International Journal of Reliability and Applications
Vol. 5, No. 1, pp. 25-36, 2004

Model for Predicting Ultrasonic NDE Reliability and Statistical Data Analysis of Piping Inspection Round Robin

Ik Keun Park
Department of Mechanical Engineering, Seoul National University of Technology, 172, Gongneung-dong, Nowon-gu, Seoul, 139-743, Korea.

Hyun Mook Kim
Precision Mechanical Engineering, Graduate School of HanYang University 17, Haengdang-dong, Seoul, 133-791, Korea

Abstract. Ultrasonic inspection system consist of the examination procedures, equipment, and operators. The reliability of nondestructive testing is influenced by the inspection environment, materials and types of defect. It is very difficult to estimate the reliability of NDT due to the various factors. Piping inspection round robin was conducted to quantify the capability of ultrasonic inspection during in-service. In this study, the models for predicting the ultrasonic NDE reliability by logistic model and linear regression model are discussed. The utility of the NDT reliability assessment is verified by the analysis of the data from round robin test with these models.

Key Words : *reliability, ultrasonic inspection, round robin, probability of detection (POD), sizing.*

1. INTRODUCTION

Nondestructive evaluation (NDE) is often the primary basis for establishing the initial flaw size that is used as the basis safe life analysis of components, structure and system. NDE methods are now applied in many industries to help guarantee the safety and reliability of components and system. Ultrasonic NDE is one of the important technologies in the life-time maintenance, and useful to inspect various types of weldments. Recently, an advanced NDE technology has been developed that allows NDE detectability and reliability issues to be treated quantitatively at all stages of the design and manufacturing.

* Corresponding Author
E-mail address: ikpark@snut.ac.kr

In order to able to quantitatively estimate the influence that NDE will have on assuring the reliability of a component, it is necessary to have three types of information: 1) a measure of the signals expected in a given NDE test of the component, 2) a measure of the variabilities associated with the NDE test, and 3) a methodology for combining those expected NDE signals and variabilities with various design parameters so that decisions can be made on likelihood of the NDE method to be detected critical flaws before the part fails.

The reliability of results in ultrasonic inspection system (equipment, procedure and operator) is affected by its ability. It is reported that frequently existing in-service inspection (ISI) ultrasonic testing (UT) methods can not detect even quite large defects in mock-up specimens, fail to size the defect, and are dependent upon an inspector's skill and physical condition. Furthermore, the reliability of ultrasonic ISI is influenced by the inspection environment, other materials and types of defect. Therefore, it is very difficult to estimate the reliability of NDE due to various factors.

In current practice, the probability of detection (POD) curve is normally given by an assumed distribution function which can be characterized by a few free parameters that are determined by empirical tests(Berens(1989)). However, if POD is obtained through the use of models, then no distribution shape need be assumed a priori. In analysis of a large number of nondestructive size data, a linear relation between the logarithm of a measured flaw size and the logarithm of the true flaw size with normally distributed deviation has proved satisfactory(Panhuise, et al.(1989)). In typical NDE reliability studies, relatively few inspections are performed on each flaw in the specimen set. The small number of samples which are typical of field inspection data lead to two types of problems with POD models. It is possible to have a sample which is not a good representation of the actual population, giving rise to poor estimate of POD. And, the confidence levels on the POD-flaw size relationship are often extremely broad for small sample sizes.

In this paper, the model for predicting the ultrasonic NDE reliability by logistic is discussed. The utility of the NDT reliability (POD and sizing performance) is verified by the analysis of the data from piping inspection round robin (PIRR) test with these models.

2. NDE RELIABILITY MODELS

2.1 POD Model

The reliability of NDE has been defined in Metals Handbook published by the American Society of Metals as a quantitative measure of the efficiency of procedures in finding flaws of specific type and size(Panhuise, et al.(1989)). The fundamental definition of probability of detection is that the ratio of the number of flaws detected by a given technique to the total number of actual flaws present in the inspected components. POD is a well established measure of inspection performance that is directly related to important issues such as accept-reject criteria, frequency and quality of inspection, etc.(Bray and Stanley(1989)).

Assume that each crack of size a in the potential population of cracks has its own distinct crack detection probability density function of the detection probabilities is given by $f_a(p)$. Figure 1 shows a schematic of distribution of detection probabilities for cracks of fixed length. The conditional probability of a randomly selected crack from the population having detection probability of p and being detected at the inspection is given by $pf_a(p)dp$. The conditional probability of a randomly selected crack from the population being detected is the sum of the

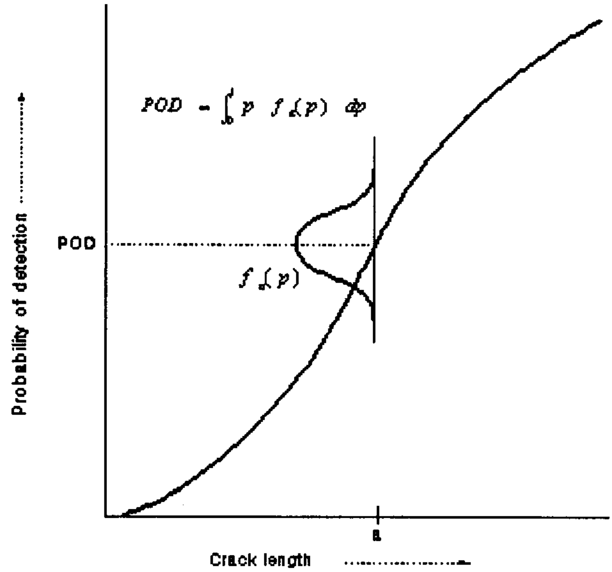

Figure 1. Schematic of distribution of detection probabilities for cracks of fixed length

conditional probabilities over the range of . That is :

$$POD\,(a) = \int_0^1 pf_a(p)dp \tag{1}$$

Therefore, $POD(a)$ is the average of the detection for cracks of size a.

Equation 1 implies that the $POD(a)$ function is the curve through the averages of the individual density functions of the detection probabilities. This curve is he regression equation and provides the basis for testing assumptions about the applicability of various $POD(a)$ models. In Berens (1989), seven different functional forms were tested for applicability to available POD data, and it was concluded that the log-logistics (log odds) function best modeled the data and provided and acceptable model for the data sets of the study. Note that the log odds model is commonly used in the analysis of binary (hit/miss) data because of its analytical tractability and its close agreement with the cumulative log normal distribution.

Two mathematically equivalent forms of the log odds model have subsequently been used. The earliest form is given by :

$$POD(a) = \frac{\exp(\alpha + \beta \ln a)}{1 + \exp(\alpha + \beta \ln a)} \tag{2}$$

This parameterization can also be expressed as :

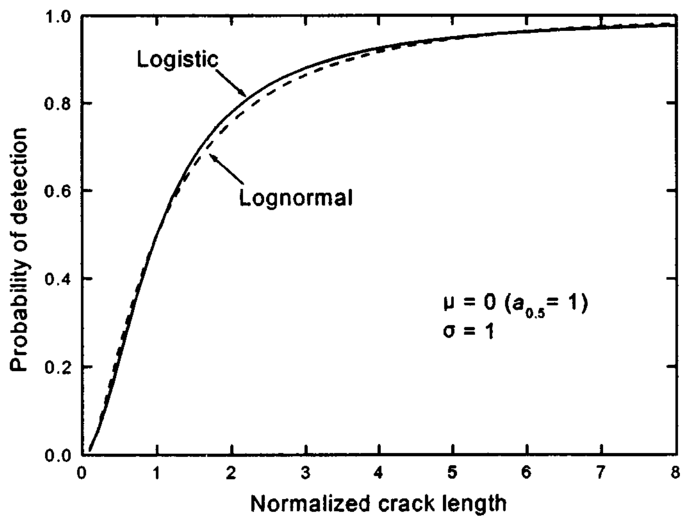

Figure 2. Comparison of logistic and cumulative log normal models

$$\ln\left[\frac{POD(a)}{1 - POD(a)}\right] = \alpha + \beta \ln(a) \tag{3}$$

In the Equation 3 form, the log of the odds of the probability of detection (the left-hand side of Equation 3) is expressed as a linear function of $\ln(a)$ and is the source of the name of the log odds models. Note that given the results of a large number of independent inspections of the model can be fit with a regression analysis. This regression approach will not be discussed further, because the maximum likelihood estimates can be applied to much smaller samples of inspection results and can give equivalent answers for large sample sizes.

Although the parameterization of Equation 2 and 3 are sensible in terms of estimation through regression analyses, α and βare not easily interpretable in physical terms. A mathematically equivalent form of the log odds $POD(a)$ model is given by :

$$POD(a) = \left\{1 + \exp - \left[\frac{\pi}{\sqrt{3}}\left(\frac{\ln a - \mu}{\sigma}\right)\right]\right\}^{-1} \tag{4}$$

In this form, $\mu = ln\ a_{0.5}$ where $a_{0.5}$ is the flaw size that is detected 50% of the time, that is, the median detectable crack size. The steepness of the *POD(a)* function is inversely proportional to σ; that is, the smaller the value of , the steeper the *POD(a)* function. The parameters of Equation 2 and 4 are related by :

$$\mu = \frac{-\alpha}{\beta} \tag{5}$$

$$\sigma = \frac{\pi}{\beta\sqrt{3}} \tag{6}$$

The log odds *POD(a)* function is practically equivalent to a cumulative log normal distribution with the same parameters, μ and σ of Equation 4. Figure 2 compares the log odds and cumulative log normal distribution functions for $\mu=0$ and $\mu=1$. Equation 4 is the form of the logistic model that will be used in the section "Analysis of Hit/Miss Data" in this article.

2.2 Sizing Model

ASME Section XI, Appendix VIII provides requirements for performance demonstration for ultrasonic examination procedures, equipment, and personnel used to detect and size flaws(Appendix VIII of ASME B&PV Code Sex. XI). The operator shall meet the requirements of Appendix VII and shall be qualified in accordance with VIII-3000. Least-squares regression analysis is used to estimate the sizing performance. Figure 3 shows the definition of statistical parameters. Line A is linear regression line, $y = a + bx$ giving the best fit of n data points $(x_1, y_1), \ldots, (x_n, y_n)$ obtained by the least square method. Where, y intercept :

$$a = \frac{\sum y_i}{N} = b\frac{\sum x_i}{N}$$

slope of the regression line :

$$b = \frac{N\sum x_i y_i - (\sum x_i)(\sum y_i)}{N\sum x_i^2 - (\sum x_i)^2}$$

And, line b is ideal line $y = x$ (perfect UT measurements)

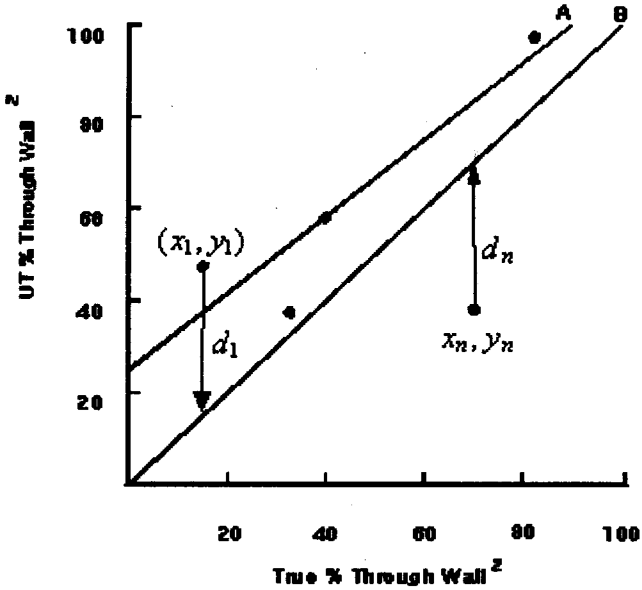

Figure 2. Definition of statistical parameter

Correlation coefficient defined as

$$r = \frac{n\sum x_i y_i - (\sum x_i)(\sum y_i)}{\sqrt{(n\sum x_i^2 - (\sum x_i)^2)(n\sum y_i^2 - (\sum y_i)^2)}}$$

is a measure of "how well" the least square regression line fits the data with respect to the ideal of $y = x$

3. PIPING INSPECTION ROUND ROBIN

Piping inspection round-robin (PIRR) was conducted in 2001 at the Korea Institute of Nuclear Safety (KINS) to qualify the capability of ultrasonic inspection for in-service and to address some aspects of reliability for this type of NDE. Two inspection groups participated in the round robin a total of 9 companies that comprised 15 commercial inspection teams employed by commercial in-service inspection companies. An individual team (consisting of level Ⅱ and Ⅲ inspectors) conducted ultrasonic examinations on welded pipes. Two different types of flaws were implanted into the specimens (EDM notches and thermal fatigue cracks). The round-robin measured the detection and sizing capabilities of fifteen inspection teams who employed procedures that met or exceeded ASME Code Section XI requirements. The specimens are inspected under conditions that simulate as closely as practical the actual application conditions. Tabulated sheets were provided for the tested personnel to keep their records. POD curves are also constructed for each company and/or their participating operator. Superiority of the NDE performance for the participating companies as well as their operator was thus revealed. Figure 3 shows the procedure of reliability analysis of round robin test results.

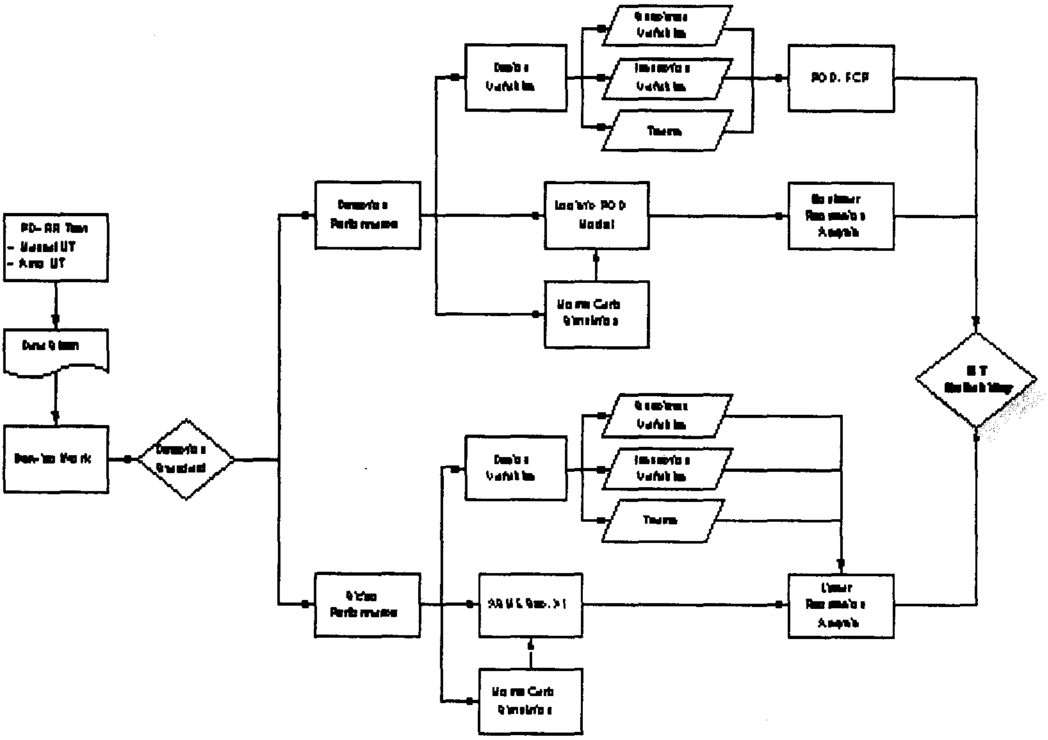

Figure 3. The procedure of reliability analysis of round robin test

3.1 Design variables

Under usual field environment, POD shows various spectrum. Design variables of a test are those which defines conditions of tests and materials influencing spectrums of such various conditions. Here, a variable is a discrete independent variable which does not influence other variables. Seven design variables were selected to uniquely define an inspection condition. While more variables could be added to this list, these seven were considered to be the most important, and should account for most of the variation in the test results. The seven variables are described in table 1.

Table 1. Design variables

Design Variable	Inspection Conditions Defined by the Variables
Material Type	304L Stainless Steel SA 312 TP347 Stainless Steel
Defect Type	Thermal Fatigue Crack(TFC)

	EDM Notch
Defect Geometric	Axial Crack Circumferential Crack
Defect Size	Blank, Size 1, Size 2, Size 3, Size 4
Inspection Group	ISI Vendor and NDT Co. Ltd.
Inspection Team	15 Teams
Procedure Type	ASME Code or Advanced

4. RELIABILITY DATA ANALYSIS

4.1 Relative Importance of Variables

The PIRR experiment was designed to determine the effect of important inspection and material variables on detection performance. The statistical significance of design variables is calculated on detection performance. The design variables evaluated are the inspection variables and material variables. Table 2 and 3 present summaries of detection data obtained for each material in the PIRR. These contingency tables present the basic data used to determine whether or not the listed variables significantly affect detection. In these tables, POD and FCP (false call probability; the probability that a blank grading unit receives any indication) detection statistics are presented, as well as the number of inspection performed. These tables demonstrates that detection in the defect size and type are highly significant to detection performance.

4.2 POD Performance

Figure 4 shows a plot of the logistic function showing detection performances of all the teams. This will means the result of detection skill of ultrasonic test in Korea. As crack size increases, detection probability increases. POD curves were constructed to plot the relationship of POD to crack depth and length as the independent variables, using mathematical regression techniques to fit the curve to

Table 2. Effect of inspection variables

	Group A		Group B	
Defect Size	ASME	Advanced	ASME	Advanced
Blank				
FCP	0.152	0.400	0.286	0.450
# Insp.	46	20	70	20
Size 1				
POD	0.350	0.063	0.071	0.063
# Insp.	40	16	56	16
Size 2				
POD	0.519	0.583	0.548	0.417
# Insp.	27	12	42	12

Size 3				
POD	0.553	0.625	0.696	0.563
# Insp.	38	16	56	16
Size 4				
POD	0.792	1.000	0.886	0.500
# Insp.	24	10	35	10

FCP : false call probability
POD : probability of detection
Insp. : Number of inspection

Table 3. Effect of material variables

Size / Type	Size 1	Size 2	Size 3	Size 4
EDM Notch				
POD	--	0.800	0.800	--
# Insp.	--	15	15	--
TFC				
POD	--	0.567	0.667	0.800
# Insp.	--	30	15	15

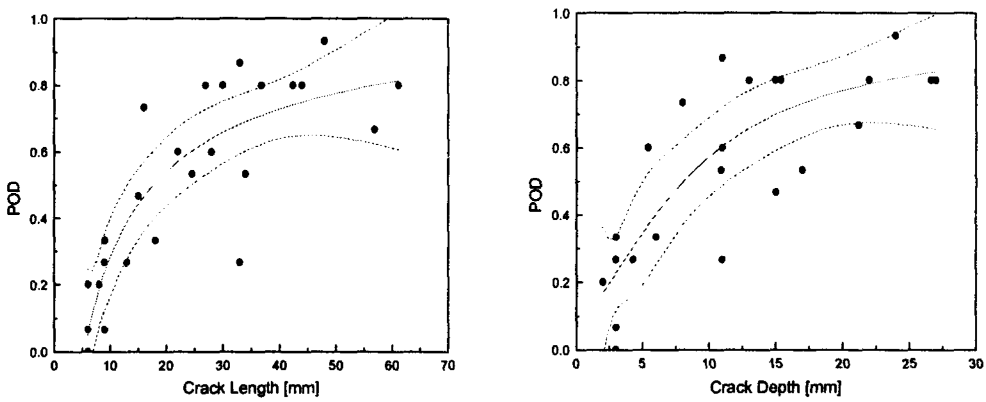

Figure 4. Logistic curve fit to POD data with 95% bounds about all teams (a) POD vs length, (b) POD vs depth

Table 4. Parameters of POD from logistic curve fit

	Length		Depth	
Team	μ	σ (1/mm)	μ	σ (1/mm)
All	−1.768	1.072	0.096	0.935
Group A	−0.428	0.854	0.171	0.899
Group B	−0.551	0.832	−0.010	0.969
ASME	−0.372	0.830	0.093	0.954
Advanced	−0.514	0.770	−0.021	0.906

the experimental data. The logistic curves are surrounded by 95% confidence bounds and the raw POD points used in the fit are illustrated on the plots. Each POD point describes the detection results on an individual defect in PIRR. In these results, statistical analysis of POD hit/miss data using logistic POD model was found to be very feasible to the reliability assessment of the NDE data sets.

Table 4 summarizes the curve fits for POD vs. depth and length. Results are shown for various inspection conditions and for various teams. The table lists the fit parameters μ for Equation 5, along with their standard deviations σ. the results of the logistic fits for the various inspection conditions are given in this table while Figure 4 is the regression curve. A comparison of the two group indicates that the POD vs depth performance of Group A is better than that of Group B.

4.3 Sizing Performance

Linear regression was employed to analyze sizing errors in both depth and length. Table 5 presents the regression fits that relate true crack length to measured length. For length sizing, the average regression slopes is close to zero, indicating relationship between the measured and true depths. we find good performance compare to depth sizing. Table 6 summarizes the fitted regression results for depth sizing. The table shows that depth sizing results were very poor. Depth sizing capabilities are poor. These plots confirm the results displayed in table 6. The visual overview of the sizing results is given in figure 5, which plot the measured vs true depth for all teams.

Table 4. Summary of linear regression fits for defect length sizing

Team	a	b	r	M.D.(mm)
All	6.445	0.888	0.778	7.955
Group A	4.006	0.933	0.808	6.743
Group B	8.747	0.842	0.750	8.762
ASME	7.353	0.836	0.759	7.955
Advanced	3.238	1.049	0.842	7.955

a = intercept of y(x)
b = slope of y(x)
r = correlation coefficient
M.D. = mean deviation

Table 5. Summary of Linear Regression Fits for Defect Depth Sizing

Team	a	b	r	RMS(mm)
All	5.234	0.251	0.274	10.473
Group A	8.168	0.181	0.209	9.975
Group B	2.348	0.342	0.371	17.455
ASME	4.552	0.168	0.252	11.168

Advanced	8.315	0.450	0.406	8.562

a = intercept of y(x)
b = slope of y(x)
r = correlation coefficient
RMS = root mean squared

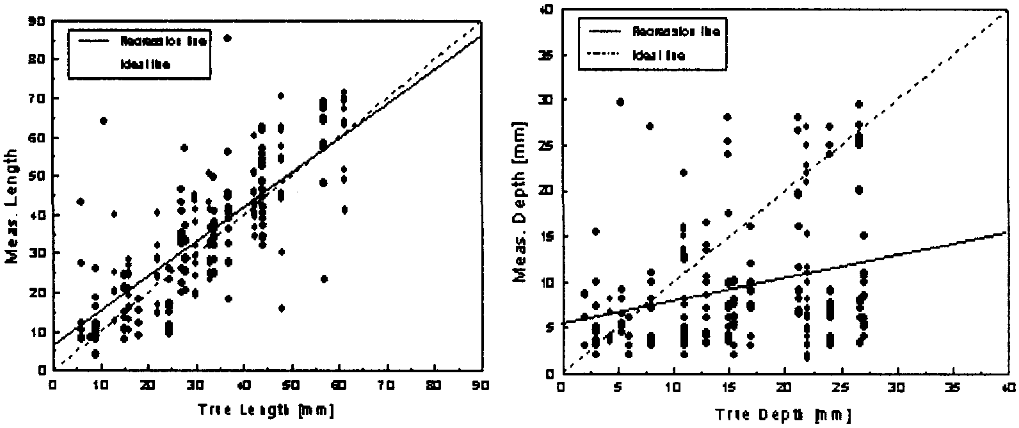

Figure 5. Regression fit of length and depth measurements of all teams

5. CONCLUDING REMARKS

The statistical reliability assessment of ultrasonic inspection system used logistic probability model. The utility of the NDT reliability assessment is verified by the analysis of the data from round robin test with this models. In case of performance of sizing, length sizing showed excellency. However, depth sizing was proven to be less accurate. It is necessary to supplement and develop methods that can improve the accuracy of length sizing. When detecting defects in actual cracks, such as thermal fatigue crack, there is higher possibility of which detection ratio and size measuring performance to get lower. Therefore, in order to improve the reliability of in-service inspection using UT, it is necessary to induce performance demonstration system and assess uncertainty of UT test results to reflect them in the inspection results.

REFERENCES

V. E. Panhuise, et al. (1989), Quantitative Nondestructive Evaluation, in Metal Handbook, 9th ed., **17**, 663-715.

Berens, A. P. (1989). NDE Reliability Data Analysis, Metals Handbook 17, ASM International, Metals Park, Ohio, 689-701.

Bray, D. E. and Stanley, R. K. (1989). *Nondestructive Evaluation*, McGraw-Hill Book Co., New York.

Appendix VIII of ASME B&PV Code Sex. XI, Performance Demonstration for Ultrasonic Examination System.

Ding Keqin (2000). A quantitative Method for Determining the Flaw Size in the Structure, 15th WCNDT, idn 107.

David S. Forsyth, Abbas Fahr, in *Review of Progress in QNDE*, **20**, 2167,

Paliou, C. and Shinozuka, M. (1987). Reliability and Durability of Marine Structures, *J. Structural Eng., ASCE,* **113-6,** 1297.

Bogdanoff, J.L. and Kozin, F. (1985). *Probabilistic Models of Cumulative Damage*, John Wiley & Sons.

Heasler, P. G. and Doctor, S. R. (1996). Piping Inspection Round Robin, NUREG/CR-5068 PNNL-10475.

Raj, B., Jayakumar, T. and Thavasimuthu, M. (1997). Practical Non-Destructive Testing, Narosa Publishing House, London, 133-139.