AECL EACL

AECL-12005

CA0000140

NDE Reliability Gains From Combining Eddy-Current And Ultrasonic Testing

Gains de fiabilité obtenus par des essais non destructifs combinant des contrôles par ultrasons et par courants de foucault

D. Horn, W.R. Mayo

1999 January

AECL-12005

NDE Reliability Gains From Combining Eddy-Current And Ultrasonic Testing

Gains de fiabilité obtenus par des essais non destructifs combinant des contrôles par ultrasons et par courants de foucault

D. Horn, W.R. Mayo

1999 January

AECL

NDE RELIABILITY GAINS FROM COMBINING EDDY-CURRENT AND ULTRASONIC TESTING

by

D. Horn and W.R. Mayo

Components and Systems Division
Nondestructive Testing Development Branch
Chalk River Laboratories
Chalk River, ON K0J 1J0

1999 January

AECL-12005

EACL

GAINS DE FIABILITÉ OBTENUS PAR DES ESSAIS NON DESTRUCTIFS COMBINANT DES CONTRÔLES PAR ULTRASONS ET PAR COURANTS DE FOUCAULT

par

D. Horn et W.R. Mayo

RÉSUMÉ

Les auteurs étudient des méthodes statistiques pour combiner les résultats de deux techniques de contrôle complémentaires : contrôles par ultrasons et par courants de Foucault. Ils comparent la fiabilité des décisions d'acceptation ou de rejet fondées sur les renseignements combinés à celle qui est fondée sur chaque technique de contrôle prise individuellement. La fiabilité calculée augmente en raison de la quantité de renseignements intégrée dans la décision.

Composants et Systèmes
Mise au point des essais non destructifs
Laboratoires de Chalk River
Chalk River (Ontario) K0J 1J0

Janvier 1999

AECL-12005

AECL

NDE RELIABILITY GAINS FROM COMBINING EDDY-CURRENT AND ULTRASONIC TESTING

by

D. Horn and W.R. Mayo

ABSTRACT

We investigate statistical methods for combining the results of two complementary inspection techniques, eddy-current and ultrasonic testing. The reliability of rejection/acceptance decisions based on combined information is compared with that based on each inspection technique individually. The measured reliability increases with the amount of information incorporated in the decision.

Components and Systems Division
Nondestructive Testing Development Branch
Chalk River Laboratories
Chalk River, ON K0J 1J0

1999 January

AECL-12005

1. INTRODUCTION

A number of nondestructive evaluation techniques, including eddy-current and ultrasonic testing (ET and UT), can be applied to the inspection of pressure tubes in nuclear power plants. The question arises whether the combination of two or more techniques can give a significantly more reliable result than a single technique taken individually. A logical response to this question suggests that incorporating more information in the decision process should lead to a better decision. This reasoning underlies the present inclusion of ET with UT in manufacturing inspections and is the basis for current discussions of multiple-technique in-service inspection capability. Nevertheless, a quantitative analysis was needed to (i) demonstrate the validity of the assumption, (ii) determine whether the reliability gain is large enough to merit implementation, and (iii) assess whether the increase in "false-alarm" rejections is acceptably small.

Quantitative means to evaluate inspection reliability for various NDE techniques have been well established for some time [1-3] and statistical approaches to combining data from separate measurements are standard tools in science and engineering [4]. Specialized techniques for making decisions based on multiple (and sometimes conflicting) inputs have recently been developed and collected under the name "data fusion" [5]; some investigators have already applied data fusion concepts to NDE [6,7]. The present report compares the reliability of combined ET and UT measurements to that of each technique taken individually for a database of 108 manufactured flaws [8].

In this paper, a description of the flaw database is followed by a review of reliability assessment techniques. We present a number of data combination methods and apply examples of them to the data. Their relative efficacy is discussed, and our findings are summarized in the final section.

2. FLAW DATABASE

A database of 108 artificial flaws was established as part of an earlier inspection reliability investigation [8]. The flaws had been produced by notching or drilling pressure-tube billets, after which pressure tubes were extruded and finished normally. The database from that work contains ET, UT, and destructive (metallographic) examination results for a variety of flaw types and sizes. While a number of different inspection criteria and setups were employed in that work, we use only data from the most recent of the eddy-current and ultrasonic manufacturing inspection procedures reported there. The ultrasonic inspection made use of the same type of transducer that is now used in Ontario Hydro's Channel Inspection and Gauging Apparatus for Reactors (CIGAR). Two manifestations of those inspection results are addressed in the present study:

(i) The accept/reject decisions as are reviewed and used to demonstrate an improved probability of flaw rejection for a logical OR of ET and UT decisions.

(ii) ET signal amplitudes from static measurements noted in the flaw catalog and UT signal amplitudes from dynamic measurements (the original strip-chart recordings) are used in a quantitative comparison of combination strategies. Though both amplitudes were originally

expressed as "percent full-scale deflection", UT amplitudes exceeding full-scale deflection have been deduced from the attenuation factor required to put them on scale; ET signals have been multiplied by a factor of 5 to give them a similar dynamic range. Table I is a compilation of the results.

3. RELIABILITY ASSESSMENT METHODS

The investigation of inspection reliability is a well-established field with an extensive literature. In this section we will outline the concepts of probability density functions, cumulative probability (*e.g.*, of detection), and relative operating characteristics, as they apply to the individual inspection campaigns (ET and UT) of our data set. These concepts will later be used to compare the reliability of various techniques of data combination.

3.1 Probability Density Function

We can describe the likelihood that a particular signal amplitude will be produced by a given phenomenon, such as a flaw of a specific depth, by a probability density function. An estimate of this probability, written *P(amplitude|depth)*, may be obtained from an analysis of the rates at which various signal amplitudes are produced from a narrow range of flaw sizes. Unfortunately, the quantity required for the application of a threshold is the converse of this, namely the likelihood of having a particular flaw size, based on a given observed signal amplitude. We can renormalize to obtain probability of flaw size given signal amplitude, *i.e. P(depth|amplitude)*, only if the *a priori* flaw size distribution is known. This presents a difficulty, since there is no reason to suppose that the flaw size distribution in pressure tubes will be the same as that for our artificially generated flaws. In this work we adopt the conservative assumption that flaws of all sizes are equally abundant. We present two alternative ways to extract analytical expressions for *P(depth|amplitude)* from the data.

Figure 1(a) illustrates the yield of ET and UT signal amplitudes as a function of flaw depth for this data set (the three flaws deeper than 0.5 mm are not plotted). In one approach, we calculate the point estimates of *P(amplitude|depth)* directly from these data and normalize to unity for each depth bin. If all depths are equally likely, then we can fit the data with a surface that is a normal distribution in amplitude (y) and has width and centroid depending on depth (x), namely,

$$P(amplitude|depth) = \exp\left\{\frac{-(y-ax)^2}{(bx+c)^2}\right\} \Big/ \{0.1772(bx+c)\}.$$

The fitted surfaces, with (*a=1086, b=364, c=58.7*) for ET and (*a=585, b=143, c= 66.7*) for UT, are displayed in Figure 2, panels (a) and (b). The transformation into a joint probability *P(depth|amplitude)* may then be effected through Bayes's Combination Rule, to be discussed in Section 4.3.

In the alternative approach, the yields may be divided by the approximate flaw depth distribution, which is known for the sample data set of manufactured flaws. The depth distribution (Figure 1b)

may be obtained from a histogram of the Figure 1(a) data; the solid curve represents the fitted function,

$$P(depth) = 2.77x\exp(-17.3x).$$

When probabilities in each amplitude bin are subsequently normalized to unity, we have *P(depth|amplitude)* under the assumption that all flaw sizes are equally abundant in an unknown sample. (Since small flaws are presumably more likely to occur than large ones, this is a conservative assumption, which will bias deduced sizes toward larger values.) A Gaussian fit to the resulting point-estimate probabilities gives surfaces of the form

$$P(depth|amplitude) = \exp\left\{\frac{-(x-ay-.036)^2}{(by+.04)^2}\right\}\Big/\left\{177.2(by+.04)\right\}.$$

which represent a normal distribution in flaw depth, x, that has width and centroid varying with signal amplitude, y. The fitted surfaces, with (a=*0.00082*, b=*0.00031*) for ET and (a=*0.00128*, b=*0.00094*) for UT, are displayed in Figure 2, panels (c) and (d).

3.2 Probability of Detection

A common form of reliability analysis in NDE is to determine, for a given threshold, the probability of detection (POD) as a function of flaw size. The POD for each flaw depth interval is the integral of *P(amplitude|depth)* from the threshold signal amplitude to the maximum possible amplitude. A POD curve therefore typically varies from near zero for undetectably small flaws to near unity for unmistakable flaws.

For our particular application, "detection" refers to detection of a rejectable flaw, *i.e.* one having through-wall extent in or above the 0.080-to-0.100 mm measurement interval, which approximates the 0.075 mm manufacturing inspection reject threshold for CANDU pressure tubes. Examples of probability-of-rejection curves obtained by ET are generated from Table II and plotted in Figure 3. The topmost curve shows the point-estimate probability of rejection. This is simply the ratio of the number of flaws rejected to the number of flaws in a given depth range. While the rejection rate appears to be 100% for the larger flaws, this result has a low statistical significance; detection of a single flaw in a size interval where only one is present establishes little confidence that the same rate will hold for subsequent cases.

For this reason, we present the lower-bound probability at the 95% confidence level (lowest curve). This gives a useful lower bound in the range where flaws are plentiful, but not for the larger, less common sizes.

The method of optimum probabilities (see, *e.g.*, [1]) may be applied where ease of detection increases monotonically with the relevant variable. This permits the grouping of the more abundant smaller flaws with the less common large ones to increase sample size. The solid line in Figure 3 is the lower-bound rejection probability at the 95% confidence level, calculated by the method of

optimized probability. This is the technique that will be used in our comparison of probabilities of rejection.

3.3 Relative Operating Characteristics

Since a POD analysis is valid only for a specific threshold, it addresses neither the family of curves produced as the threshold is varied nor the trade-off between type I and type II inspection errors: missed rejections and false alarms. An operating characteristic is a statistical quantity dependent on the error probability in hypothesis testing; a relative operating characteristic (ROC) curve is a parametric expression comparing the legitimate rejection rate to the false-alarm rate as the threshold parameter is varied [2].

Rejectable and acceptable sample regions may have different but overlapping signal distributions, *P(amplitude|rejectable)* and *P(amplitude|acceptable)*, as is illustrated schematically in Figure 4(a). The curves in the illustration are logistic functions, similar in shape to Gaussian distributions, but more convenient to integrate analytically. The ease of discrimination increases with the separation between the distributions, and the optimum threshold value is chosen to balance the expense of false alarms against the consequences of a missed rejection. To quantify this trade-off, an ROC curve may be produced.

Figure 4(b) shows a family of ROC curves generated analytically from the logistic distributions in Figure 4(a). The solid curve is the ROC based on the distribution centroids and widths of the left-hand panel. Decreasing the overlap between the acceptable and rejectable signal distributions by starting with narrower distributions would indicate improved sensitivity for the inspection technique; the resulting short-dashed curve comes close to the ideal (0.0, 1.0) value. Conversely, increasing the overlap by moving the acceptable and rejectable signal centroids closer together produces the long-dashed curve, which indicates decreased sensitivity. The diagonal "random chance" line represents the case of signal distributions that are identical for acceptable and rejectable sample regions, and hence provide no distinction between the two.

4. PRINCIPLES OF DATA COMBINATION

The goal of this work is to assess the reliability of various combinations of ET and UT data. To arrive at the optimum combination, we survey the methods available, ranging from a simple OR of the individual inspection results to more complex techniques collected under the label "data fusion".

4.1 Elementary Combination

The most elementary data combination process uses only the outcome of yes/no decisions resulting from the individual measurements; OR logic serves to reject any sample for which one or more of the individual measurements indicates rejectability. This is a zeroth-order process in that it does not directly bring the first moment (centroid) of the flaw size distributions into the decision. A simple logic table (see Table II) indicates the combined outcomes of the possible individual decisions.

4.2 Averaging

Decisions based on a simple average of signal amplitudes represent a first-order combination process to the extent that the signal amplitudes for each measurement type depend similarly on flaw size. The mean amplitude then corresponds to the average of the first moments of the flaw size distributions.

An average weighted by the widths, or second moments, of the flaw size distributions brings still more *a priori* information into the decision process. The weighted average and variance are commonly taken as $\mu = \frac{\sum_i x_i / \sigma_i^2}{\sum_i 1/\sigma_i^2}$ and $\sigma^2 = \frac{1}{\sum_i 1/\sigma_i^2}$, respectively.

Higher moments, involving skewness and other detailed shape parameters, may also be included in the combination process; they require a correspondingly higher level of knowledge about the flaw size distributions for each measurement technique at each signal amplitude.

4.3 Data Fusion

The desire to achieve a uniform and reliable outcome in complex situations has prompted a great investment of effort in the area of "data fusion" by the U.S. military and others. While the military concept of data fusion encompasses the entire process, from sensors to final decision, the particular items of interest to us are the mathematical techniques for combining incommensurate, uncertain, and sometimes conflicting information into a single result. See Hall [5] for a review. Three techniques relevant to decision-level identity fusion will be considered here: classical inference, Bayesian inference, and the Dempster-Shafer combination rules. In each case, the outcome of interest for NDE applications is a joint probability distribution on which an acceptance or rejection decision may be based.

Classical inference deals with the probability, $P(D|H_i)$, of observing certain data (D) whenever hypothesis H_i is true. It does *not* give the probability of the hypothesis being true. In classical inference, probability distributions are fused by the creation of multidimensional joint probability densities, $P(D_1, D_2, \ldots D_k | H_i)$.

Bayesian inference permits the transformation of the probability, $P(D|H_i)$, into the probability, $P(H_i|D)$, of hypothesis H_i being true whenever D is observed. The transformation requires knowledge of the *a priori* probability that H_i is true, namely $P(H_i)$, independent of any specific observation. Then,

$$P(H_i|D) = \frac{P(D|H_i)P(H_i)}{\sum_j P(D|H_j)P(H_j)},$$

which is known as Bayes's Theorem. The transformation qualitatively described in Section 3.1 to obtain *P(depth|amplitude)* from *P(amplitude|depth)* made use of this theorem with $P(H_i)$ constant

for all i. The joint probability, $P(H_i|D_1,D_2,...D_k)$, that hypothesis H_i is true, given k different declared data from k different sensors is given by Bayes's Combination Rule:

$$P(H_i|D_1,D_2,...D_k)=\frac{P(D_1|H_i)\cdot P(D_2|H_i)\cdot...\cdot P(D_k|H_i)\cdot P(H_i)}{\sum_j P(D_1|H_j)\cdot P(D_2|H_j)\cdot...\cdot P(D_k|H_j)\cdot P(H_j)}.$$

Bayesian inference would be the ideal method for assessing an exhaustive, mutually exclusive set of hypotheses for which the *a priori* probabilities are known. Unfortunately, while NDE outcomes may be defined in terms of exhaustive, exclusive hypotheses (*e.g.* within a certain flaw depth interval or not within it), the *a priori* depth distribution of real flaws in working systems cannot be deduced from the measured distributions in a population of manufactured flaws. Application of Bayesian inference to NDE data fusion therefore requires assumptions which, if conservative, dilute the power of the method and, if aggressive, diminish confidence in the result.

The Dempster-Shafer method avoids the requirements that criteria be exhaustive and exclusive by assigning probability masses to propositions (rather than to hypotheses). The total probability of proposition A_i is then

$$P(A_i)=\sum_{A_i\subset\theta,2^\theta} m(\theta,2^\theta),$$

i.e. the sum of the probability masses of the elementary proposition A_i in the "frame of discernment" (set of elementary propositions), θ, and of all general propositions containing A_i in the set of general propositions, 2^θ.

Information from two sensors may be fused by Dempster's rule of combination to give the total probability mass for the proposition u_n:

$$m(u_n)=\frac{\sum_{A_i\cap B_j=u_n} m_1(A_i)m_2(B_j)}{1-\sum_{A_k\cap B_l=\varnothing} m_1(A_k)m_2(B_l)},$$

which represents the sum of the products of probability masses for which the two sensors agree, divided by a normalization factor based on the cases in which they conflict.

The advantage of the Dempster-Shafer approach is that it provides a means of dealing with conflict between measurements and with ignorance of their implications. Two means of combining probability distributions are available in this context. One approach assumes that all probabilities outside a given measurement interval are in conflict with the probability inside the interval; *i.e.* they represent disbelief of the proposition that the true value lies within the measurement bin. The other takes the view that the probabilities outside the bin represent no particular belief about the primary interval (non-belief). In the special case of normal distributions, the former procedure is the equivalent of combining Gaussian belief functions [9]; the latter is that commonly used in data fusion algorithms, including the nondestructive

evaluation work of Gros *et al.*[6]. The non-belief combination of two probability masses, *e.g.* that a flaw size is in depth interval x_i, then becomes:

$$m(x_i) = \frac{m_1(x_i)m_2(x_i) + m_1(x_i)(1-m_2(x_i)) + m_2(x_i)(1-m_1(x_i))}{1-[0]},$$

or

$$m(x_i) = m_1(x_i) + m_2(x_i) - m_1(x_i)m_2(x_i).$$

Note that, in this case, the sum of the joint probability masses over all the depth intervals, x_i, is not normalized to 1.

5. APPLICATION OF COMBINATION TECHNIQUES

We now apply the combination techniques outlined above to the flaw data base established from the work of reference [8].

5.1 Logical -or Combination

Table III contains the number of flaws for each depth interval, the number rejected by each of ET and UT, and the number rejected by one or both of the techniques. Note that relative to the original work, two flaws have been reassigned to larger depths and the combination technique changed, so that the entries in this table differ from the corresponding values in Table 1 of the original work. The rightmost column of the table, divided by the number of flaws in the corresponding depth bin, therefore represents a point estimate of the "OR"ed rejection rate.

We calculate the lower-bound probability of rejection at the 95% confidence level by the optimized-probability method as in reference [1]. Figure 5 shows the individual and joint rejection probabilities and indicates that in the critical region around 0.1 mm, at and just above the nominal rejection threshold, the combined results are over 50% more effective than either technique individually.

The union of two sets of rejection criteria will give at least as high a rejection rate as either set alone, so a higher probability of rejection is not in itself surprising. The issue in reliability analysis is whether the increased probability of rejection comes at an unacceptable cost in "false alarms". For this data set, ET and UT wrongly reject 8 and 9 out of 39 acceptable indications, respectively, while the combined criterion rejects 12 out of 39.

The data from which Table III and Figure 5 are derived represent the fixed-threshold decisions made in reference [8]. One might ask whether changing the combined-data threshold criterion to reproduce the single-technique false-alarm rate would still give an improved probability of rejection, or, conversely, whether a given single-technique probability of rejection could be achieved with an improved false-alarm rate by the combined data. To investigate this, we need to access the actual signal amplitudes measured (Table I). This permits evaluation of the number of successful rejections (out of 69 rejectable flaws) and the number of incorrect rejections (out of 39

acceptable flaws) as a function of the thresholds for ET and UT. It is also possible, by varying both thresholds, to find the best rejection rate at each false-alarm rate for the logical-OR combination. The results are easily tallied in a spreadsheet.

The appropriate comparisons are best made with a relative operating characteristic analysis, as described in Section 3. ROC curves generated from the point-estimate probabilities of correct rejections and false alarms are shown in Figure 6. Point estimates are adequate for this purpose as the data are effectively binned into just two classes – accept and reject – by the varying threshold parameter; the sample size of each bin is thus more generous than in a POD analysis, reducing the need for an optimized probability calculation. In the figure we see that a false call rate of 0.3 (for flaws between 0.020 and 0.080 mm, not for flawless material) corresponds to a 90% point probability of rejection (of flaws larger than 0.080 mm) for ET and UT individually (filled circles and squares), as compared with 95% for the combined data (filled triangles).

We conclude that for a fixed false-alarm rate, logical-OR combination of ET and UT rejection decisions significantly improves rejection of flaws. Conversely, at a fixed probability of rejection, the false-alarm rate for the logical-OR combination is generally lower than that for either technique individually.

5.2 Averaging

As noted previously, the arithmetic mean of two signal amplitudes will give similar weight to the two only if each has the same amplitude-to-flaw-size relationship. Since correct weighting is most important near the size corresponding to the rejectability threshold, the ET data of Table I have been scaled to be commensurate with the UT data in that depth range. For example, our database contains 20 flaws between 80 and 120 μm in through-wall extent. Their UT signal amplitudes range from 70 to 281, with a mean of 123 and a standard deviation of 52; their scaled ET signal amplitudes range from 0 to 325, with a mean of 129 and a standard deviation of 99, indicating that the two quantities can meaningfully be averaged. The calculated ROC values for a simple average (arithmetic mean) are shown by the apex-down triangles in Figure 6. Note that, in most cases, these approach the "ideal" (0,1) values in the upper left more closely than do the individual and the OR results.

The slightly greater width of the signal amplitude distribution for ET than for UT near the rejectability threshold suggests that the latter should be given somewhat more weight in determining the average. Choice of constant $\sigma_{ET} = 60$ and $\sigma_{UT} = 50$ and use of the weighted average formulae in Section 4.2 gives a further improvement in the ROC curve (open circles) in Figure 6.

More detailed knowledge of the signal properties is best incorporated by direct combination of the probability distribution functions.

5.3 Combined Probability Distributions

Classical Inference: Probability density functions (PDFs) for observing ET and UT signal amplitudes from a specific flaw depth, *P(amplitude|depth)*, may be combined by classical

statistical inference to give a joint PDF. If the signal amplitude variables for ET and UT are independent (*i.e.*, if the two are uncorrelated for fixed depth), the joint probability density is the product of the individual PDFs. Such a joint probability density function, based on our fitted distributions for a 0.100-mm flaw depth, is shown in Figure 7. Note that this bivariate distribution, $P(amp_{ET},amp_{UT}|depth)$, indicates the response of inspection signals to a specific situation but may not be directly used in deciding flaw depth based on signal amplitudes. A transformation of the probabilities, needed to obtain $P(depth|amp_{ET},amp_{UT})$, may be accomplished by applying Bayes's Theorem to the monovariate distributions, or by using Bayes's Combination Rule to produce the joint distributions.

Bayesian Inference: PDFs, $P(amplitude|depth)$, obtained by the fits in Section 3.1 and transformed to $P(depth|amplitude)$, by Bayes's Theorem, assuming a uniform depth distribution, are plotted in Figure 8(a) for arbitrary ET and UT amplitudes of 100. We compare these to the joint distribution, $P(depth|amp_{ET},amp_{UT})$, obtained by Bayes's Combination Rule as formulated in Section 4.3 (solid curve in figure). The joint distribution is evidently narrower than those for the individual techniques, indicating graphically the gains possible by combining the two measurement techniques. Also shown in the figure is the joint cumulative probability function (divided by 10). If, instead of a uniform depth distribution, one assumes the empirical distribution,

$$P(depth) = 2.77x\exp(-17.3x),$$

from Section 3.1, the individual and joint probability functions of Figure 8(b) are obtained. These are more compact than the curves in panel (a) of the figure, since multiplication by the *a priori* likelihood enhances the probabilities in the 0.100 mm region. However, analysis by ROC curves (Figure 9) shows little improvement in performance, since such analysis relies on relative probability amplitudes, which are preserved by the transformation. A more fundamental reason to avoid use of the empirically obtained depth distribution is that the artificially generated flaw set may not be representative of accidentally occurring flaws, and may thus inappropriately bias the joint distribution function.

Dempster-Shafer Combination: Again, the PDFs are transformed to $P(depth|amplitude)$ by Bayes's Theorem, assuming a uniform depth distribution. Following the method of Gros *et al.*[6], we obtain a joint probability from Dempster's rule of combination for each depth interval, subject to the requirement that probabilities outside that depth interval represent no belief (as opposed to disbelief) about the probabilities within the interval. ROC analysis of the results gives the curve in Figure 9 labeled "Dempster-Shafer, non-belief". The same procedure, but with the assumption of conflict between depth intervals, would represent the "disbelief" scenario; however, since our probability distribution functions are normalized Gaussians, that joint probability distribution should be the same as the one obtained by Bayesian combination.

6. DISCUSSION

Our results demonstrate the significant improvement in inspection reliability achievable by combining the results of the two inspection techniques with simple OR logic. Figure 5 shows the

large gain in probability of rejection obtainable at fixed threshold. As an example, the lower-bound rejection probability at 95% confidence for a threshold-depth flaw is 44% for ET, 49% for UT, and 78% for the OR combination. Figure 6 demonstrates that this is also a gain relative to the false-alarm rate, and is applicable over a broad range of thresholds.

Additional gains were achieved by the implementation of better combination techniques. The arithmetic mean and the weighted mean of the ET and UT signal amplitudes give a somewhat improved ROC curve in Figure 6 relative to the curve for logical-OR combination. To include more detailed knowledge of the probability density functions for ET and UT, joint probability distributions must be computed.

There are many statistical approaches to the creation of joint probability distribution functions, including classical statistical inference, Bayesian inference, and the Dempster-Shafer method. These more complex combination techniques also give good results (the Dempster-Shafer results are plotted in Figure 6 for comparison), but offer little improvement over the mean and weighted mean methods.

The measured reliability of inspection outcomes increases with the amount of information incorporated in the decision process. Large gains are achieved for zeroth-order OR combination; further improvement is obtained by combining the first moments (amplitudes) of the probability functions by averaging and, to a lesser degree, the second moments (widths) in a weighted average. The higher moments accessible by more sophisticated combination techniques would be significant only for very asymmetric probability density functions and do not appear to improve reliability in this case.

7. SUMMARY

We have compared a series of progressively more complex techniques for combining two types of inspection data. These techniques include:

- the logical-OR combination of ET and UT rejection decisions,
- the mean (both simple and weighted) of ET and UT signal amplitudes, and
- the combination of the PDFs for each into a joint probability distribution function.

Significant gains in inspection reliability are demonstrated for the combination of ET and UT data from a large and realistic set of simulated flaws. While the reliability increases with the amount of information incorporated in the decision, this effect seems to saturate with the introduction of weighted averages of ET and UT results for the present data. More detailed data could potentially benefit from more sophisticated combination techniques.

8. ACKNOWLEDGEMENT

This work was supported in part by the CANDU Owners Group (COG).

9. REFERENCES

[1] Packman, P.F., S.J. Klima, R.L. Davies, J. Malpani, J. Moyzis, W. Walker, B.G.W. Yee, and D.P. Johnson, *Reliability of Flaw Detection by Nondestructive Inspection*, ASM Metals Handbook, 8th Ed., **11** (ASM, 1976).

[2] Heasler, P.G., S.R. Doctor, and T.T. Taylor, *Quantifying Inspection Performance Using Relative Operating Characteristic Curves*, 10th Int'l. Conf, on NDE in the Nuclear and Pressure Vessel Industries, M. J. Whittle *et al.*, eds. (ASM International, 1990) 481.

[3] Rummel, W.D., *Probability of Detection as a Quantitative Measure of Nondestructive Testing End-To-End Process Capabilities*, Materials Evaluation **56** (1998) 29.

[4] See, for example, S.L. Meyer, Data Analysis for Scientists and Engineers, (Wiley, Toronto, 1975).

[5] Hall, D.L., Mathematical Techniques in Multisensor Data Fusion, (Artech House, Boston, 1992) and references therein.

[6] Gros, X.E., P. Strachan, and D.W. Lowden, *Theory and Implementation of NDT Data Fusion*, Res. Nondestr. Eval. **6** (1995) 227.

[7] Dromigny, A. and Y.M. Zhu, *Improving the Dynamic Range of Real-Time X-Ray Imaging Systems via Bayesian Fusion,* J. Nondestr. Eval. **16** (1997)147.

[8] Mayo, W.R., *Reliability of Nondestructive Testing*, CSNDT Journal **12**, (1991)14.

[9] Liping, Liu, *A Theory of Gaussian Belief Functions,* International Journal of Approximate Reasoning, **14** (1996)95.

Table I. Depth (through-wall extent), ET amplitude, and UT amplitude for manufactured flaws of reference [8].

FLAW label	DEPTH (mm)	ET (scaled)	UT (dynamic)	FLAW label	DEPTH (mm)	ET (scaled)	UT (dynamic)	FLAW label	DEPTH (mm)	ET (scaled)	UT (dynamic)
D1	1.700	500	224	C8	0.150	175	178	OD9	0.076	50	83
C2	1.020	500	281	K2	0.150	65	100	A7	0.064	150	178
D3	0.572	500	178	N8	0.150	130	150	ID8	0.064	50	100
ID15	0.480	480	224	C12	0.140	275	141	M3	0.060	30	100
N9	0.460	350	150	D6	0.140	475	100	A9	0.051	150	58
OD16	0.410	400	178	ID22	0.140	100	224	B9	0.051	65	30
B0	0.406	500	178	OD6	0.140	80	100	ID2	0.051	40	100
OD17	0.380	410	224	D5	0.130	500	178	L2	0.051	65	60
D2	0.343	500	278	K3	0.130	100	150	N7	0.051	75	150
L9	0.300	165	150	B6	0.127	175	80	OD5	0.051	0	100
ID13	0.280	480	251	C9	0.127	100	200	A8	0.050	90	40
A4	0.267	425	100	OD3	0.127	90	200	B10	0.050	60	45
B5	0.254	300	141	OD7	0.127	80	126	B12	0.050	70	0
D4	0.254	400	158	D9	0.114	300	112	D11	0.050	125	100
K8	0.254	270	150	D12	0.114	175	126	ID12	0.050	60	55
A3	0.250	500	126	OD2	0.114	80	100	ID17	0.050	500	158
A6	0.230	500	100	A11	0.102	130	70	ID20	0.050	100	224
ID14	0.230	280	178	B3	0.102	325	100	L7	0.050	35	0
OD14	0.230	290	178	C7	0.102	300	112	N2	0.050	15	0
B2	0.229	500	126	D10	0.102	300	100	N3	0.050	15	30
C5	0.229	300	224	ID11	0.102	100	281	N6	0.050	50	80
C11	0.229	300	200	ID19	0.102	90	224	OD18	0.050	0	40
B4	0.216	325	141	K1	0.102	70	150	N4	0.040	0	0
B7	0.216	250	100	M6	0.102	50	150	B11	0.038	70	20
D8	0.216	500	200	OD10	0.102	80	90	OD15	0.038	70	68
OD11	0.216	160	200	C10	0.100	125	92	A12	0.030	60	0
A5	0.203	400	72	K5	0.100	55	80	ID4	0.030	0	30
D7	0.191	500	200	K6	0.100	75	86	ID7	0.030	0	25
A1	0.178	350	100	M5	0.100	0	100	ID9	0.030	0	43
B8	0.178	175	100	A10	0.089	150	95	ID16	0.030	280	158
C6	0.178	225	158	ID5	0.080	70	158	L3	0.030	0	20
K7	0.178	180	150	L6	0.080	60	90	OD4	0.030	0	0
L8	0.178	190	150	N5	0.080	50	150	OD13	0.030	110	60
M8	0.178	90	150	K4	0.076	25	34	ID21	0.025	0	0
OD23	0.178	130	100	OD1	0.076	80	112	M7	0.025	0	70
OD21	0.165	110	65	OD8	0.076	60	95	ID18	0.020	0	50

Table II. Logical-OR combination of inspection decisions

	UT: accept	UT: reject
ET: accept	accept	reject
ET: reject	reject	reject

Table III. ET, UT, and combined rejection rates as a function of depth range (±0.020 mm) for flaw database [8].

Depth (mm)	Total # of flaws	Rejection		
		ET	UT	Combined
0.50	4	4	4	4
0.46	1	1	1	1
0.42	2	2	2	2
0.38	1	1	1	1
0.34	1	1	1	1
0.30	2	2	2	2
0.26	5	5	5	5
0.22	11	11	10	11
0.18	9	8	8	8
0.14	13	10	12	13
0.10	20	14	13	19
0.06	26	6	8	10
0.02	13	2	1	2

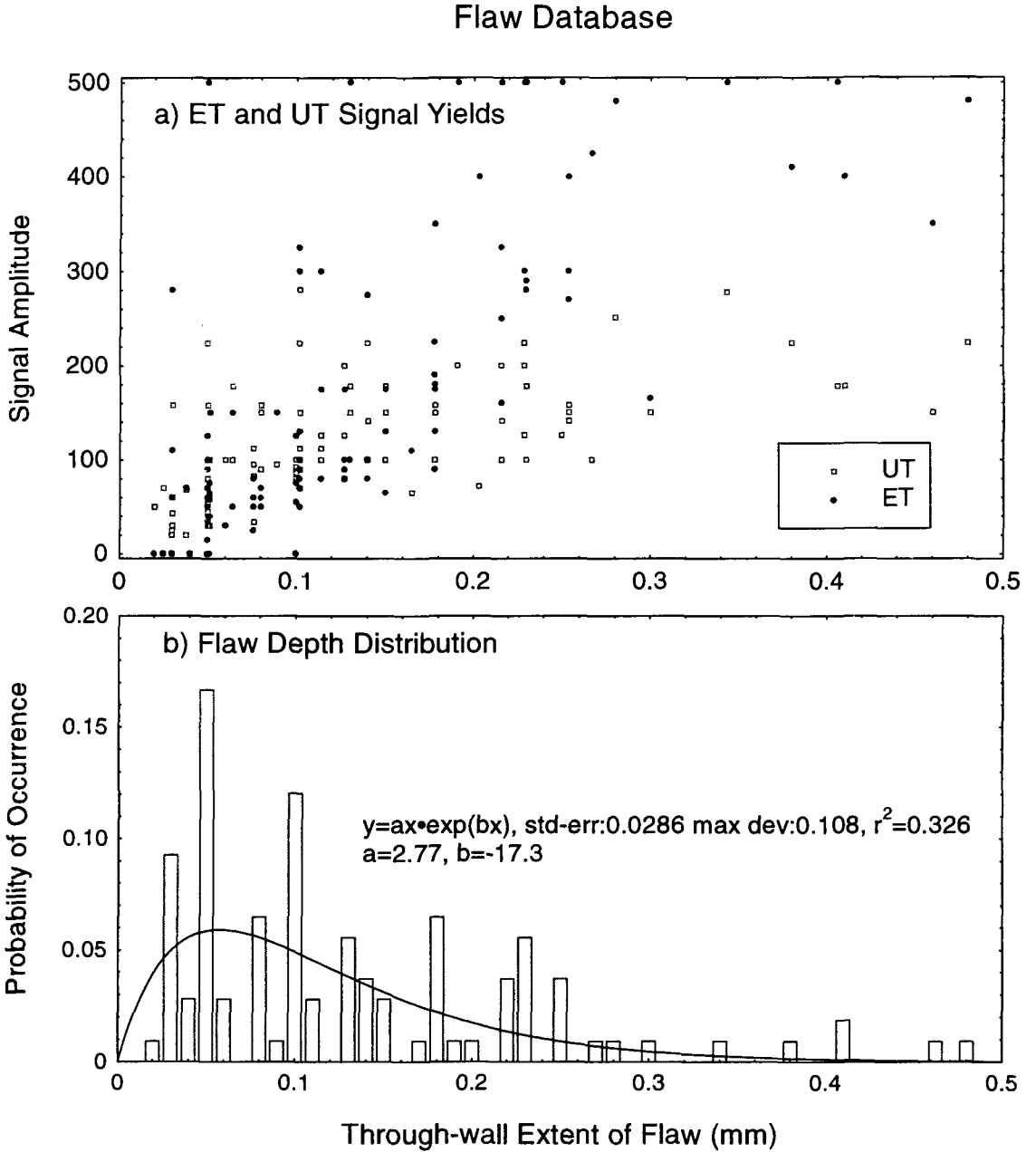

Figure 1. a) Measured ET and UT signal amplitudes for flaw depths up to 0.5 mm. b) Flaw depth distribution histogram, projected from top panel and normalized to unit total probability. Fitted curve is of the form $P = ax\ exp(bx)$, with $a = 2.77$, $b = -17.3$.

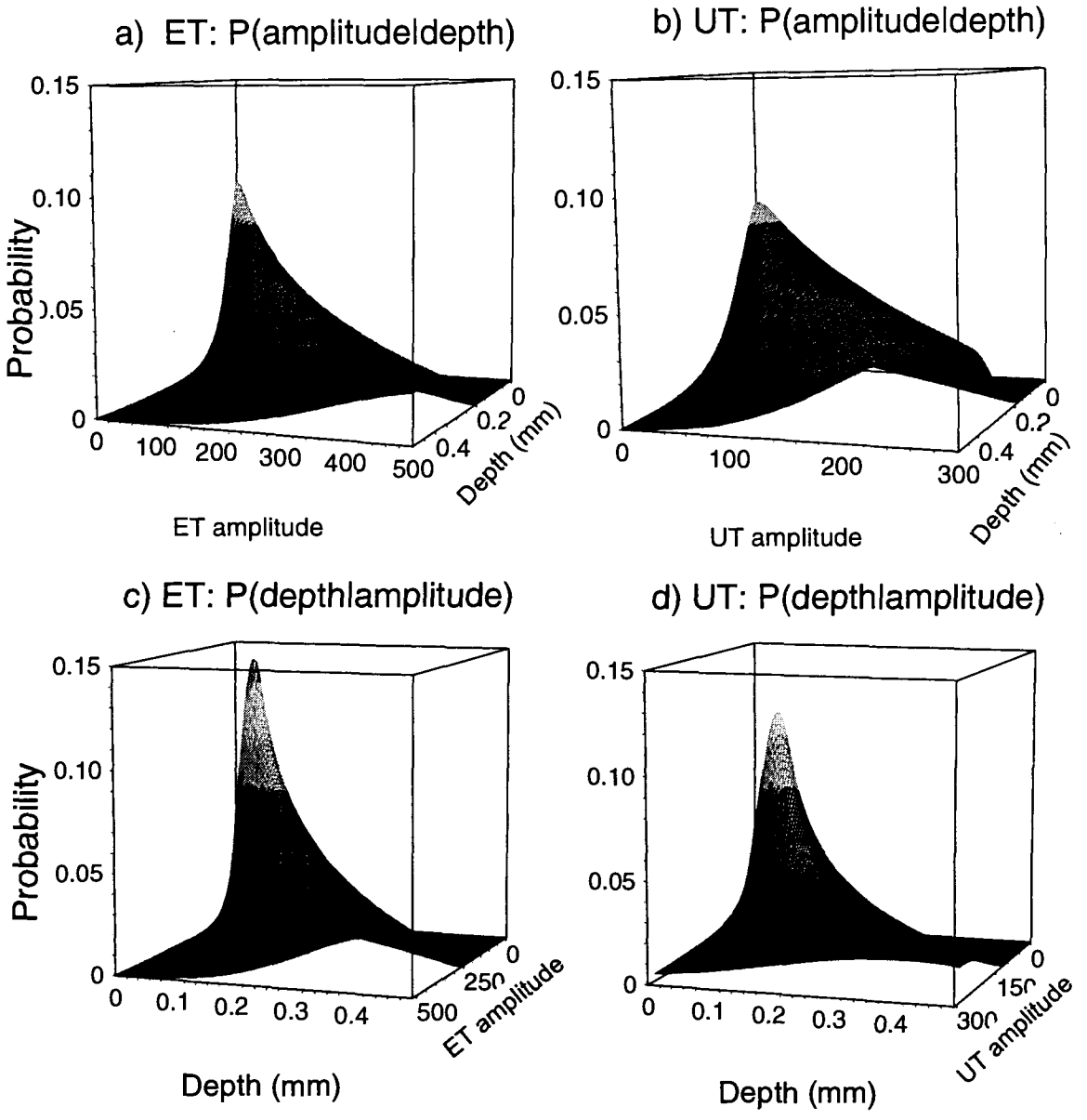

Figure 2. Gaussian fits to the point probabilities generated from Figure 1. a) Fit to probability as a function of ET signal amplitude, normalized to unity for each depth. b) Fit to probability as a function of UT signal amplitude, normalized to unity for each depth. c) Fit to probability as a function of depth, normalized to unity for each ET signal amplitude. d) Fit to probability as a function of depth, normalized to unity for each UT signal amplitude. See text for fit parameters.

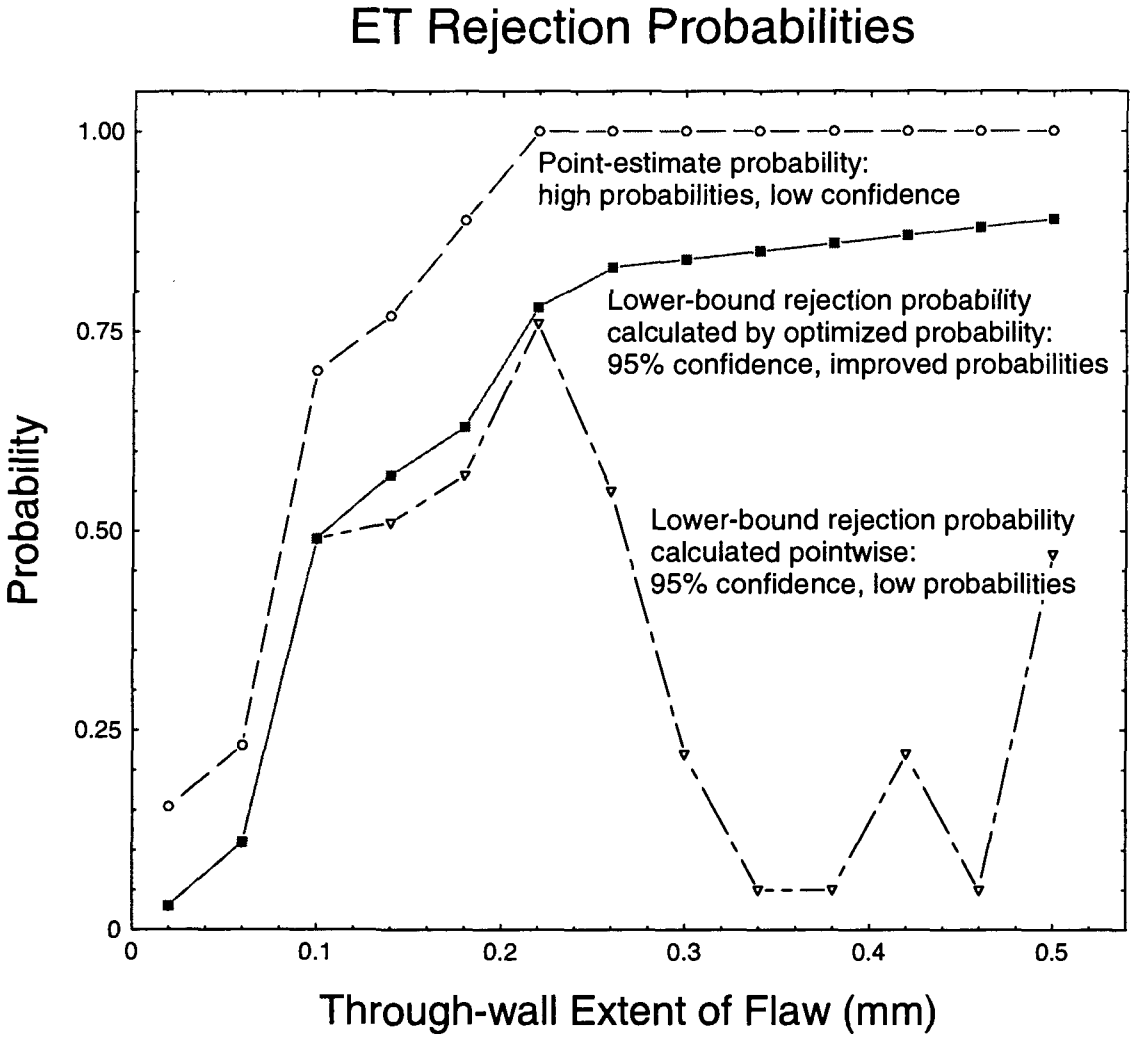

Figure 3. Examples of various estimates of flaw rejection probabilities as a function of flaw depth (through-wall extent).

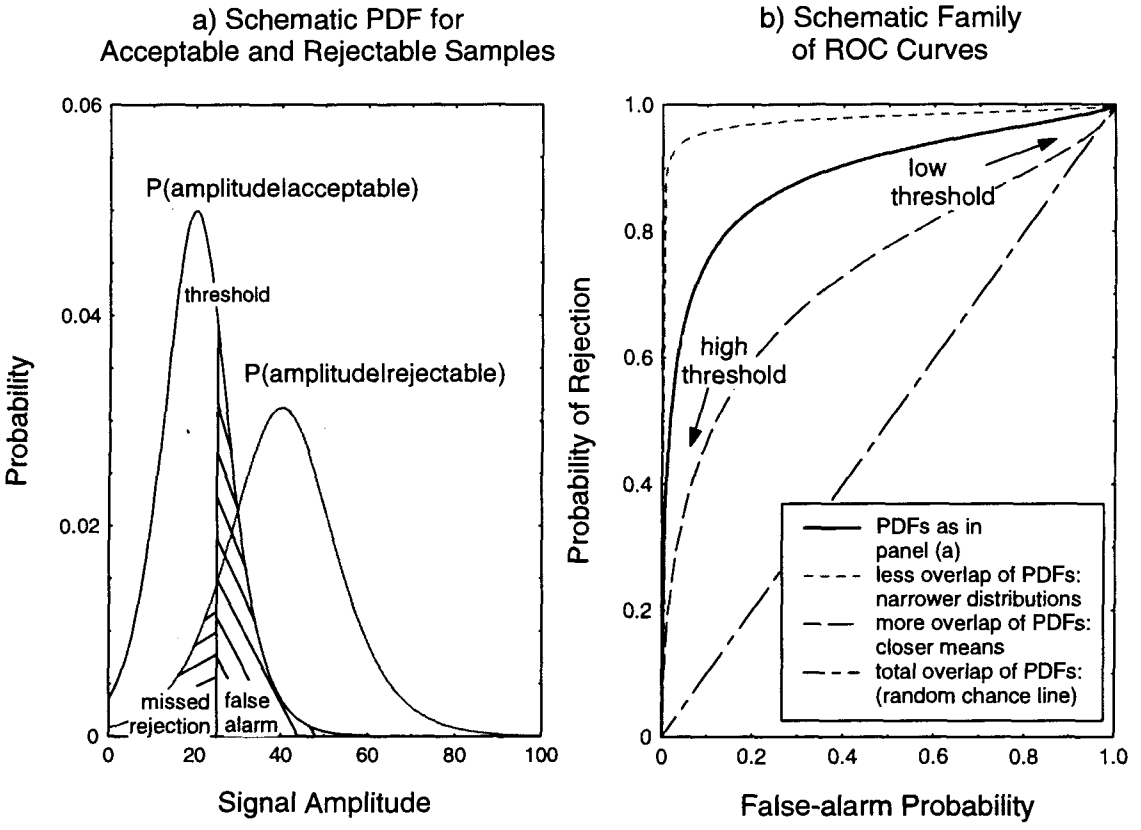

Figure 4. a) Schematic probability density functions generated from the logistic distribution to demonstrate the dependence of error rates on choice of threshold. b) Relative operating characteristic curve obtained from the distributions in the left-hand panel (solid line), curve for a pair of narrower distributions (short-dashed line), curve for the distributions moved closer together (long-dashed-line), and the random chance line (diagonal).

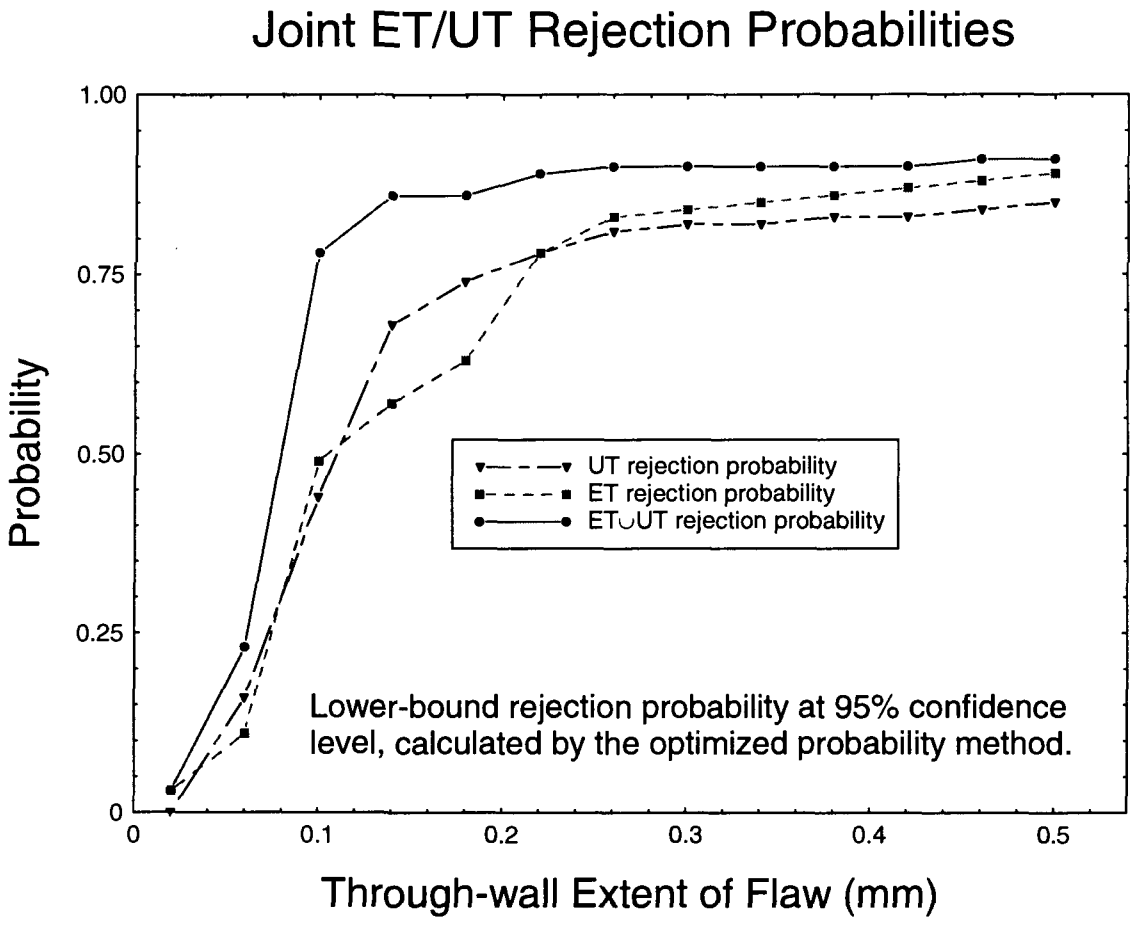

Figure 5. Lower-bound ET, UT, and combined rejection probabilities, calculated for 95% confidence level. Data from Mayo [8].

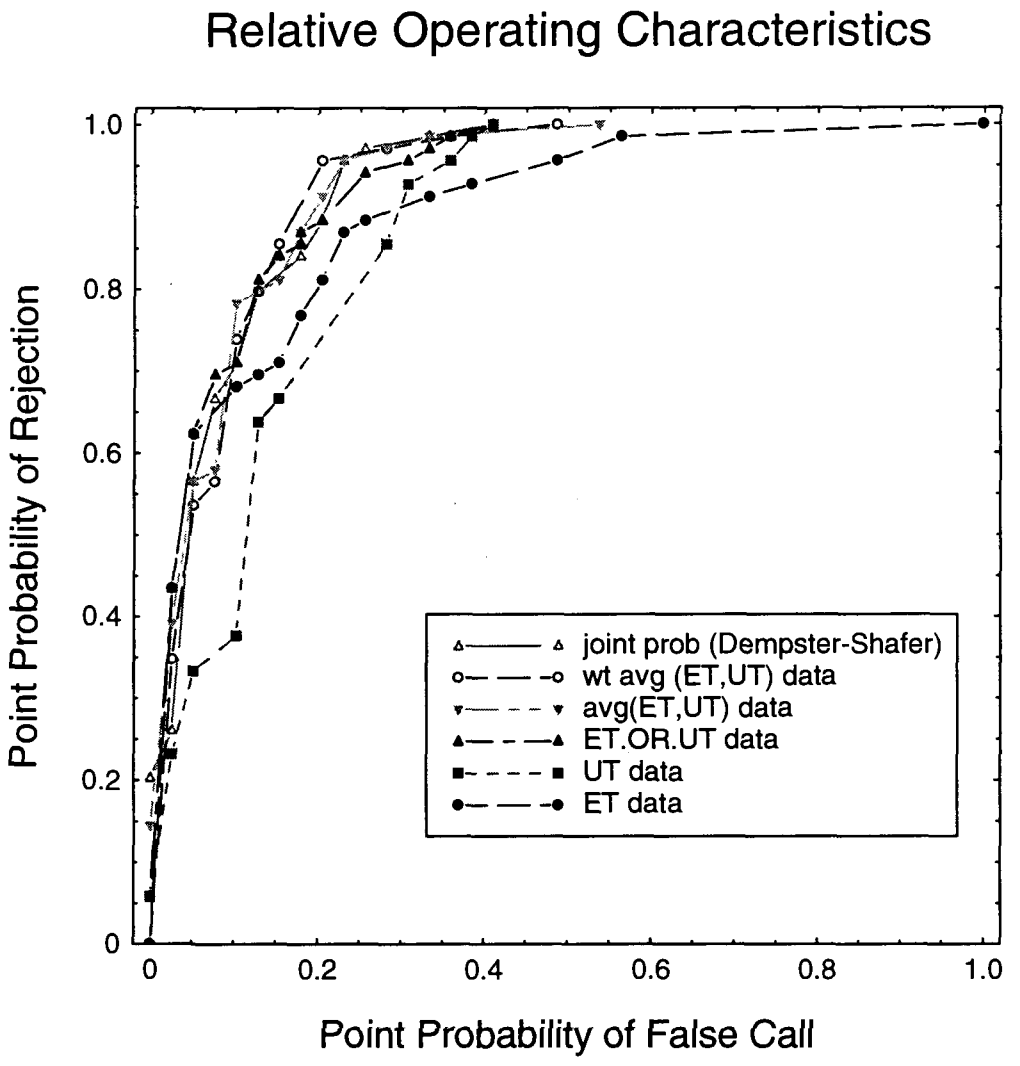

Figure 6. Relative operating characteristics for decisions based on ET and UT data and on their combination by various means. The upper right and lower left are regions of low and high threshold, respectively; merit may be judged by proximity to the upper left corner, representing "perfect" inspection data with a 1.00 probability of rejecting a defect and a 0.00 probability of making a false call.

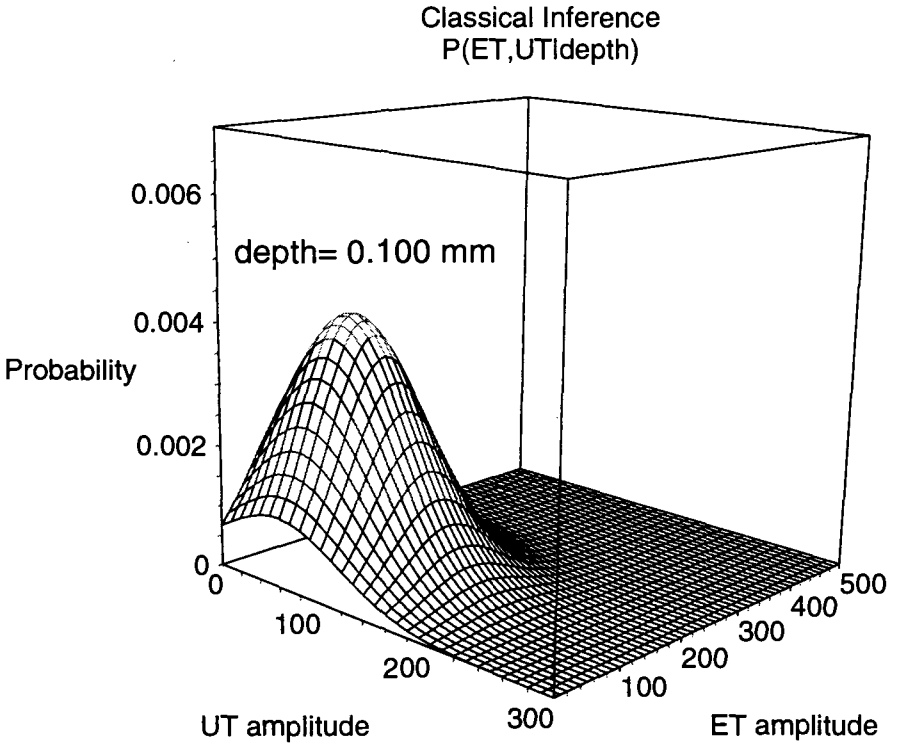

Figure 7. Joint probability density function, computed by classical inference assuming a correlation coefficient of zero, for obtaining various signal amplitudes from a flaw that is 0.100 mm in through-wall extent.

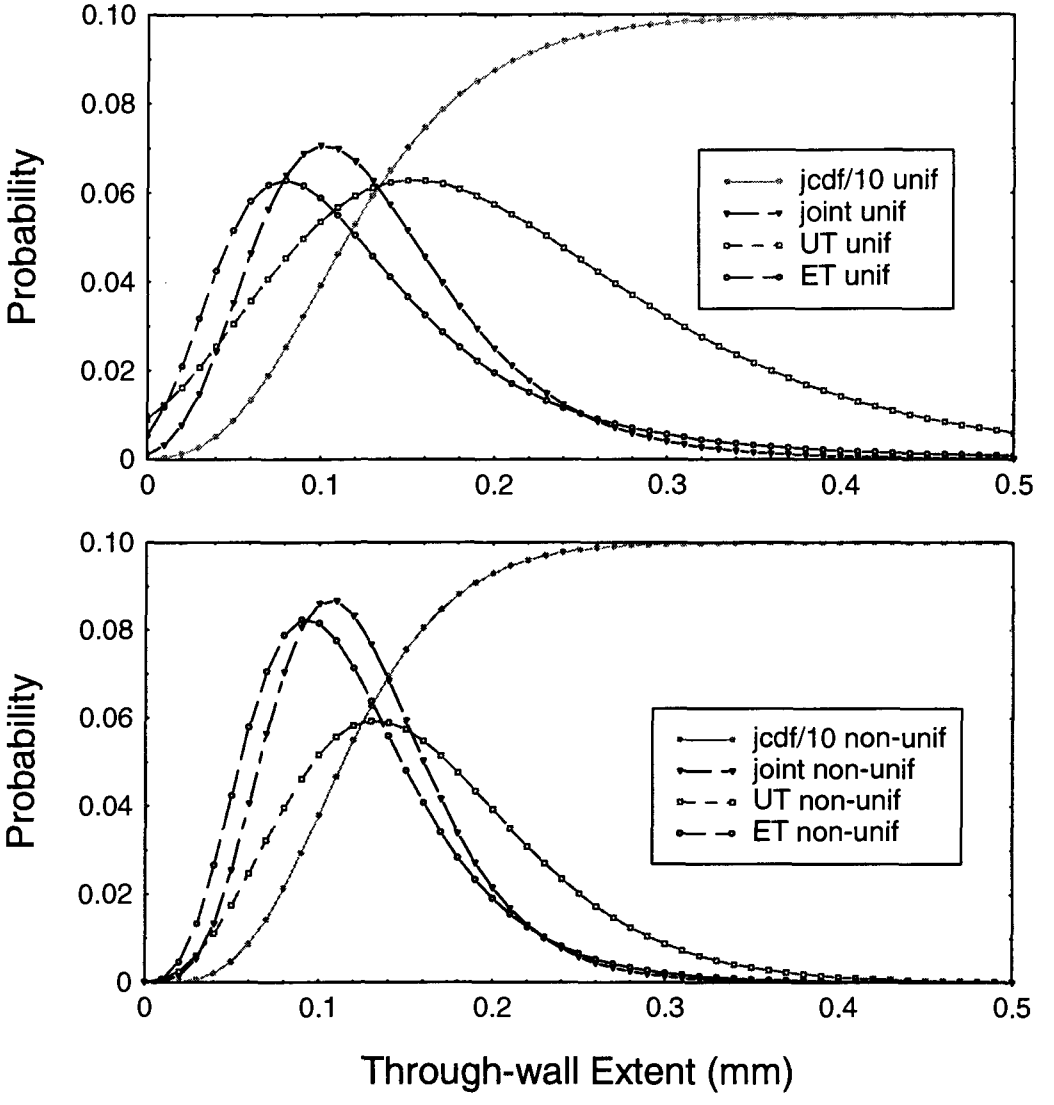

Figure 8. Individual and joint probability functions for ET and UT amplitudes of 100, obtained with transformation by Bayes's Theorem and Bayes's Combination Rules. a) Uniform depth distribution assumed. b) Non-uniform depth distribution taken from fit to Figure 1(b).

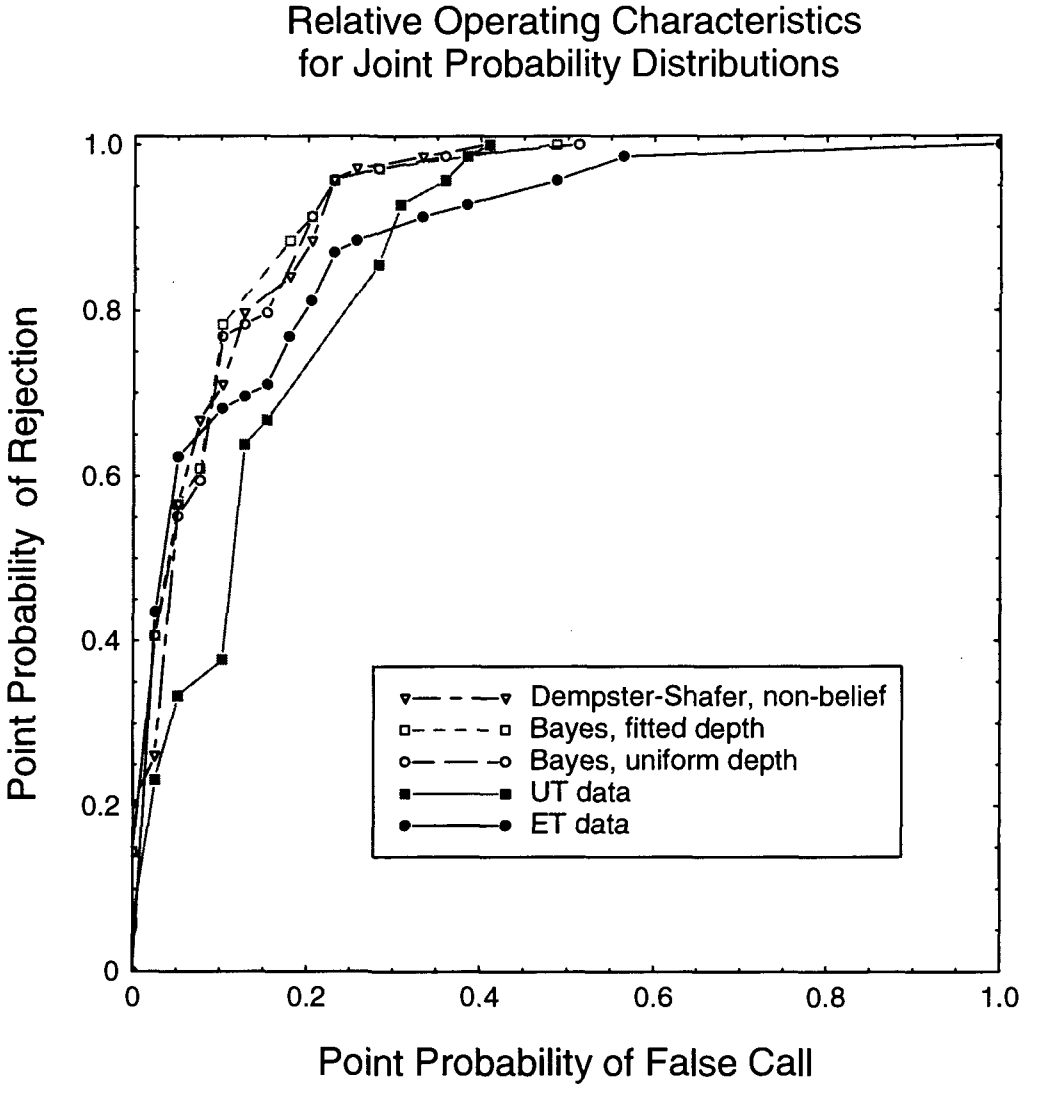

Figure 9. Relative operating oharacteristics for decisions based on ET and UT data and on their fusion by Bayes's Combination Rule and by the Dempster-Shafer method. The upper right and lower left are regions of low and high threshold, respectively; again, merit may be judged by proximity to the upper left corner.

ISSN 0067-0367

To identify individual documents in the series, we have assigned an AECL- number to each.

Please refer to the AECL- number when requesting additional copies of this document from:

Document Centre
AECL
Chalk River, Ontario
Canada K0J 1J0

Fax: (613) 584-8114 Tel.: (613) 584-3311
ext. 4623

Price: A

Pour identifier les rapports individuels faisant partie de cette series, nous avons affecté un numéro AECL-à chacun d'eux.

Veuillez indiquer le numéro AECL- lorsque vous demandez d'autres exemplaires de ce rapport au:

Service de Distribution des Documents Officiels
EACL
Chalk River (Ontario)
Canada K0J 1J0

Fax: (613) 584-8114 Tel.: (613) 584-3311
poste 4623

Prix: A