the total leak rate is the outcome of a series of technical discussions held with the NRC. The method differs from that reported in some prior WCAP reports, wherein the predictive distribution was simulated and covariance terms were ignored. It is noted that, although both methods yielded similar results (within ~3%) for one domestic plant analyzed, the method described herein is more statistically accurate. This small difference in the total leak rate results is because the contribution of the covariance terms relative to the variance terms is relatively small for the correlations used herein. In summary, random versions of the POL and leak rate correlations are generated and used to calculate the sum of the leak rates for all of the indications in a SG to obtain a single simulated value of the total leak rate. This process is repeated to obtain a distribution of the total leak rate from at least 10,000 simulations of the correlation equations. A non-parametric 95% confidence bound on the total leak rate is then estimated from the distribution of total leak rates.

At the start of each SG simulation, i.e., the calculation of a single total leak rate, a random value for the *standard deviation of the errors for the population* is calculated from the χ^2 distribution, the degrees of freedom from the data, and the standard deviation of the regression errors. This is used to calculate random values for the parameters of the regression equation, which remain constant for the entire SG simulation. The variation of the regression predictions are accounted for by randomly estimating the POL from a uniform distribution, and by adding the product of a random normal deviate and the standard deviation of the errors for the population to the predicted logarithm of the leak rate, for each individual indication in the SG distribution. The total leak rate for the SG simulation is calculated as

the sum of the leak rates from all of the indications in the SG. The expression for the total leak rate is

$$T = \sum_{i=1}^{N} \zeta_i R_i(\beta_1, \beta_2) Q_i(\beta_3, \beta_4, \beta_5), \tag{5.16}$$

where N = the total number of whole or partial indications in the SG at EOC,

ζ_i = the proportion of the indication, e.g., 1 for a whole indication,

$R_i(\beta_1, \beta_2)$ = 0 or 1 is the POL for a single indication, i, in a tube,

$Q_i(\beta_3, \beta_4, \beta_5)$ = is the conditional leak rate of indication i, i.e., the leak rate if the indication is leaking,

β_1, β_2 = the coefficients of the POL equation,

β_3, β_4 = the coefficients of the leak rate versus bobbin amplitude equation, and

β_5 = the standard error of the log of the leak rate about the correlation line, also referred to herein as σ.

To simulate the total leak rate from all of the indications in the generator, random coefficients for the probability of leak, POL, and leak rate correlation equations are generated, and then those coefficients are used to simulate the POL and leak rate for each indication. The POL, R_i, for each indication, i, is simulated as,

$$R_i(\beta) = \begin{cases} 1 & \text{if } U_i < logit(\beta_1 + \beta_2 \log(V_i)) \\ 0 & \text{otherwise} \end{cases}, \tag{5.17}$$

where U_i is an independent draw from a uniform distribution. The step of determining an integer value for the POL accounts for the variation of the distribution of probabilities about the log-logistic regression line. Discussion of the generation of β_1 and β_2 is left until after the discussion of the coefficients for the leak rate equation.

5.4.1 Leak Rate versus Bobbin Amplitude Simulation

To simulate the leak rate from the regression line, random coefficients β_3 and β_4 must be simulated. Each of these has a variance that is dependent on the variance of the error of the log of the leak rate about the regression line. Thus, the first step is to simulate a random error variance by picking a random χ^2 deviate for n-2 degrees of freedom and then calculating a random error variance, σ^2, for the correlation equation from the regression error variance as

$$\sigma^2 = \frac{(n-2)}{\chi^2_{(n-2),random}} \hat{\sigma}^2 = f_V \hat{\sigma}^2, \tag{5.18}$$

where n is the number of data pairs used to calculate the regression coefficients, and f_V is defined by equation (5.18). This is now one possible variance for the *population* of log-leak rates about a correlation equation. Thus, it is appropriate to use the normal distribution to obtain random values for the parameters of the correlation equation. The distribution of β_3 and β_4 will be bivariate normal. Since they are correlated, although each is normally distributed marginally, they are not free to vary independently. If a value for the slope is determined first, then the distribution of the intercept values will be conditional on that value of the slope. The degree of correlation is indicated by the off-diagonal entry in the parameter covariance matrix calculated from the regression analysis. The entries of the covariance matrix of the parameters, V_{11}, V_{12}, and V_{22}, for the correlation equation to be used for a SG simulation are obtained from the corresponding estimated matrix obtained from the regression analysis as

$$V_{ij} = f_V \hat{V}_{ij}, \tag{5.19}$$

where the caret, "^", is used to indicate an estimate from the regression data. A bivariate normal intercept for the simulation correlation is then calculated from the regression equation intercept as

$$\beta_3 = b_3 + Z_1 \sqrt{V_{11}}, \tag{5.20}$$

and the bivariate normal slope is calculated from the regression slope as

$$\beta_4 = b_4 + Z_1 \frac{V_{12}}{\sqrt{V_{11}}} + Z_2 \sqrt{V_{22} - \frac{V_{12}^2}{V_{11}}}, \tag{5.21}$$

where Z_1 and Z_2 are random univariate normal deviates, i.e., from a population with a mean of zero and a variance of one. We now have β_3, β_4, and σ for use in simulating all of the leak rates from each of the indications in the SG for one

simulation of the total leak rate. For each simulation of an individual indication, i, the leak rate from an indication with a proportion, ζ_i, of unity will be,

$$Q_i(\beta) = 10^{\beta_3 + \beta_4 \log(V_i) + \beta_5 Z_i}, \tag{5.22}$$

with Z_i representing the i^{th} value from N independent draws from a standard normal distribution. Once the probabilities of leak have been calculated, the total leak rate for one simulation is then calculated using equation (5.16). It is noted that each simulation of T requires the generation of one β vector, N binomial variates R_i, and a maximum of N log-normal variates Q_i. In practice, a value for the leak rate only needs to be generated for each indication that is leaking, i.e., when $R_i = 1$.

Simulation When the p-Value is Greater than 5%

If the p value from the regression analysis is greater than 5%, it is assumed that the leak rate is independent of the bobbin amplitude and the leak rate model, as discussed in Section 4.0, is,

$$S_i = b_3 + \varepsilon, \tag{5.23}$$

where S_i represents the common logarithm of the leak rate, and b_3 is the mean of the leak rate data. The simulation proceeds similar to that when the leak rate correlation is used. For each simulation of total leak rate from a SG, a random estimate of the population standard error, β_5, is obtained using equation (5.18) with the numerator of the fraction and the degrees of freedom for the random selection of χ^2 being (n-1). A random value of β_3 is then calculated from b_3 as,

$$\beta_3 = b_3 + Z \frac{\beta_5}{\sqrt{n}}, \tag{5.24}$$

where Z is a random normal deviate, ~ N(0,1). An individual simulation of the leak rate from indication i with a proportion of unity, Q_i, is then given by,

$$Q_i = 10^{\beta_3 + Z\beta_5}. \tag{5.25}$$

The total leak rate from all of the indications in the SG is then calculated as per equation (5.16), except that the leak rate is a function of β_3 and β_5 only.

5.4.2 Probability of Leak Simulation

The generation of the coefficients of the POL relation to be used in the simulation of the total leak rate proceeds in the same manner as for the coefficients of the leak rate relation. The elements of the covariance matrix are obtained from the GLM regression analysis and used with the estimated coefficients in equations like (5.20) and (5.21) to obtain β_1 and β_2 for a random population POL equation. However, for the simulation of the POL, there is no term of the form σZ_i, e.g., like $\beta_5 Z_i$ in equation (5.22), in the simulation of the total leak rate. This exception is due to the fact that the data are binary. In effect, this additional term is being simulated through the use of the random sampling to determine if R_i is 0 or 1 in equation (5.17).

It is noted that the elements of the covariance matrix obtained from the GLM regression are scaled to a mean square error (mse) of 1. This is because the mse for the binary variables is asymptotically 1. A check of this assumption may be made by calculating an estimate of the square root of the mse, also referred to as the Pearson standard deviation, from the regression results as

$$\hat{\sigma} = \sqrt{\frac{1}{n-2}\sum_i \frac{(y_i - \mu_i)^2}{\mu_i(1-\mu_i)}}\,, \tag{5.26}$$

where the y_i's are the observed probabilities of leak, either zero or one, from the leak and burst testing, and the μ_i's are the calculated probabilities of leak from the *logistic* regression equation. A significant departure from 1 for this quantity could be indicative of an inadequate model.

5.4.3 SLB Tube Burst Pressure Simulation

The simulation of the burst pressure is performed in a manner similar to that for simulating the leak rate. For the burst pressure, however, an additional simulation must be made of the flow strength of the material of the tube containing the indication. This is because the correlation of the burst pressure to the bobbin amplitude was performed for test burst pressures which were adjusted to correspond to a reference flow stress.

For each simulation of the SG, a random estimate of the standard deviation of the residuals of the regression errors about the regression line is generated. This is followed by generating random estimates of the parameters of the regression equation. The burst pressure, P_{ref}, for each individual indication in the SG is then calculated as,

$$P_{ref} = \alpha_0 + \alpha_1 \log(V_i) + Z_i \alpha_3 , \tag{5.27}$$

where α_3 is the estimated standard error of the population of the residuals, α_0 and α_1 being the bivariate normal estimates of the parameters of the correlation, and Z_i being a random standardized normal variate. The result thus obtained is valid for the reference flow stress, S_{ref}, of the adjusted data used to calculate the estimates of the parameters of the burst pressure correlation. A random estimate of the flow stress of the tube material is then made as,

$$S_i = S_m + t\sigma_S , \tag{5.28}$$

where S_m and σ_S are the mean and standard deviation rspectively of the flow stress from a database of the materials of fabrication of the tubes in the SG or from an expanded database for tubes from a population of SGs. The final estimate of the burst pressure is then calculated as,

$$P_i = P_{ref} \frac{S_i}{S_{ref}} . \tag{5.29}$$

The value of the burst pressure is then compared to the SLB differential pressure to determine if the tube would be likely to burst during a postulated SLB event.

5.5 SLB Leak Rate Analysis Methodology

Once the simulations have been performed it is necessary to estimate the total leak rate from the SG during a postulated SLB event. The ouput from the simulations is a distribution of total leak rates that might be expected. The current accepted methodology is to estimate a 95% confidence bound on the total leak rate and compare that value to accepted limits. The total leak rate values from the Monte Carlo simulation are ordered from the lowest to the highest. A one-sided distribution-free 95% confidence bound for the 95^{th} percentile of the population of total leak rates is then calculated. Thus, there is a 95% confidence that at least 95% of the potential population of total leak rates will be less than the estimated value.

A confidence interval for value, ξ_p, bounding the $100 \cdot p^{th}$ percentile of the distribution from a drawn sample of size n is constructed by means of the binomial distribution. A one-sided conservative upper $100 \cdot (1-\alpha)\%$ confidence bound for the

$100 \cdot p^{th}$ percentile of the sampled population is obtained as $\xi_p = x_u$, where u is chosen as the smallest integer such that

$$P(x_i > \xi_p) = \sum_{j=0}^{u-1} \binom{n}{j} p^j (1-p)^{n-j} \geq 1 - \alpha . \tag{5.30}$$

Since the binomial distribution is computationally difficult for the extremely large number of simulations performed, an equivalent approach using the F-distribution is to find the smallest value of n such that,

$$\frac{1}{1 + \frac{N-n+1}{n} F_{1-\alpha, 2(N-n+1), 2n}} \geq P \tag{5.31}$$

where N is the total number of simulations performed.

Note that for equation (5.31), the index is found such that the desired confidence of $100 \cdot (1-\alpha)\%$ is maintained constant while the actual percentile, P, will be greater than or equal to that desired. The result obtained using equation (5.30) maintains the percentile, P, constant so the actual confidence level, $100 \cdot (1-\alpha)\%$, will be greater than or equal to that desired. For example, if the number of SG simulations performed was 1,000, a one-sided upper 95% confidence bound on the 95th percentile of the total leak rates of the population of possible leak rates would be given by the 962nd ordered (from smallest to largest) total leak rate. This value is a 95.7% confidence value for the 95.0 percentile by equation (5.30), or a 95% confidence value for the 95.1 percentile by equation (5.31).

5.6 SLB Tube Burst Probability Analysis Methodology

During the simulation of the burst pressures, the number of tubes, based on the indication proportion, with burst pressures less than the SLB differential pressure is counted. The value for each SG simulation is retained. Thus, at the end of the simulation the number of SGs experiencing zero bursts is known, the number of SGs experiencing one burst is known, et cetera. The method of accounting for fractional tubes is to add the fractional part to the next integer number of tubes.. Thus, if 1.3 bursts result from the simulation, it is reported as 1 occurrence of 1 burst tube, and 0.3 occurrences of two burst tubes. Using these results, confidence bounds are determined for the likelihood of one, two, or more bursts to occur, i.e., a

one-sided 100·(1-α)% upper confidence bound for the Monte Carlo results is found from the following equation:

$$Pr_U = \frac{1}{\dfrac{N-n}{(n+1)F_{1-\alpha,2(n+1),2(N-n)}} + 1} \tag{5.32}$$

where N is the total number of Monte Carlo trials, n is the number of observed occurrences of $P_B \leq P_{SLB}$, i.e., predicted bursts, and F is from the F-distribution for the specified number of degrees of freedom for the numerator and denominator respectively. For zero occurrences in the Monte Carlo simulation, equation (5.32) can still be used to find an upper confidence bound on the probability. The value of the upper confidence bound relative to the observed fraction of occurrences is larger when fewer occurrences are predicted. If n/N is significant, e.g., on the order of 10^{-2}, and n is large, the upper bound might be only a few percent higher than the mean result. However, if n/N is not significant, say 10^{-6}, and n is very small, the upper bound could be an order of magnitude greater than the mean estimate. Since the probability, i.e., relative frequency, of multiple ruptures is expected to be very low, the upper confidence bound on the probability will be relatively higher than that for a single burst.

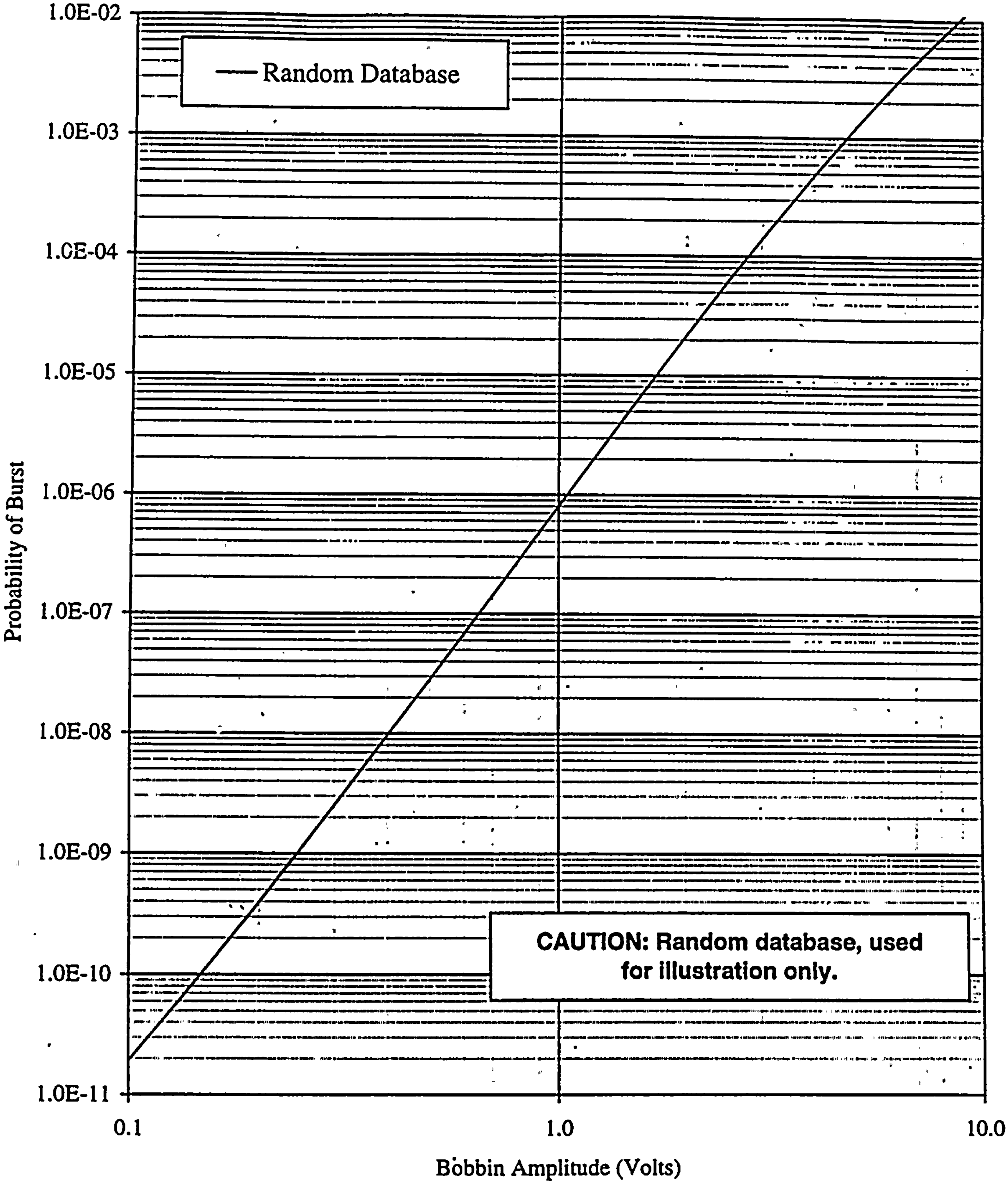
For information, since it would rarely be of interest relative to the probability of burst, a 100·(1-α)% one-sided lower confidence bound on the Monte Carlo results can be found as

$$Pr_L = \frac{1}{1 + \left(\dfrac{N-n+1}{n}\right)F_{1-\alpha,2(N-n+1),2n}}. \tag{5.33}$$

For zero occurrences in the Monte Carlo simulation, the limit of equation (5.33) is also zero.

Figure 5-1: Probability of Burst vs. Bobbin Amplitude

3/4" OD x 0.043" Thick, Alloy 600 MA, SG Tubes @ 650°F



6.0 Example Analysis Results

The purpose of this section is to provide illustrative example results from the analyses described in the previous sections. Since the database used for the examples is not to be kept up to date by continuously revising this document, the numerical results are not applicable to any one plant and should not be used to perform plant specific APC analyses.

6.1 Simulation Code Description

In order to estimate the probability of burst (PoB) and the total leak rate during a postulated SLB event, two Westinghouse proprietary computer codes, EOC_VQB and LEAKTOTL, were initially written. These were later combined into a single Westinghouse proprietary code named CYCLESIM to perform the same calculations. Input to the code consists of a beginning of cycle distribution of indications, the length of the cycle to be simulated, a distribution of indication growth rates, and the length of the cycle used to determine the growth rate distribution. The parameters of the burst versus volts correlations (for 3/4" and 7/8" diameter tubes), limiting material properties, the POL versus volts correlations, and the leak rate versus volts correlations are coded into the program. The regression standard error, the values of the variance-covariance matrices, and other values pertinent to the regressions are also contained within the code.

Several options for running the code are also provided, e.g., inputting the EOC distribution of indications, the plugging distribution, followed by the application of the POD as described in section 3.0. In practice, the program may be first run to develop a projected EOC distribution of indications. This is achieved by simulating the EOC distribution of the indications in the SG several thousand times. Once the EOC distribution is obtained, it is used as a basis for simulating the total leak rate and for estimating the probability of burst during a postulated SLB event. The simulation methodology is described in Section 5.0 of this report.

An alternative to specifically developing the EOC distribution followed by the simulation of the total leak rate and the distribution of the burst pressures is to base the entire simulation on the BOC distribution of indications. In this case the simulation proceeds as follows. The parameters of the POL, the leak rate, and the burst pressure correlations to the bobbin amplitude are randomly established using the methodology described in Section 5.0. Once the set of correlating equations is established, the entire distribution of indications in the SG is sampled to obtain for each indication, a random BOC voltage, a random NDE uncertainty, a random growth, a random POL, a random leak rate if the POL is 1, and a random burst pressure. The total leak rate for the entire SG is calculated as the sum of the leak rates of all of the indications in the SG. Again, the leak rate from a partial indication is weighted by the proportion of an indication being simulated. The number of indications exhibiting an EOC burst pressure less than the SLB differential pressure is counted. The resultant values from each SG simulation are retained.[1]

6.2 Example Correlations and Distributions

Regression analyses results were given in Section 4.0 for randomized data sets. For those examples, a reference set of data was randomly adjusted to develop new data sets. The results of the analyses were that the burst pressure can be predicted as a function of the bobbin amplitude by the relation,

$$P_{B_i} = 7.832 - 3.126\log(V_i). \tag{6.1}$$

Likewise, the POL of an indication as a function of the bobbin amplitude was estimated as,

$$P(leak\,|V_i) = \frac{1}{1 + e^{-[-4.947 + 8.337\log(V_i)]}}, \tag{6.2}$$

and the equation for predicting the leak rate as a function of bobbin amplitude was estimated as,

$$Q_i = 10^{-2.000 + 3.259\log(V_i) + \frac{\ln(10)}{2}\sigma_i^2}, \tag{6.3}$$

where σ_i is estimated from the predictive distribution as discussed in Section 5.0. The other parameters associated with the estimated correlating functions were provided in tables in Section 4.0.

A typical BOC distribution of indications is illustrated as Figure 6-1. A typical growth distribution for the indications is illustrated as Figure 6-2. Finally, a typical EOC distribution of indications is illustrated as Figure 6-3.

6.3 Deterministic Analysis Results

As discussed in Section 5.0 of this report, deterministic analyses may be performed for screening purposes. The estimation of total leak rate from the SG at a 95% confidence level may be expected to be reasonably accurate, e.g., within about 5% of

---

[1] The total number of simulations of the SG is a function of the particular result desired, e.g., for the total leak rate the number of simulations might be on the order of 50,000 to 100,000. The number of simulations to characterize the burst pressure distribution would likely be significantly larger.

the values obtained from Monte Carlo simulations, however, the estimates for higher confidence levels may diverge rapidly from those obtained by simulation. The deterministic values are based on only a first order approximation of the covariance that exists between the estimated leak rates for the individual indications. Relatively good agreement with the simulation results at a 95% confidence level has been observed for all analyses performed to date. There is, however, no rigorous theoretical justification for the efficacy of the results obtained, hence, the restriction on usage to scoping analyses only.

A BOC deterministic leak rate estimate is illustrated in Table 6-1, and an EOC estimate is illustrated in Table 6-2. For the case illustrated the estimated BOC leak rate during a postulated SLB is 2.6 gpm and the estimate EOC leak rate is 4.6 gpm. For this level of leak rate, the deterministic estimate at a 95% confidence level would be expected to match the Monte Carlo result within about 25 to 30%. For some specific distributions the agreement has been within 3%. If the EOC leak rate was predicted to be on the order of 0.5 gpm, the deterministic estimate could be in error on the order of 50% or more. For even lower values the error can increase to an order of 100%. However, the relative magnitude of the error becomes less important for the lower values since they would be expected to be significantly less than the allowable leak rate.

An EOC deterministic estimate of the probability of burst of one or more tubes is illustrated in Table 6-3. For the example shown the POB is greater than $1 \cdot 10^{-2}$, thus, the result would require reporting to the NRC if it was obtained from a Monte Carlo simulation. For low probabilities of burst, as in the example, the POB obtained from a simulation using the same data will generally result in an estimate about two orders of magnitude less. This is because the random estimate of the POB is the product of two assumed normal distributions. The resulting distribution is not normal, although it is treated as such for the deterministic estimate. An examination of the actual distribution as obtained from Monte Carlo simulations indicates it to be skewed right, i.e., the upper tail is longer than the lower tail. Hence, the probability of a burst pressure being lower than a specified value is less in the tail than would be estimated assuming the distribution to be symmetric.

6.4 SLB Leak Rate Analysis Results

The output from the Monte Carlo simulation code consists of a check of the input data, a deterministic estimate of the total leak rate (a check value), statistics of the simulated NDE uncertainties and voltage growth rates, statistics of the simulated leak rate distrigution and the 95% confidence estimate of the 95th percentile value of the leak rate. An example of the deterministic check of the analysis in provided as Table 6-4. The descriptive statistics of the simulated SG total leak rate distribution is provided as Table 6-5. A check value of the 95% confidence value of the total leak rate during a postulated SLB event is calculated based on assuming a log-normal distribution. For the example cited, the check value was 2.37 gpm. The

minimum and maximum total leak rates simulated are reported as shown in Table 6-6 along with the appropriate sorted index number and the 95% confidence total leak rate. These represent the extreme tails of the distribution and are not statistically significant. For the case shown, the leak rate from the simulation, 2.87 gpm, was about 30% less than the deterministic estimate. Finally, a representative listing of the sorted leak rates is printed, see Table 6-7[2], to gain an understanding of the magnitude of the leak rates in the upper tail of the distribution. If the values in the tail are found to vary significantly, it would be an indicator that the analysis should be repeated with the number of simulations increased. The example results presented do not indicate that a repeat of the analysis is necessary.

6.5 SLB Tube Burst Probability Analysis Results

The output from the simulation of the burst pressures of the tubes for the example case is illustrated in Table 6-8. The results are based on estimating the probabilities of one, two, three, etc., bursts in a single SG as the fraction of occurrences divided by the number of simulations. A 95% upper confidence bound on the respective probabilities of burst is also calculated and reported. Finally, the statistics of the simulated burst pressure distribution are calculated and reported as shown in Table 6-9. For the example distribution, the 95% upper confidence bound probability of a single tube burst was found to be $2.9 \cdot 10^{-2}$. The corresponding estimate from the deterministic analysis was $3.6 \cdot 10^{-2}$. The probability of two bursts in the same SG during a SLB was estimated to be $7.3 \cdot 10^{-4}$. Finally the probability of three bursts was $7.8 \cdot 10^{-5}$ at an upper bound 95% confidence. This was however for 0.3 tubes, hence the probability would be expected to be about 1/3 of this value for a whole indication.

[2] The values in this table are not from a simulation of a previously presented data set, e.g., they are not from a simulation of the distribution on indications as given in Table 6-2.

Table 6-4: Deterministic Estimate of the Total Leak Rate	
Parameter	Value
$\Sigma N_i E(Q_i) P_i$	$1.485 \cdot 10^2$ lph^2
$V[N_i V(P_i Q_i)]$	$2.002 \cdot 10^5$ lph^2
$Cov[N_i V(P_i Q_i)]$	$1.935 \cdot 10^4$ lph^2
V + Cov	$2.195 \cdot 10^5$ lph^2
Standard Deviation	$4.686 \cdot 10^2$ lph
Z - Deviate	1.645
Q_{total}	919.3 lph (4.05 gpm)
Covariance Contribution	3.8 %

Table 6-5: Descriptive Statistics of the Monte Carlo EOC Leak Rate Distribution	
Parameter	Value
Maximum Qsum	91.62 gpm
Minimum Qsum	$6.475 \cdot 10^{-3}$ gpm
Sum of Qsum	$4.112 \cdot 10^4$ gpm
Average Qsum	0.8223 gpm
Std Dev Qsum	2.716 gpm
Avg log(Qsum)	-0.4297
StD log(Qsum)	0.4885
Approx. Bound	2.366 gpm

Table 6-6: Monte Carlo Total Leak Rate 50,000 Simulations of the SG 95% Confidence	
Minimum Total Leak Rate	$6.50 \cdot 10^{-3}$ gpm
Maximum Total Leak Rate	91.4 gpm
Confidence Index Number	47580
Bounding Total Leak Rate	2.87 gpm

Table 6-7: Monte Carlo Estimates of the Total Leak Rate

Bin Index	Sorted Index	Number in Bin	Leak Rate (GPM)
649	47566	48	2.871
650	47614	25	2.900
651	47639	49	2.929
652	47688	25	2.959
653	47713	73	2.988
654	47786	24	3.019
655	47810	24	3.049
656	47834	74	3.079
657	47908	24	3.110
659	47932	74	3.173
660	48006	24	3.205
661	48030	25	3.237
663	48055	1	3.303
665	48056	72	3.370
666	48128	24	3.403
669	48152	48	3.507

Table 6-8: Monte Carlo Results for the Simulation of Tube Burst (the Upper Bounds are at 95% confidence.)				
Number of Bursts	Frequency of Occurrence	Probability of Burst	Upper Bound Pr(Burst)	Upper Bound Cumulative Pr(Burst)[1]
1	1397.8	$2.796 \cdot 10^{-2}$	$2.920 \cdot 10^{-2}$	$2.920 \cdot 10^{-2}$
2	26.6	$5.320 \cdot 10^{-4}$	$7.356 \cdot 10^{-4}$	$2.974 \cdot 10^{-2}$
3	0.3	$6.000 \cdot 10^{-6}$	$7.851 \cdot 10^{-5}$	$2.975 \cdot 10^{-2}$
Notes: 1. Based upon cumulative PoB values calculated using the cumulative frequency of occurrence.				

Table 6-9: Statistics of Simulated Burst Pressures	
Parameter	Value (ksi)
Minimum	0.582
Maximum	14.854
Mean	7.677
Mode	7.650
Median	7.750
Standard Deviation	1.187
Mode to SLB Margin	5.090

Figure 6-1: Example BOC Distribution of Indications

Indications Adjusted for PoD = 0.6

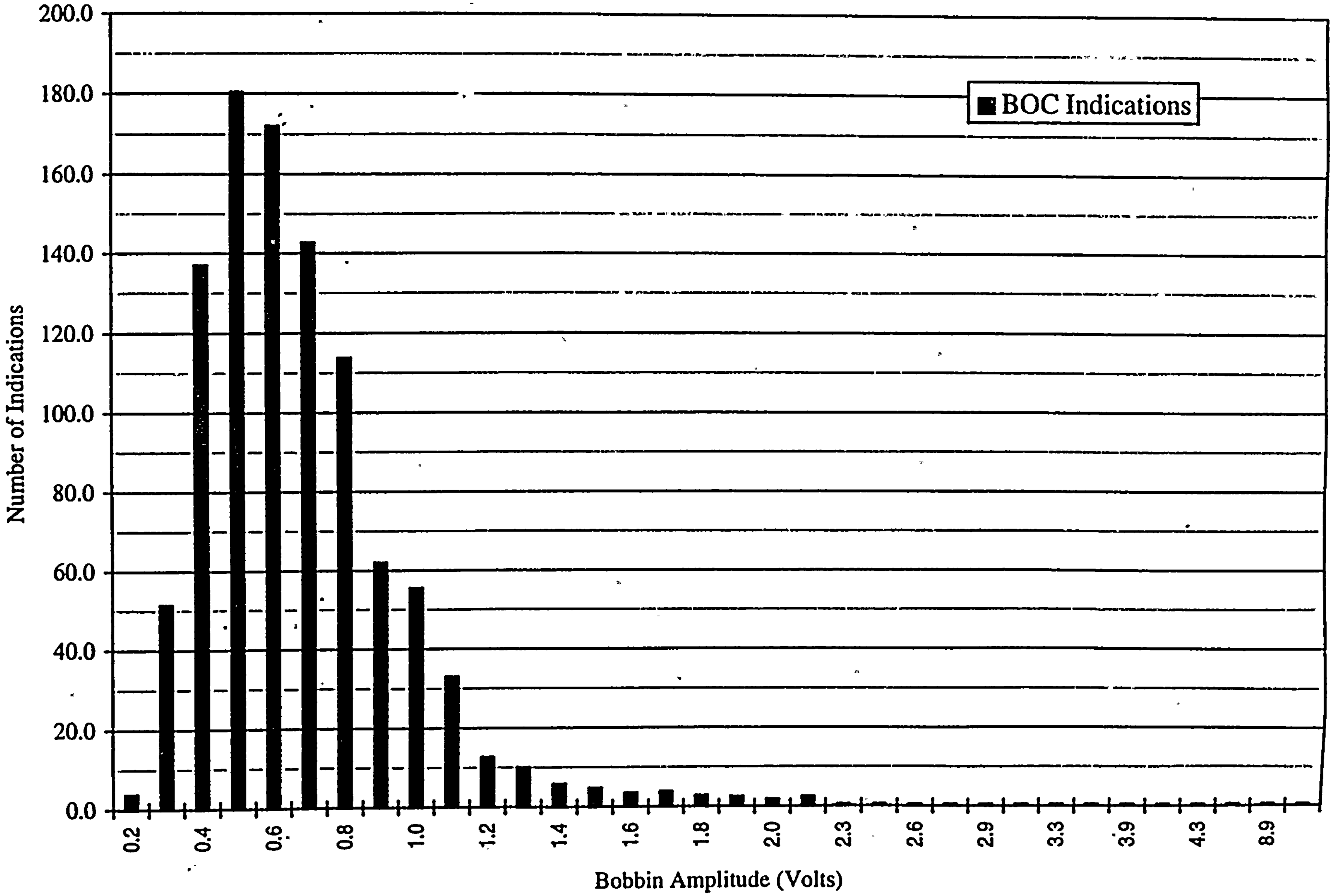


Figure 6-2:Example Cumulative Growth Distribution Used for Monte Carlo Analyses

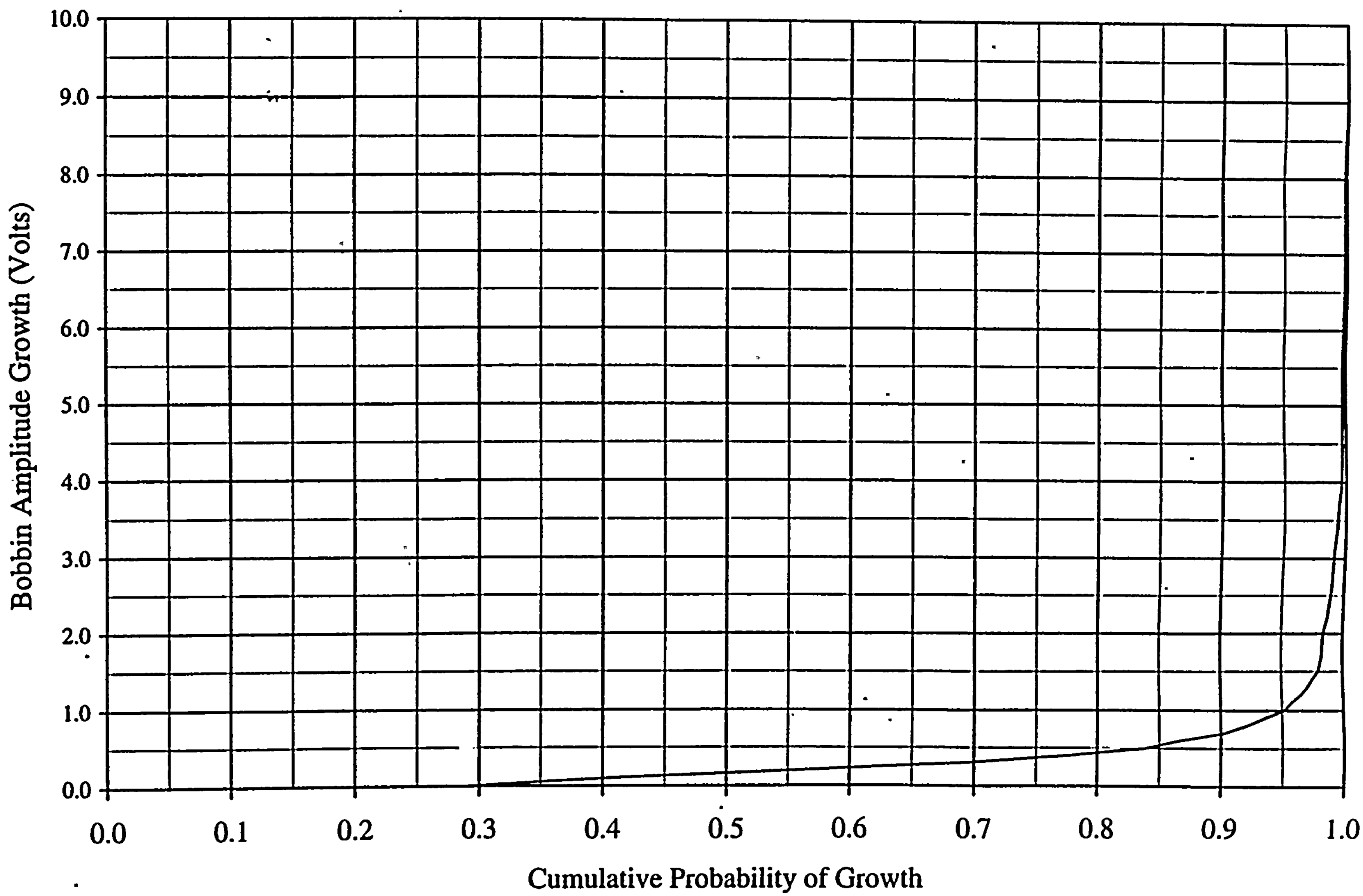
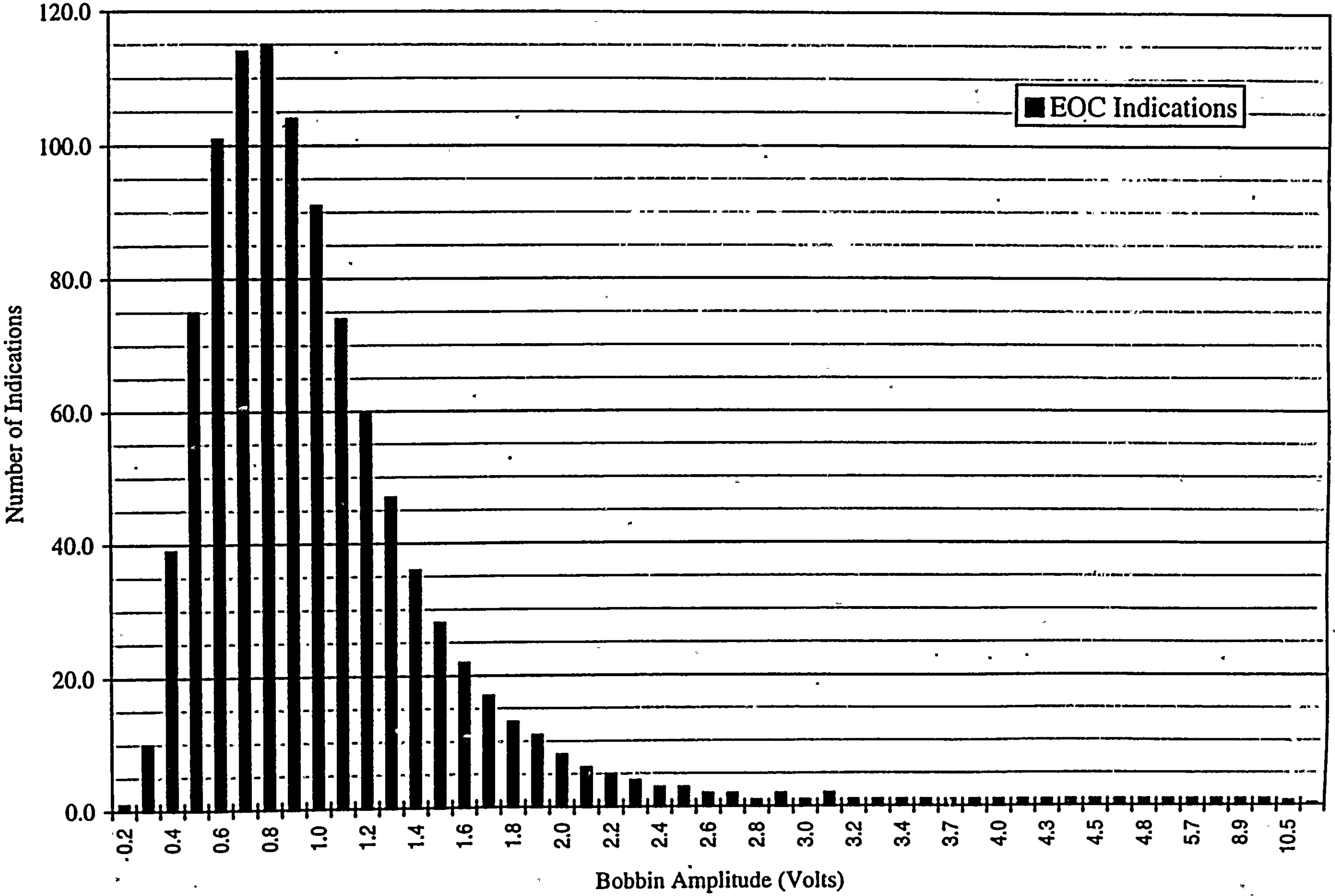


Figure 6-3: Example of SG EOC Distribution of Indications



Appendix A: Regression Analysis

A.1 Introduction

The analysis of the relationship between two variables is generally termed either *regression* analysis or *correlation* analysis. In addition, one may also find the term *confluence* analysis in the literature [1,2,3,4,5,6,7]. For each, the objective is to establish a mathematical model describing a predictive relationship between the variables. The use of the term regression is frequently interpreted to imply that some sort of causal relationship exists while correlation has been reserved for non-causal relationships. Other differentiations between the two terms involve the nature of the variables, i.e., whether or not one or both is stochastic. In addition, the term regression is also frequently used to mean the process by which the parameters of a relationship are determined.

For the purposes of the evaluations reported herein the name *regression analysis* is used in the broad sense of covering the aspects of the fitting of a curve, i.e, equation, referred to as the regression curve or line, to observed data points, where concern is with the slope and position of the curve that best fits the data, and to the analysis of how well the data points can be represented by the curve, i.e., the correlation analysis. The correlation analysis has two aspects, one is a measure of the degree of covariability between two variables, and the second is as a measure of the closeness of fit of a regression line to the distribution of the observations. The statistical analysis is performed for the purpose of establishing a stochastic dependence, and does not, nor does it have to, demonstrate the existence of a causal dependence.

For the analyses dealing with the APC it is desired that models be developed relating the burst strength and leak rate of degraded tubes to the morphology, or some element of the morphology, of the degradation. Unfortunately, a complete description of the degradation morphology is only known exactly for tubes which have been destructively examined. However, a third variable, based on the non-destructive examination of the tubes, which is directly related to the morphology of the degradation is available. Each degradation state is taken to correspond to a set of quantifiable characteristics or variables, such as the burst strength (measured by a burst pressure test), the leak rate (measured as a function of differential pressure), and a non-destructive examination (NDE) response, e.g., eddy current bobbin coil signal amplitude in either an absolute or differential mode. Since the field examination of the tubes is based on the NDE response it is appropriate to examine the relationships between the first two variables and the third.

The experimental and field data for outside diameter stress corrosion cracking (ODSCC) at tube support plates (TSPs) consists of bobbin coil voltages and measured tube burst pressure, and leak rates at differential pressures corresponding to normal operating conditions and steam line break (SLB) conditions. As noted, these data are correlated, but not causally related. For example, high burst pressures correlate with low voltages but high burst pressure does not cause low voltage. Similarly, low leak rates are correlated with low voltage, but low leak rates do not cause low voltage.

The degradation process determines the magnitude of the evolution of each variable, however, the degradation process is complex and the morphology and time history will vary even under conditions which would normally be termed identical. Thus, it is expected that the correlation between any pair of the three variables may have significant scatter. This is expected even if each of the variables is measured with perfect accuracy and contains no measurement error.

In order to predict burst pressures and leak rates under postulated conditions for degraded tubing, confirmed by field inspection by eddy current test, it is necessary to develop regression lines which relate average burst pressure to measured voltage and average leak rate to measured voltage. The "conventional" regression lines are usually determined by considering the variable which is to be predicted in the future, e.g., burst pressure, as the regressed variable, and the variable which will be measured in the future, i.e., voltage, as the regressor variable. While regression lines can also be established to predict voltages from measured burst pressures or leak rates, there is no particular reason to do so as these "inverse" correlations do not usually provide useful information beyond that which is obtained by the conventional regression lines.

It is to be noted that the causative factor relative to the magnitude of each variable is the crack morphology, and that none of the three characteristic variables can be considered to be the cause of the other. This means that for any pair, either may be treated as the predictor and the remaining variable treated as the response. Once a correlating relationship has been established, either variable may be used to predict an expected value for the other. For example, a correlating relationship may be mathematically determined using burst pressure as the response and bobbin amplitude as the predictor. Once the relationship is known, a mean bobbin amplitude associated with a given burst pressure or leak rate can be calculated.

Confidence limits for predicted burst pressure or for predicted leak rate can then be established about the regression line using standard statistical methods. The confidence limits which are determined directly from the regressions of burst pressure or leak rate on voltage will be narrower, for a fixed probability level, than the corresponding limits which could be deduced from the inverse regression lines. these correlations can then be used to determine high confidence values for the

structural limit or leak rate, corresponding to the postulated SLB differential pressure.

A.2 The Linear Regression Model

The general, linear (meaning linear in the coefficients), first order regression analysis model relating two variables is given by

$$y_i = a_0 + a_1 x_i + \varepsilon , \tag{A.1}$$

where y_i is taken here as the response or predicted variable, and x_i as the predictor, or regressor. The ε, or error, term accounts for deviations from a perfect prediction. In order to establish confidence and prediction limits on y_i, the error is assumed to be normally distributed with a mean value of zero and a variance that is uniform over the range of interest. An analysis is then performed to determine the best values of a_0 and a_1 to use in equation (A.1). Three methods are commonly used for the analysis, maximum likelihood estimation, least squares (LS), and weighted least squares (WLS). For maximum likelihood analysis the values of a_0 and a_1 are found that maximize the probability of obtaining the observed responses. The use of maximum likelihood analysis is formally correct, however, if the errors are normally distributed, the maximum likelihood estimators (MLE) will be identical the estimators obtained using least squares. If both variables are stochastic and the errors are normally distributed, the application of least squares still leads to the maximum likelihood estimators of a_0 and a_1.

The LS method is based on minimizing the sum of the squares of the errors, also referred to as residuals, between the observed and predicted values, thus, the best values of a_0 and a_1 are those that make

$$\sum_{i=1}^{n} (Y_i - \hat{Y}_i)^2 \tag{A.2}$$

a minimum, where the caret indicates the predicted value,

$$\hat{Y}_i = a_0 + a_1 x_i . \tag{A.3}$$

Expression (A.2) is differentiated with respect to a_0 and a_1 and the resulting expressions set equal to zero and solved for the coefficients. For WLS, the same expression for the errors is established by considering the error term, ε, to be weighted non-uniformly, i.e., the error distribution is

$$\varepsilon_i \sim N(0, \Sigma_i \sigma^2) \tag{A.4}$$

and the expression to be minimized becomes

$$\sum_{i=1}^{n} w_i (Y_i - \hat{Y}_i)^2, \tag{A.5}$$

where the Σ_i and hence the w_i are known. In situations where the variance of the response is not uniform it is possible to find appropriate weights such that the resulting estimators are MLEs. Such a case is the dependence of the probbility of leak on the bobbin voltage. In this case the response is either 0 for no leak, or 1 for leak. A predictive model based on the logistic function can be fitted by transforming the variables and iteratively solving the resulting weight least squares problem.

For the unweighted LS analysis the slope of the regression or correlation line is found to be

$$a_1 = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sum (x_i - \bar{x})^2}, \tag{A.6}$$

where the summation limits are understood. The intercept is then found as

$$a_0 = \bar{y} - a_1 \bar{x} \tag{A.7}$$

if y has been regressed on x. If x is regressed on y the slope will be

$$\frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sum (y_i - \bar{y})^2} \tag{A.8}$$

relative to the ordinate, or y, axis. If this is reckoned to the x axis, i.e., the abscissa of the original coordinates, the slope is

$$a_1 = \frac{\sum (y_i - \bar{y})^2}{\sum (x_i - \bar{x})(y_i - \bar{y})}. \tag{A.9}$$

If the data used for the analysis contains significant scatter, the values found by (A.6) and (A.9) can be quite different. A rough visualization of this can be obtained by picturing the smallest ellipse that can be drawn that envelopes all of the data points. A line connecting the largest and smallest abscissa values of the ellipse will approximate the regression of the y variable on the x variable, while the line connecting the maximum and minimum ordinate values will approximate the regression of the x variable on the y variable.

For the APC analyses, the objective is to relate burst pressure and leak rate to bobbin voltage. This means that bobbin amplitude is depicted as the abscissa variable while burst pressure and leak rate are depicted as ordinate variables respectively. For the conventional regression analysis, these are the corresponding choices for the regressor and regressed variables. However, if conditions dictate, an inverse regression may be performed, thus the depiction does not necessarily imply the direction of the regression analysis performed. The considerations discussed in the introduction indicate that the inverse regression is only worthwhile if additional useful information can be gained from such an analysis.

The expansion of the model to include more terms, e.g., considering burst pressure to be related to the logarithm of the bobbin amplitude by a second order polynomial is still linear regression analysis. If the assumption of constant variance of the residuals is verified, the application of least squares still results in the maximum likelihood estimators of the coefficients of the equation. If the prediction equation is non-linear in the coefficients, e.g., exponential, a transformation may be made to result in a linear equation, or non-linear regression techniques may be necessary. The use of a logarithmic transformation is common, and may result in a stabilization of the variance, i.e., a non-uniform variance before the transformation may become uniform as a result of the transformation. Consideration of a non-linear regression model, e.g., logistic regression [13,14,15], which can be transformed into a linear model, is contained in the body of this report relative to determining the probability of leak as a function of bobbin amplitude.

A.3 Consideration of Variable Error

If the values of the regressed variable, say y, are subject to error, but the regressor, x, is free from error, no bias will be introduced into the mean, i.e., regression predicted mean value of y for a given x, although the variance will be greater due to the errors in the measurement of y. The calculated values of y are then unbiased estimates of the true values of y, assuming the error to also be normally distributed. The only effect of the errors in the measurement of y is to increase the variance of the residuals and render the estimate of y less reliable, i.e., the estimate will have larger inference bounds. If now x is also subject to measurement error the regression will be of observed values on observed values instead of true values on true values. If there is measurement error present in the predictor variable, the slope obtained from the regression analysis will be biased [2,3,6,7,8,9,10,11], but the regression line will still pass through the centroid of the data. The standard regression analysis assumes that the regressor variable is known without error and that the regressed variable is a measured value subject to uncertainty. Thus, for example, the regression of burst strength, P on the logarithm of the bobbin amplitude, $\log(V)$, estimates the mean value P_k for which the observed value of bobbin amplitude is $\log(V_k)$. If the bobbin examination and evaluation technique were to

be changed in the future to reduce the measurement errors the correlations based on current technology might have to be repeated.

If there is significant error present in the measurement of the variables, the regression analysis may be performed using what is termed as the *error in variables* model. In this case, it is assumed that the data measurements are of the form

$$X = x + \eta \quad \text{and} \quad Y = y + \delta\,, \tag{A.10}$$

where X and Y are the measurements corresponding to the true values of the variables x and y, and η and δ are their corresponding errors of measurement. For the predictor, say X, the total variance will be

$$\sigma_X^2 = \sigma_x^2 + \sigma_\eta^2\,. \tag{A.11}$$

It can be shown that when the measurement error is independent of the true value, the expected value of the calculated slope, a_1, will be

$$a_1 = \frac{\alpha_1}{1 + \dfrac{\sigma_\eta^2}{\sigma_x^2}}\,, \tag{A.12}$$

where α_1, is the true value of the slope, or the value that would result if no measurement error was present, and

$$\sigma_x^2 = \frac{\sum (x_i - \bar{x})^2}{n - 1}\,. \tag{A.13}$$

It is noted that a_1 would be found from equation (A.7) as before. A key point to note is that the calculated slope under predicts the true slope (without measurement error). If the measurement error is known, and is uniform, its effect on the analysis slope can be calculated directly and the appropriate slope to be used for prediction would be

$$\alpha_1 = a_1\left(1 + \frac{\sigma_\eta^2}{\sigma_x^2}\right). \tag{A.14}$$

When the error variance is known and can be expressed as a fraction of the variable variance, the slope will be affected by a like amount.

When the error variance is not known, which is usually the case, an estimate of the true slope can be made using the partitioning technique developed by Wald [10] and subsequently improved upon by Bartlett [11]. The technique consists of partitioning the data into three groups based upon the ordered regressor variable. The line joining the centroids of the upper and lower groups is an unbiased and consistent estimator of the true slope. If the slope thus found is close to the slope determined without considering measurement errors then the measurement errors are considered to be not significant. The application of this technique must be done with caution since the order of the true values of the regressor variable(s) is not known, only the order of the measured variables. For the APC analyses the application of the Wald-Bartlett technique is restricted to estimating whether or not significant measurement error is present. For the correlations examined through the writing of this report, it has been concluded that the measurement errors are not significant. Moreover, it may be assumed that the measurement errors are not significant, and a standard regression analysis performed. If the residuals are normally distributed about the regression line, inference bounds may be determined using the standard inference methods.

It is noted that, if the magnitudes of the measurement errors associated with each of the variables, or their ratio, is not known, an "errors-in-variables" analysis does not lead to a criterion for the selection of the best regression direction. In general, the need for performing an "inverse" regression can be based on the determination of whether or not useful information beyond the conventional regression analysis will result.

A.4 Detection of Outliers

If the errors are normally distributed the application of LS to determine the coefficients of the regression equation minimizes the variance of these estimators. The coefficients are also the MLEs. A drawback of the LS technique is that it is not very robust. This means that the fitted line may not be the best estimator of the correct relationship because it can be significantly influenced by potentially outlying data. In addition, the resulting fit may be such that potential outliers become hidden if examined after the analysis is performed. There are established methods for identifying influential data that may result in a distortion of the regression line. Such methods fall into the categories of regression diagnostics and robust regression. Robust regression methods are designed to be insensitive to potential outliers, and can be used to identify outliers based on the residual errors from the robust regression line. A rather simple example of improving the robustness of the fit would be simply minimize the sum of the absolute values of the residuals instead of the sum of the squares. This provides significant improvement if the outlier is in the y-direction for a y on x regression, but is not resistant to outliers in the x-direction.

One very robust technique is termed the "least median of squares," or LMS [12] regression: The *best* regression line (or polynomial) is the one for which the median of the squared residuals is a minimum. The drawbacks to this technique are that there is no closed form solution and techniques for the determination of inference regions would be difficult to apply. However, the determination of a reasonable solution is quite easy using a computer. The algorithm proceeds by drawing sub-samples of a given size from the data set. For each sub-sample,. regression line coefficients and the median of the squared residuals are calculated. The coefficients of the minimum median solution are designated as the LMS solution. A median based scale estimate (analogous to the standard deviation) is determined for the identification of outliers at a two-sided 98% confidence level, or a one-sided 99% confidence level.

The data for the APC were examined using the LMS robust regression program PROGRESS by Rousseeuw and Leroy. It is noted that the application of robust regression is not intended to be used for the justification of the deletion of improbable data points, only for the identification of potential outliers. The rejection of any data must be based on an evaluation of the circumstances surrounding the data collection to search for possible sources of error.

A.5 Selection of a Regression Coordinate System

For the analysis of continuous variable data four, alternatives were examined for each correlation. These choices are listed in Table A.1. For each case, the *correlation coefficient, r,* measuring the "goodness-of-fit" of the regression line was calculated. The correlation coefficient is a measure of the variation of the data explained by the regression line, thus the largest value is indicative of the best fit.

Table A.1: Fitting Options Considered for LS Regression

Abscissa	Ordinate	Relation
Linear	Linear	$y_i = a_0 + a_1 x_i$
Logarithmic	Linear	$y_i = a_0 + a_1 \log(x_i)$
Linear	Logarithmic	$\log(y_i) = a_0 + a_1 x_i$
Logarithmic	Logarithmic	$\log(y_i) = a_0 + a_1 \log(x_i)$

The expression for the square of the correlation coefficient, known as the *index of determination*, is

$$r^2 = \frac{\sum (\hat{y}_i - \bar{y})^2}{\sum (y_i - \bar{y})^2}. \tag{A.15}$$

The index of determination is the proportion of the total variation about the mean of the predicted variable that is explained by the regression line. The scale combination yielding the largest index of determination, and, hence, correlation coefficient, was selected for the analysis. In the event that the predicted variable for the regression is the logarithmic transformation of a physical variable, the above calculation is performed on the untransformed variable. It is readily apparent, however, that for data with a range of several orders of magnitude, e.g., bobbin amplitudes ranging from O(0.1 volt) to O(100 volts), the use of a logarithmic scale is appropriate. It is also to be expected that the variation of observed voltages would be normally distributed about the log of the voltage. The same is true for the leak rate which ranged from O(0.02 l/hr) to O(500 l/hr) for specimens for which leaking was observed. It is to be noted that the use of a logarithmic transformation is commonly used for data with a large range as a variance stabilizing technique. If the dependent variable has been logarithmically transformed, regression line predictions will be of the expected median of future values, not the expected mean of future values.

A.6 Selection of a Regression Direction

As noted in the introduction to this appendix, the bobbin amplitude does not cause the observed burst pressure and vice versa. The same is true for the relation between bobbin voltage and leak rate. Thus, the regression direction is not specified by the choice of variates.

The objective of performing the regression analysis is prediction. For all practical purposes the bobbin voltage will be used as a predictor of burst pressure and leak rate. However, the intended use does not automatically dictate the designation of the predictor and response variable roles for the regression analysis. The LS fit simply finds the line such that the variance of the responses is minimized relative to the regression line. As previously noted, once the LS fit has been performed either variable can be predicted from the other. In addition, inference regions or bands established for prediction in one direction may be similarly use in the reverse direction (although the terminology is changed to discrimination).

For a regression of y on x, the means of all future values of y_0 corresponding to all given x_0 are simultaneously bounded (confidence) with a level of confidence of ≥(1-α)·100% by

$$\left[y_0 - a_0 - a_1 x_0\right]^2 \geq \left(2F_{1-\alpha/2,2,n-2}\right)^{1/2} s^2 \left[\frac{1}{n} + \frac{(x_0 - \bar{x})^2}{\sum (x_i - \bar{x})^2}\right]. \quad \text{(A.16)}$$

Where s^2 is the"standard error of regression," i.e.,

$$s^2 = \frac{\sum (y_i - \hat{y}_i)^2}{n-2} \quad \text{(A.17)}$$

and $t_{1-\alpha/2,\, n-2}$ is found from the Student's t-distribution. Similarly, an individual future value of y_0 for a given x_0 is bounded (prediction) with a level of confidence of ≥(1-α)·100% by

$$\left[y_0 - a_0 - a_1 x_0\right]^2 \geq t^2_{1-\alpha/2,n-2}\, s^2 \left[1 + \frac{1}{n} + \frac{(x_0 - \bar{x})^2}{\sum (x_i - \bar{x})^2}\right], \quad \text{(A.18)}$$

However, for a given y_0, the bounds on x_0, referred to as discrimination bounds, are found by solving equation (A.18) for the values of x_0 that satisfy the equality, although care must be taken relative to the solution, since real roots of equation (A.18) may not exist depending on the results of the data analysis.

If the scatter of the data is small, as for the burst pressure to bobbin amplitude correlation, the regressions of x on y and y on x will yield slopes that are similar. However, for APC analyses the data exhibit significant scatter for the leak rate to bobbin amplitude correlation and the two regression lines have significantly different slopes. In this case it is appropriate to select the regression line based on non-statistical considerations. Such considerations may be known end points of the regression line, e.g., burst pressure for non-degraded tubes, or comparison of the slope with theory based results. For either regression direction, inference regions can be determined.

As noted, equation (A.18) can be used to determine inference bounds regardless of the direction of the regression. In general the magnitudes of the inference bounds will not be identical. However, the confidence level statements are true for both bounds, i.e., one bound is not invalidated by the other. Thus, if a ≥(1-α)·100% lower bound on the burst pressure from the regression of the burst pressure on the logarithm of the bobbin amplitude is higher than the corresponding lower bound from the inverse regression, it simply means that the confidence level of the inverse regression is >(1-α)·100%. Thus, if the residuals are verified to be normally distributed, the lower prediction bound may be taken as the higher of the prediction bounds established by performing the regression analysis in each directions.

A.7 Significance of the Regression

The significance of the regression is evaluated by calculating the improvement in the estimate of the predicted variable based on knowledge of the regressor variable. For the APC analyses this is the same as determining whether or not the estimate of the burst pressure or leak rate for a tube is improved by knowing the bobbin coil voltage amplitude. For a linear, 1st order regression this is the same as testing to determine if a zero slope is probable. If the confidence interval for the slope includes zero then the relationship between the predicted variable and the regressor could be accidental, i.e., due to random error. The actual determination may be made by calculating the confidence interval on the slope, a_1, to see if it includes zero, or by testing the null hypothesis that the true slope is zero. In practice this is stated as

$$\begin{aligned} H_0 &: a_1 = 0 \\ H_1 &: a_1 \neq 0 . \end{aligned} \tag{A.19}$$

If the null hypothesis, H_0, is true, then ratio of the mean square due to regression (SSR) to the mean square due to error (SSE), i.e., the mean square of the residuals, follows an "F" distribution with the regression degrees of freedom (DOF) in the numerator and the residuals degrees of freedom in the denominator. For a linear analysis with k regressor variables and n data points, then

$$\frac{\text{MS Regression}}{\text{MS Error}} = F_{1-\alpha,k,n-k-1}, \tag{A.20}$$

where $100\cdot(1-\alpha)\%$ is the associated confidence level. (Note that α is the area in the tail of the distribution.) If the true value of the slope is zero then both mean square (MS) Regression and MS Error are independent estimators of the true value of the error variance. Since they are both estimates they would not be expected to be exactly equal, however, it would be expected that they would be nearly equal so that their ratio would not be too far from unity. It is noted that the F ratio and the Index of Determination, r^2, are both calculated from the sums of squares of the variables, so

$$F_{1-\alpha,k,n-k-1} = \frac{SSR/k}{SSE/(n-k-1)} = \frac{r^2}{1-r^2}\,\frac{n-k-1}{k}, \tag{A.21}$$

and a critical value of r^2 for a selected critical α can be found as

$$r_{crit}^2 = \frac{k F_{1-\alpha,k,n-k-1}}{(n-k-1) + k F_{1-\alpha,k,n-k-1}}. \tag{A.22}$$

If the value of r^2 found from the regression is greater than the critical value from (A.22) the null hypothesis, H_0, i.e, that the slope is zero, would be rejected, and the alternate hypothesis, H_1, that the true slope is not zero would be accepted. For example, consider k=1, n=15, and α=0.01. Then, from equation (A.22) we find a critical value of r^2 of 0.411 and a critical value of r of ±0.641. If the regression value of r^2 exceeds 0.411 the regression is significant at a level greater than 99%.

For a 1st order regression, equation (A.21) can be rearranged as

$$t_{1-\alpha/2,n-2} = \frac{r}{\sqrt{\dfrac{1-r^2}{n-2}}}, \tag{A.23}$$

i.e., a t distribution with n-2 DOF's. Given a value of r from the regression analysis, a value of t can be calculated and a significance level determined. For the same example as above, we consider r^2=0.411 and n=15. From equation (A.23) we find t=3.013, and a significance level of 100·(1-α)=99.1%, which agrees with the above determination. It is to be noted that for a small number of regressor variables and a large number of data points, the square of the correlation coefficient does not have to be very close to one to reject the null hypothesis and accept the alternate hypothesis that the slope is not equal to zero, thus implying that a correlation does exist. This test is identical to testing the hypothesis of equation (A.19) by calculating t as

$$t_{1-\alpha/2,n-2} = \frac{b_1}{\sigma_{b_1}}, \tag{A.24}$$

where the denominator is the estimated standard deviation of the slope parameter, and then determining the probability of obtaining t at random. The probability thus obtained, i.e., the probability that the slope of the equation for the entire population is zero, is referred to as the p value for the coefficient. If the p value is less than a selected criterion value, e.g., 5%, use of the regression equation is considered justified.

A.8 Analysis of Regression Residuals

To use the results of the least squares analysis, it is assumed that:

1. the expectation function is correct, that the response is given by the expectation function plus a disturbance,

2. that the disturbance is independent of the response function,

3. that each disturbance has a normal distribution about the response function value, that each disturbance has zero mean,

4. that the disturbances (or weighted disturbances) have equal variances, and

5. that the disturbances are independently distributed.

The purpose of analyzing the residuals, i.e., the differences between the actual variable value and the predicted variable value, is to verify each of the assumptions inherent in performing the least squares analysis. There are a variety of plots that can be used for the analysis of the residuals, although not all may be judged necessary for each analysis. A plot of the residual values against the predicted values should be nondescript since the residuals should not be correlated with the predicted values. Such results indicate that the variance is approximately constant (as assumed), that there is no systematic departure from the regression curve, and that the number of terms in the regression equation is adequate. A frequency plot (histogram) of the residual values should appear to be similar to a normal distribution. A plot of the ordered residuals on normal probability paper should approximate a straight line. Any of these plots may be used to verify that the regression residuals are normally distributed, although the results are not obvious from the scatter plot. The normal probability plot offers the advantages that it can easily be used to determine if the mean is approximately zero, and a reasonable estimate of the standard deviation of the residuals may be read directly.

To prepare the cumulative normal probability plot, the residuals are sorted in ascending order and then plotted against an ordinate cumulative percent probability value given by

$$100\,\frac{\left(i - \frac{1}{2}\right)}{n}, \qquad \text{(A.25)}$$

where n is the number of data points used in the regression and i is an index ranging from 1 to n. If a small number of outliers have been omitted from the regression analysis, but the depiction of their residuals is desired, n may be taken

as the total number of data points and the residuals of the outliers included accordingly. This has the effect of compressing the spread of the outliers along the probability axis, but generally will not affect the conclusions relative to the linearity of the plot. The rationale for the cumulative probability values used is if the unit area under the normal curve is divided into n equal segments, it can be expected, if the distribution is normal, that one observation (residual) lies in each section. Thus, the i^{th} observation in order is plotted against the cumulative area to the middle of the i^{th} section. The factor of 100 is used to convert the scale to percent probabilities.

If the plotted residuals approximate a visually fitted straight line, it may be concluded that they are normally distributed about the regression curve. The residual value where the line crosses the 50% probability value is an estimation of the mean of the residuals, and can be used to verify that the mean is approximately zero. The residual distance from the 50% point to the 84% point is an approximation for the standard deviation of the residuals. If the residuals for the outliers have been included in the plot they will distort the results obtained for the mean and standard deviation, with the mean value being less affected. For this type of plot, the outliers in the data, if any, will tend to appear on the far left in the lower half of the residual normal plot and on the far right in the upper half, i.e., large negative and positive residual values. The results from the normal probability plot may be used to determine the need for preparing any of the other plots, i.e., it may be apparent that no additional information would be available from a scatter plot.

A.9 References

1. Yamane, T., Statistics: An Introductory Analysis, 2nd Edition, Harper & Row, New York (1967).

2. Draper, N.R., and Smith, H., Applied Regression Analysis, Second Edition, Second Edition, John Wiley & Sons, New York (1981).

3. Hald, A., Statistical Theory with Engineering Applications John Wiley & Sons, New York (1952).

4. Weisberg, S., Applied Linear Regression, John Wiley & Sons, New York (1985).

5. Lipson, C., and Sheth, N. J., Statistical Design and Analysis of Engineering Experiments, McGraw-Hill, New York (1973).

6. Deming, W. E., Statistical Adjustment of Data, Dover Publications, New York (1964).

7. Mandel, J. The Statistical Analysis of Experimental Data, Dover Publications, New York (1984).

8. Davies, O. L. (Editor), *Statistical Methods in Research and Production*, Oliver and Boyd, London, England (1957).

9. Fuller, W.A., Measurement Error Models, John Wiley & Sons, New York (1987).

10. Wald, A. *Fitting a Straight Line When Both Variables are Subject to Error*, Annals of Mathematical Statistics, Vol. 11, pp. 284-300 (1940).

11. Bartlett, M.S., *Fitting a Straight Line When Both Variables are Subject to Error*, Annals of Mathematical Statistics, Vol. 5, pp. 207-212 (1949).

12. Rousseeuw, P.J., and Leroy, A.M., Robust Regression and Outlier Detection, John Wiley and Sons, New York (1987).

13. Hosmer, D.W., and Lemeshow, S., Applied Logistic Regression, John Wiley & Sons, New York (1989).

14. CSS: Statistica Users Manual, Statsoft (1991).

15. SAS/STAT User's Guide, Revision 6, Fourth Edition, Volume 2, Chapter 27, *The Logistic Procedure*, SAS Institute, Inc., Cary, North Carolina (~1990).