SKI Report 2007:04

Research

Link Between RI-ISI and Inspection Qualification: Relationship between Defect Detection Rate and Margin of Detection

Barrie Shepherd
Sophie Goujon
John Whittle (John Whittle & Associates)

January 2007

ISSN 1104-1374
ISRN SKI-R-07/04-SE

SKi

SKI perspective

Background

Quantitative risk-informed in-service inspection (RI-ISI) requires a quantitative measure of inspection effectiveness if the risk change associated with an inspection is to be determined. General methods for this are not available today and the work in this report address the problem of how the information generated in the qualification process used today can be used in RI-ISI models.

Purpose of the project

This project propos one way to extract risk change data from NDT systems formed for the qualification process of today. The most common method for in service inspection in nuclear power plants is ultrasonic and this work investigates how the defect response signal in relation to the noise level is related to the probability of detection.

Results

A relationship has been established between defect detection ratio and margin of detection based on blind trails. The results are based on a rather small amount of data produced by personnel from one inspection laboratory but give us indications of how signal to noise ratio is possible to relate to the probability of detection (POD) used in RI-ISI models.

The relationship need to be developed with more data to put the results on a more solid basis.

Project information

Responsible for the project at SKI has been Peter Merck.
SKI reference: 14.43 - 200443110

SKI Report 2007:04

Research

Link Between RI-ISI and Inspection Qualification: Relationship between Defect Detection Rate and Margin of Detection

Barrie Shepherd
Sophie Goujon
John Whittle (John Whittle & Associates)

January 2007

CONTENTS

SUMMARY

Quantitative risk-informed in-service inspection (RI-ISI) requires a quantitative measurement of inspection effectiveness if the risk change associated with an inspection is to be determined. Knowing the probability of detection (POD) as a function of defect depth (through wall dimension) would provide ideal information.

However the main in-service inspection method for nuclear plant is ultrasonics, for which defect detection capability depends on a wide variety of parameters besides defect depth, such as defect orientation, roughness, location, shape etc. In recognition of this the European approach to inspection qualification is generally based on some combination of technical justification, and practical trials on a relatively limited number of defects.

This inspection qualification process involves demonstrating that defects of concern will generate responses in excess of the specified recording level or noise, depending on the inspection. It is not currently designed to quantify the probability with which defects will be detected.

The work described in this report has been performed in order to help address the problem of how the information generated during inspection qualification can be used as an input for RI-ISI. The approach adopted has been to recognise that as the defect response increases above the recording or noise level, the probability of detecting defects is likely to increase.

The work therefore involved an investigation of the relationship between POD (strictly speaking defect detection rate) and margin of detection. It involved blind manual and automated ultrasonic trials on artificial defects in testplates designed to generate a range of signal responses. The detection rate for defects which provided signals at a particular level above noise or above a recording level was then measured.

A relationship between defect detection rate and margin of detection has been established based on these trials. In addition to establishing a stronger link between RI-ISI and inspection qualification, the results should provide useful information to support inspection qualification, even when the inspection programme is not based on RI-ISI

Recommendations are provided for further work to put the results on a more solid basis, and to confirm or otherwise the extent to which the relationship established is generic.

SAMMANFATTNING

Kvantitativt Risk Baserat Provningsurval för mekaniska komponenter (RI-ISI, från engelskans risk-informed in-service inspection) kräver en kvantitativ metod att mäta effektiviteten av provningen om man ska kunna fastställa hur risken förändras när en provning utförs. Att veta provningseffektiviteten (Engelskans Probability Of Detection, POD) som en funktion av defektdjupet skulle vara idealt.

Den vanligaste provningsmetoden i kärnkraftverk är ultraljud: Sannolikheten att finna en defekt beror på ett flertal olika parametrar såsom defektens djup, orientering, ytans jämnhet samt dess position och utformning, mm. Med detta i åtanke är det europeiska tillvägagångssättet för provningskvalificering vanligtvis baserat på en kombination av teknisk motivering och praktiska demonstrationer på ett fåtal typer av defekter.

Denna kvalificeringsprocess för provningssystem inkluderar därför oftast att demonstrera att de defekter som systemet kvalificeras för, genererar signaler över de specificerade upptagnings- och brusnivåerna beroende på systemets utformning. Kvalificeringsprocessen för tillfället är inte kopplad till att kvantitativt fastställa sannolikheten med vilken defekter är detekterbara.

Denna rapport beskriver det arbete som har gjorts för att adressera hur information som alstrats under provningskvalificeringar kan användas inom modeller för riskinformerat provningsurval. Den metod som valts tar i beaktande att sannolikheten att finna en given defekt ökar i samband med det att signalstyrkan överskrider upptagningsnivåerna eller brusnivåerna.

Arbetet har inneburit en utredning av förhållandet mellan POD-värden (till vilken grad defekten kan upptäckas) och erhållna marginaler av detekterbarheten. Detekterbarheten för defekter som genererade signaler av en given nivå över brus eller upptagningsnivån mättes därefter.

Sambandet mellan graden av detekterbarhet och erhållna marginaler av detekterbarheten fastställdes därmed under dessa försök. Försöken borde utöver att fastställa ett starkare samband mellan RI-ISI och provningssystemens kvalificeringar, också bidra med information for att stödja kvalificeringarna, även när provningsprogrammet inte är baserat på RI-ISI.

Förslag till vidare arbete är rekommenderat för att underbygga resultaten och för att fastställa eller bevisa, till vilken grad sambandet är allmänt gällande.

1. INTRODUCTION

This report describes the work performed and presents the results and conclusions from the project "Link between RI-ISI and Inspection Qualification: Relationship between POD and Signal to Noise Ratio" funded by the Swedish Nuclear Power Inspectorate (SKI).

Ultrasonic inspections are generally designed to detect those defects judged, on the basis of analysis or previous experience, to be possible in the component under test. Any inspection has a finite chance of failing to detect some of these defects. These failures may be as a consequence of systematic faults in the inspection which means that it is incapable of detecting some of the defects. Alternatively, random faults normally associated with human error may be responsible. For high integrity plant it is important to keep the chance of detection failure small. A commonly used measure in this context is the probability of detection.

POD = 1- probability of failing to detect defects.

Conventionally POD is expressed as a function of defect size since it is the size of the defect which is generally of concern from a structural integrity viewpoint. Knowing POD as a function of defect size can be used as an input to risk-informed in-service inspection (RI-ISI) programmes, in order to calculate risk change associated with an inspection.

However, many variables apart from defect size can influence the response signal from a defect and hence detection. These variables include defect position, tilt, skew, roughness and shape. In the case of austenitic stainless steel and inconel welds, there are additional variables due to the influences of welding process, procedure and position on ultrasonic properties.

For high integrity inspections it is common to use Inspection Qualification as a means of providing confidence in inspection capability, taking these various defect and component attributes into consideration. A European approach to this has been defined by ENIQ (European Network for Inspection and Qualification).

Qualification of an NDT procedure establishes whether the inspection is capable of meeting the requirements set for it. These include the detection of defects of specified type, position, orientation and size. Depending on the inspection, there may be additional targets relating to positional and sizing accuracy.

Qualification is based on a combination of methods depending on the particular inspection. These usually include a technical justification containing all the evidence supporting the inspection, often including modelling calculations for the amplitude of the target defects. Practical trials to show that the predicted performance is achieved in practice are also usual. For practical reasons these trials can generally only include a relatively small number of defects.

The output of a qualification is a demonstration that the defects in question generate responses in excess of the set recording level or noise depending on the inspection. The greater the margin of detection, the greater the confidence that the inspection will detect the defects in practice. If defects are detected independently by several probes, such redundancy adds further to the confidence in the capability of inspection.

However at present it is not possible to make quantitative statements from qualification about the probability with which defects will be detected. This is because it is generally impractical to generate sufficient data from practical trials to achieve this, bearing in mind all the parameters which influence detection capability. Even then the results would only be applicable to that specific component, that specific inspection procedure (which is often unique to the inspection vendor) and those specific defect types. This prevents qualification making an objective input into calculations of risk except in very limited cases.

The work described in this report has been performed in order to help address this problem of how the information which comes from inspection qualification can be used as an input to risk informed in-service inspection.

The approach adopted has been to recognise that although inspection qualification is not designed to generate POD data, it often involves determining whether defects above a given size (and within a specific range of orientations, shapes, locations etc.) will be detected with adequate margin. This margin may be expressed in terms of signal to noise ratio, or in terms of signal above report threshold.

It seems obvious that POD will fall as signal to noise ratios approach unity or as defect signals fall towards the threshold recording level set in the inspection procedure. Once again, the POD is influenced by more features of the data than just the response amplitude, (including e.g. proximity of geometric reflectors, experience of inspection personnel etc.) but it is clear that signal amplitude is one of the most significant.

However, there has previously been no information on how rapidly POD falls as signal levels fall. If this relationship can be established, then it may be possible to take the information produced during inspection qualification (margin of detection for defects within a specific range) and convert it to the information required by RI-ISI (POD for defects above a certain size)

This project therefore covered experimental work which was performed to establish the relationship between margin of detection and POD.

In addition to establishing a stronger link between risk informed in-service inspection and inspection qualification, the project results should provide useful input to inspection qualification even when not used in conjunction with RI-ISI. This is because inspection qualification often concludes that there will be high reliability of detection based on calculated or measured signal levels, but there has previously been no objective data to back this assumption.

2. OBJECTIVES

The objective of the project was to conduct a systematic investigation to establish the relationship for ultrasonics between margin of detection for a defect, and its probability of detection. It should be noted that due to the relatively small numbers of defects studied, the term probability is not correct in its strict statistical sense, and "defect detection rate" would be a more accurate term than "probability of detection". However for convenience the more common terms probability and POD are used throughout this report.

The work was based on a preliminary investigation to determine appropriate testpiece materials and defect sizes for the study, followed by a series of blind trials i.e. the UT inspectors did not know the locations or numbers of the defects present.

The intention for the main blind trials was to manufacture testpieces containing a large number of artificial defects which would be inspected by both manual UT and automated UT. It was originally proposed that 10 identical defects would be produced for each signal to noise level to be investigated, in the range 0dB to 18dB. However the preliminary investigation identified that a slightly different approach to defect design would be required in order to achieve sufficiently low signal to noise ratios. This is discussed in more detail below.

3. DESCRIPTION OF TRIALS

3.1 Preliminary Investigation

The aim of the preliminary investigation was to establish the defect types and sizes, and the ultrasonic beam characteristics, which would produce signal amplitudes and signal to noise ratios in appropriate ranges. Note that the nature of the defects themselves and the testpiece material are unimportant since their purpose is simply to be the source of signals. It had been anticipated that small spark machined slits in ferritic carbon steel plate, in combination with particular probe types would be suitable.

Initial exploratory trials were performed on a carbon steel testplate (testplate A) which was 22mm thick and contained three slots at the inner surface which were 1mm, 3mm and 5mm deep. Inspections were performed using 45º and 70º 4MHz, 10mm diameter single crystal probes. The ratios for the signal to noise levels were in the range 17dB to 54dB which were much higher than required, suggesting that a ferritic testplate was unlikely to be suitable, even if thickness was increased (within a practical limit).

The investigation was therefore extended to testplates made from ultrasonically noisier material, since it might be easier to increase the noise level than reduce the signal amplitude. An alternative approach would have been to generate the noise electronically but since this would have added complications to the instrumentation used it was decided to keep this as a final option to fall back on if necessary.

Three further testpieces were therefore manufactured which were designed to produce higher noise levels. These were:

- Testpiece B: Manufactured stainless steel plate 316 type, 22mm thick which contained three slots at the inner surface. The slots were: 2mm, 1mm and 0.5mm deep.

- Testpiece C: Manufactured cast iron disk 25mm thick 105mm diameter. The disk contained three slots at the bottom surface 2mm, 1mm and 0.5mm deep and one 4mm deep slot at the upper surface.

- Testpiece D: Manufactured ferritic plate with 6mm stainless steel cladding, total thickness 26mm, containing three circular saw cut slots at the surface. The slots were 0.5mm, 1mm and 2mm deep, with lengths 11mm, 16mm and 23mm respectively

The ultrasonic probes used in these trials were as follows (not all probes were used for all testpieces):

Shear wave probes:

38°, 4MHz, Ø10mm single crystal probe
45°, 4MHz, Ø10mm single crystal probe
70°, 4MHz, Ø10mm single crystal probe
60°, 2MHz, Ø10mm single crystal probe
70°, 2MHz, Ø10mm single crystal probe
45°, 4MHz, Ø20mm single crystal probe
45°, 4MHz, Ø6mm sub-miniature single crystal probe

Angle compression wave probes:

45°, 4MHz, 10x10mm single crystal probe
60°, 4MHz, 10x10mm single crystal probe
70°, 4MHz, 10x10mm single crystal probe

The stainless steel and the cast iron materials appeared more appropriate for the S/N range of interest. Cast iron was considered the best for the lowest S/N ratios.

It should be noted that the material used does not need to be representative of that used in nuclear plant - it is the inherent relationship between S/N and POD which is required.

Since cast iron seemed to be the most appropriate choice for the main trials, two iron castings were ordered to provide further confidence in the design of the defects and testpieces before starting the main phase of the project. These were both 680mm diameter x 55mm thick. One was spheroidal graphite (SG) composition, the other was flake iron. Ultrasonic testing on them confirmed that the noise levels in the SG casting appeared satisfactory, while the noise /

attenuation in the flake iron casting were higher (as expected) – in fact so high that it was difficult even to detect a corner reflection from the edge of the testplate.

Nine slits of various sizes were machined into the SG cast iron testiece which was inspected by 45° and 60°, 4Mhz, 10mm diameter single crystal shear probes. The signal to noise ratios were around the levels required.

3.2 Organisation of Blind Trials

Four rectangular spheroidal graphite cast iron plates were ordered, each 1m long, 500mm wide and 55mm thick. These were referenced TP1 to TP4.

Spark eroded slits were machined in TP1, TP2 and TP3. The fourth testpiece TP4 was held in reserve. The slits were machined in batches, with the first batch produced in TP1. After UT checks confirmed that the signal to noise levels were suitable, a second batch was machined into TP1. The third and fourth batches were machined into TP2 and TP3 respectively.

Tables 1 – 3 provide details of the sizes and locations of the spark machined slits in these three testpieces, and Figures 1 – 3 present drawings of the testpieces. The slits were randomly located so that UT operators could not deduce where defects were likely to be, apart from ensuring that no two slits were so close together that it would be difficult to know which had been detected, and also none was closer to the edge of the testplate than 60mm (to allow sufficient scan distance and to avoid masking of the defect signal by the edge of the testplate).

All slits were 10mm long, corresponding to the nominal width of the UT beam. If any of the slits had been significantly longer than this, the probability of detecting it would increase since the operator would have more than one opportunity to detect it as the probe scanned back and forwards along its length. It would therefore become difficult to relate POD directly to signal to noise (or signal above threshold).

Slit depth was one of the following:

- 0.5mm (A series)
- 1mm (B series)
- 2mm (C series)
- 3mm (D series)
- 5mm (E series)

The first testpiece to be inspected blind was TP1 and it was inspected by manual techniques. All UT operators were Doosan Babcock staff qualified to PCN level 2 and there were seven in total. They were provided with the same information and instructions. The background to the project was explained to them and the importance of ensuring that the trials were blind was stressed. It was emphasised that there was no interest in checking the performance of an individual operator, only in the combined results over all operators. They should therefore not be tempted to look under the block or to discuss results with each other. (In fact the testblocks would have been too heavy for them to lift anyway, and it was clear from their results afterwards that there had been no conferring between operators.)

They were told that the defects were machined slits at the inner surface and that they were oriented parallel to the minor edge of the plate. Scanning was to be performed initially towards the top of the plate only, using a specific 45°, 4MHz, Ø10mm single crystal probe, and a specific flaw detector.

Operators were instructed to report the location of every indication which they considered to be a slit, using x and y coordinates with respect to a specific corner of the testpiece. No reporting threshold was specified so that the results could retrospectively be analysed in terms of detection as a function of any given report threshold. For every defect reported, the signal amplitude was recorded by measuring the flaw detector gain required to set the signal to 80% full screen height, and the noise was recorded by measuring the gain required to set the noise adjacent to the defect location to 80% full screen height.

It had initially been intended that the effects of combining the results from diverse or redundant techniques would be investigated, by combining the results from scans in opposite directions, and from scans using different probes. However as described in Section 4.1 an alternative approach was adopted to investigate this, since it soon became apparent that scans in different directions could be subject to significantly different noise levels so that they would not be "equivalent". Also the use of an additional probe to the 45º probe referenced above provided no useful additional information (e.g. with a 60º 4Mhz probe too few defects were detected for results to be useful, with a 45º 2MHz probe too many defects were detected for results to be useful).

Automated blind trials were performed using the same probe (mounted in a specially manufactured probe holder) but using a Micropulse 4 flaw detector, an AWS-6 scanner, a motor controller, and MIPS/GUIDE control and display software. A broadly similar approach to the manual trials was adopted except that scans were performed in both directions. The results were independently analysed by three different UT data interpretation engineers. There was no need to repeat the physical scan for each of these engineers since any difference would have been attributable to minor changes in set-up.

A lot of useful data was generated from TP1. The defects in TP2 and TP3 had been manufactured partly as a contingency in case the trends which emerged from TP1 indicated that a wider population of defects was required – slit machining can be slow so it was decided to make sure these testpieces were available without delay.

In fact the noise levels in both TP2 and TP3 appeared to be higher than in TP1 to the extent that very few of the defects were detectable using the 4MHz probe referenced above (and too many were detected with high S/N ratio at lower frequency). Some limited trials were still performed on these testpieces although the results are not included in this report since population sizes for detected defects were rather small.

4. RESULTS AND ANALYSIS

4.1 Manual UT Signal to Noise relation with "POD"

Each operator's results were analysed by marking the locations of the indications they had reported on drawings showing the actual locations of the slits. No specific criteria were defined regarding how close to an actual defect a reported indication had to be to qualify as a "hit". However there were no instances where there was significant doubt related to location, since indications either coincided almost exactly (within a few mm) of a slit, or consistently plotted (for a given operator) a maximum of 10mm in the same direction with respect to actual slit locations, or were clearly false calls (which were never closer than around 30mm to slits). False calls could be caused by local peaks in the noise or possibly by pitting of the far surface of the casting.

Table 4 provides an overview of the manual UT results from the 33 defects in TP1. The left hand column lists the defect number. The next seven columns list the signal to noise measurements for each of these defects (if detected) for each of the seven manual UT operators M1 to M7. Where no S/N value is given, the defect was not reported.

Initially it was intended that an operator with prior knowledge of the defects would measure the signal and noise levels and these would then act as the "actual" or benchmark values. However if a defect is detected during a blind trial, then any signal and noise measurements made by that operator once the defect has been found would be just as valid. The signal to noise ratio for each defect has therefore been taken as the average calculated from each operator's results. This also has the advantage of minimising the subjectivity which can be associated with measuring noise using manual techniques.

The column headed "S/N mean A" lists the mean signal to noise ratio for each defect across all operators where S/N is taken as 0dB when not detected. However it would be very misleading to relate this mean to POD since each "miss" would artificially lower the mean S/N – e.g. a defect can have a S/N ratio of 12dB but still be missed by an individual operator. The miss does not mean that the S/N is any lower.

The column headed S/N mean B therefore lists the mean S/N for each defect, across all operators who detected that defect. The final two columns list the number of hits and the number of misses for each defect.

For each individual defect, seven operators scanned over it which is not a statistically significant population. The results for individual defects have therefore been batched according to the range that S/N meanB is in. The results are presented in Table 5 which lists the S/N meanB value, the total number of hits (by manual UT) for defects in that range, the total number of misses for defects in that range, and the "POD". Note that as previously noted "defect detection rate" is a more technically correct term in the context of this project, since the populations involved are still too low for these values to be considered as probabilities in a strict statistical sense. There is clearly a correlation between S/N and POD.

It had initially been intended to investigate how POD increases as additional independent scans are combined, based on similar scans in opposite directions, and using different combinations of probe. However the preliminary trials described in Section 3.1 had lead to the conclusion that the 45º 4 MHz probe was the most suited for the blind trials, with other probes resulting in too many or two few defects being detected for meaningful analysis. An alternative approach was therefore adopted based on combining scan results from different operators. An added advantage of this is that there was complete independence between the results – if the same operator had employed two different scans, the second scan could not be considered truly blind.

The scan results of two operators, M2 and M3 were combined. Table 6a lists whether these combined scans resulted in a hit or a miss for each defect. The results from another two operators, M5 and M6 were combined in the same way. Table 6b presents the total number of hits, misses, and "POD" for the combined scans, as a function of mean S/N (in batches as previously). Although the populations are small there is still a very clear trend of increased POD compared to the single scan PODs in Table 5.

4.2 Manual UT report threshold relation with "POR"

It is also useful to consider the relationship between how high a defect signal is above a reporting threshold, and the probability that it will be reported. An operator may notice a signal and recognise that it is from a defect, so that the defect has been detected in a literal sense, but there is unlikely to be any record of this if the operator considers it to be below the report threshold. The term "Probability of Reporting" (POR) is therefore used here to distinguish from POD. ("Report frequency" would be a more technically correct term than "probability of reporting").

A report threshold was not specified during the trials so that the influence of reporting threshold could be investigated retrospectively, without data from any of the defects being lost because the operator had noticed the defect signal but not recorded it. Clearly applying a report threshold will tend to filter out some of the defects, starting with the lowest amplitude ones.

The POR values will therefore decrease compared to the POD values in Table 5, with increasing report threshold (for no report threshold POR = POD). On the basis that noise is fairly uniform across the plate then as the report threshold increases, POR will reduce first for those defects with the lowest S/N ratio, e.g. in the band 5.0 – 7.4, then for defects in the band 7.5 – 9.9 and so on.

Even for defects which provide signals which are well above the noise level and are noticed by the operator, the probability that all defects above the report threshold are reported is expected to be less than 100%. This is because there will be variations in the report threshold which is established by operators even when using the same inspection procedure, equipment and calibration set-up. There will also be variations in the measured amplitude of defects. One operator may therefore assess a defect as being above the report threshold while another operator might assess it as being below. These discrepancies are obviously likely to be more pronounced for defects which produce signals close to the report threshold (as opposed to well above or well below).

This has been investigated by setting an arbitrary* report threshold retrospectively. Table 7 lists the amplitude of each defect as recorded by each of the operators (manual operator 2 has been excluded since he used a different flaw detector). In each case the gain (flaw detector amplification) required to set the defect signal to 80% full screen height has been recorded. Note that this means that the lower the number, the higher the signal. The spread in values appears surprisingly high, although this may be attributable to local variations in noise resulting in variations in attenuation. The report threshold was selected to be 67dB.

The column headed "mean" lists the mean amplitude for each defect and the number of "hits" (amplitude above report threshold) recorded. A "miss" corresponds to the recorded amplitude either being below the report threshold, or the defect not being detected by the operator. The number of operators (six) is too low to be statistically significant so the results have been batched into 2.5dB bands and presented in Table 8.

In Table 8, the amplitude bands are listed as absolute values, and also as dB above or below the 67dB report threshold. The number of hits and misses within each band are listed, and these have been converted to POR values.

The POR value increases with increasing amplitude, as expected. It is also clear that a significant proportion of defects below the threshold would still be reported, and that some of the defects above the report threshold would not be reported (even though the operator had detected them).

4.3 Automated UT Signal to Noise relation with "POD"

The automated UT data interpretation engineers were able to view the UT data as a colour coded C-scan (plan view) of the testblock, and dynamically adjust the record threshold until the defect images appeared clear with noise being below the threshold. A typical example of one of the C-scans is presented in Figure 4.

Analysis of the automated UT results was performed in a similar manner to that described in section 4.1 for the manual results, i.e. by marking the reported locations of defects on drawings showing the actual locations.

Table 9a presents the results. The mean signal to noise figures for the defects are those calculated from the manual results, to maintain consistency when comparing the manual and automated results. A hit or a miss is represented by 1 or 0 respectively. There is a greater degree of consistency between the different sets of automated UT results than for the manual. This is due to all data interpretation engineers analysing the same data, and the reduced subjectivity

* Not entirely arbitrary since it was selected to be at around the midrange of the amplitudes for the purpose of illustration.

involved in interpreting automated UT data (It can be seen from Figure 4 that interpretation is relatively straightforward).

Table 9b presents the results batched into 2.5dB ranges, together with POD. Even though the sample sizes are small, there appears to be a clear trend of increased POD compared to the manual UT results. The POD for S/N in the range 7.5 to 9.9 is 75% which is lower than for the 5.0 to 7.4 range (89%). Although this may be partly attributable to the low population of defects in the 5.0 to 7.4 range, it should be noted that none of the automated UT interpretation engineers detected defect 4C and it was only detected by one out of seven manual UT operators. If this indication reported by the manual UT operator is a false call which happens to coincide with the location of 4C then the actual signal to noise from 4C is probably lower than the 7.5 to 9.9 range, in which case the corrected POD would be based on 18 hits and only 3 misses, i.e. 86%

There was little point in combining the results from different data interpretation engineers in the way that was done for manual UT, because of the high degree of consistency between the analyses.

5. DISCUSSION

A lot of effort was devoted to establishing a combination of material, defect type and ultrasonic probe characteristics which provided signal to noise ratios in the range of interest. One of the main aims of the project was to establish a relationship between signal to noise ratio and probability of detection. The results suggest that this has been achieved.

The relationship between signal to noise ratio and POD is expected to be generic in the sense that it should not matter what types of artificial or real defect are the sources of the signals, or what the material or probe type is. However caution should still be exercised when applying the results of this project for the following reasons:

- The blind trials were performed under workshop conditions and the operators knew that the testblocks contained defects. Both factors are expected to increase POD compared to inspections on real plant at site.

- The operators were aware that the defects were at the inner surface whereas if the defects could have been situated anywhere throughout the depth of the testblock this might have made the inspection more challenging (although service induced defects are generally surface breaking).

- Although as stated above the material from which the testblocks were made should not be particularly relevant, the noise pattern within the blocks is likely to have been more representative of austenitic welds and castings than ferritic material. However it is usually in such materials that signal to noise ratio is the limiting factor regarding defect detection – in ferritic material it is more likely to be the margin by which the defect signal exceeds the report threshold.

- Although signal to noise ratio is an important aspect when considering POD it is not the only aspect which affects flaw detectability. The proximity of geometric discontinuities and other sources of signal (machined counterbore, weld cap, weld root, etc) can result in defect signals being masked or overlooked. A manual operator may miss a defect because of a gap in the scan pattern, or lack of concentration. The inspection equipment may not be reliable, or the inspection procedure may have inherent weaknesses (inspection qualification should identify these shortcomings). Pattern recognition can aid detection, and the longer a defect is, the more often the probe is likely to scan over it, effectively increasing the number of attempts at detection.

- It is important to recognise that although a lot of data were gathered, the results are based on relatively low sample sizes. In particular the POD of 100% for S/N values above 15dB presented in Table 5 should not be taken literally since it is only based on seven results. Averaging signal to noise ratios and assigning them into batches was considered to be a sensible way of analysing the results, but there are associated risks that incongruous results are masked and that the POD values calculated are sensitive to the ranges selected for the batches.

However regardless of the above cautionary statements, the results of the analysis reveal clear and consistent trends which provide confidence in the approach adopted.

Establishing the relationship between signal to noise ratio (or margin above report threshold) should help to translate inspection qualification results into the type of information which can be used by quantitative RI-ISI.

As an example, the RI-ISI analysis might be based on the assumption that defects above 15% wall thickness will be detected with a POD of at least 80%.

The role of the inspection qualification body would then be to make a judgement on the detectability of defects which exceeded this depth, and which were within certain other ranges determined by metallurgical considerations (covering tilt range, skew range, roughness, shape, location etc.)

Based on some combination of (possibly very limited) practical trials, theoretical modelling, physical reasoning and previous experience, the qualification body might conclude that all such defects should be detected with signal at least 14dB above noise.

Using the relationship suggested by Table 5 (in the case of manual UT), there would then be supporting evidence that the POD for such defects is around 86%, and the inspection system should therefore achieve the target POD required by the RI-ISI analysis.

Note that this simple example is for illustrative purposes only, and the other factors which can affect POD (not just signal to noise) would need to be kept in mind.

Combining independent scan results increased calculated POD as expected, and automated UT also increased POD. In the latter case the fact that automated inspection allows defects to be identified through pattern recognition (see Figure 4) is a major advantage compared to manual UT. With automated UT the concentration of signals at, or even below, the amplitude of the randomly distributed noise can facilitate detection. However a manual UT operator may not recognise that a signal is consistently rising and falling at the same position as the probe is scanned back and forwards along the length of a defect.

The results also demonstrate the importance of recognising that there will be a spread of results when different operators measure the amplitude of a defect signal. Although local variations of noise and attenuation may have exaggerated the spread in the current project, the results still highlight the importance of ensuring that when defining a report threshold, there is enough contingency to allow for this spread.

As stated in the introduction, the work performed is relevant to inspection qualification (and also inspection procedure design) even if RI-ISI is not adopted as the basis for the inspection programme.

6. CONCLUSIONS AND RECOMMENDATIONS

Blind trials have been performed on a range of artificial defects designed to provide a range of signal to noise ratios. A relationship has been established between margin of detectability (signal to noise ratio or signal above threshold) and POD (strictly speaking defect detection rate), although caution should be exercised when applying the results due to the relatively small populations of data corresponding to some of the POD values, and the other considerations discussed in Section 5.

The results can be used to help translate the results from inspection qualification into the type of information required by quantitative RI-ISI. They should also be useful even if RI-ISI is not being adopted, since they can be used both during inspection procedure design and inspection qualification by relating margin of detection to inspection performance (POD).

Even though the defect population sizes were relatively small, there appears to be a clear trend of POD increasing when scans are combined (as expected) or when performing automated inspection instead of manual inspection.

For example the proportion of defects detected when signal to noise ratio was in the range 10.0 – 12.4 dB was:

66% for single manual scans

94% for combined manual scans

100% for automated scans

Signal to noise ratio is often the limiting factor in ultrasonic inspection of austenitic welds. If this ratio is not sufficient to achieve a high enough POD then the inspection procedure should be revised (different probes, additional scans, automated scanning etc.)

For other components such as ferritic welds signal to noise ratios are generally much higher so that there is scope for increasing POD by reducing the report threshold.

When defining report thresholds, it is important to include sufficient contingency to allow for the spread which can occur when measuring signal amplitude. The results of this project have highlighted this point.

The relatively low defect populations studied, together with the other issues discussed in Section 5 provide a strong incentive to perform further work, in order to confirm or otherwise the validity of the results.

This work could include:

- Further trials using similar material, probe and defects to those studied in this project in order to increase the statistical significance of the results

- Further trials using different material, probes and defects to investigate the extent to which the relationship between margin of detection and POD is generic, and its sensitivity to other factors.

Slit	Depth (mm)	Length (mm)	Location slit centre(x, y)mm
1A	0.5	10	210, 800
2A	0.5	10	140, 480
3A	0.5	10	340, 420
4A	0.5	10	120, 220
5A	0.5	10	400, 160
6A	0.5	10	200, 940
7A	0.5	10	90, 840
8A	0.5	10	340, 740
9A	0.5	10	200, 600
10A	0.5	10	300, 580
11A	0.5	10	100, 460
12A	0.5	10	400, 340
13A	0.5	10	260, 200
14A	0.5	10	310, 140
1B	1.0	10	380, 880
2B	1.0	10	400, 660
3B	1.0	10	80, 560
4B	1.0	10	180, 360
5B	1.0	10	180, 160
6B	1.0	10	140, 860
7B	1.0	10	240, 680
8B	1.0	10	420, 500
9B	1.0	10	90, 320
10B	1.0	10	300,300
11B	1.0	10	80, 140
12B	1.0	10	100, 60
1C	2.0	10	310, 850
2C	2.0	10	420, 780
3C	2.0	10	140, 700
4C	2.0	10	240, 460
5C	2.0	10	200, 280
6C	2.0	10	380, 260
7C	2.0	10	440, 100

Tolerances: Depth 0.1mm, Length 1mm, Location 5mm

TABLE 1 Defect details TP1

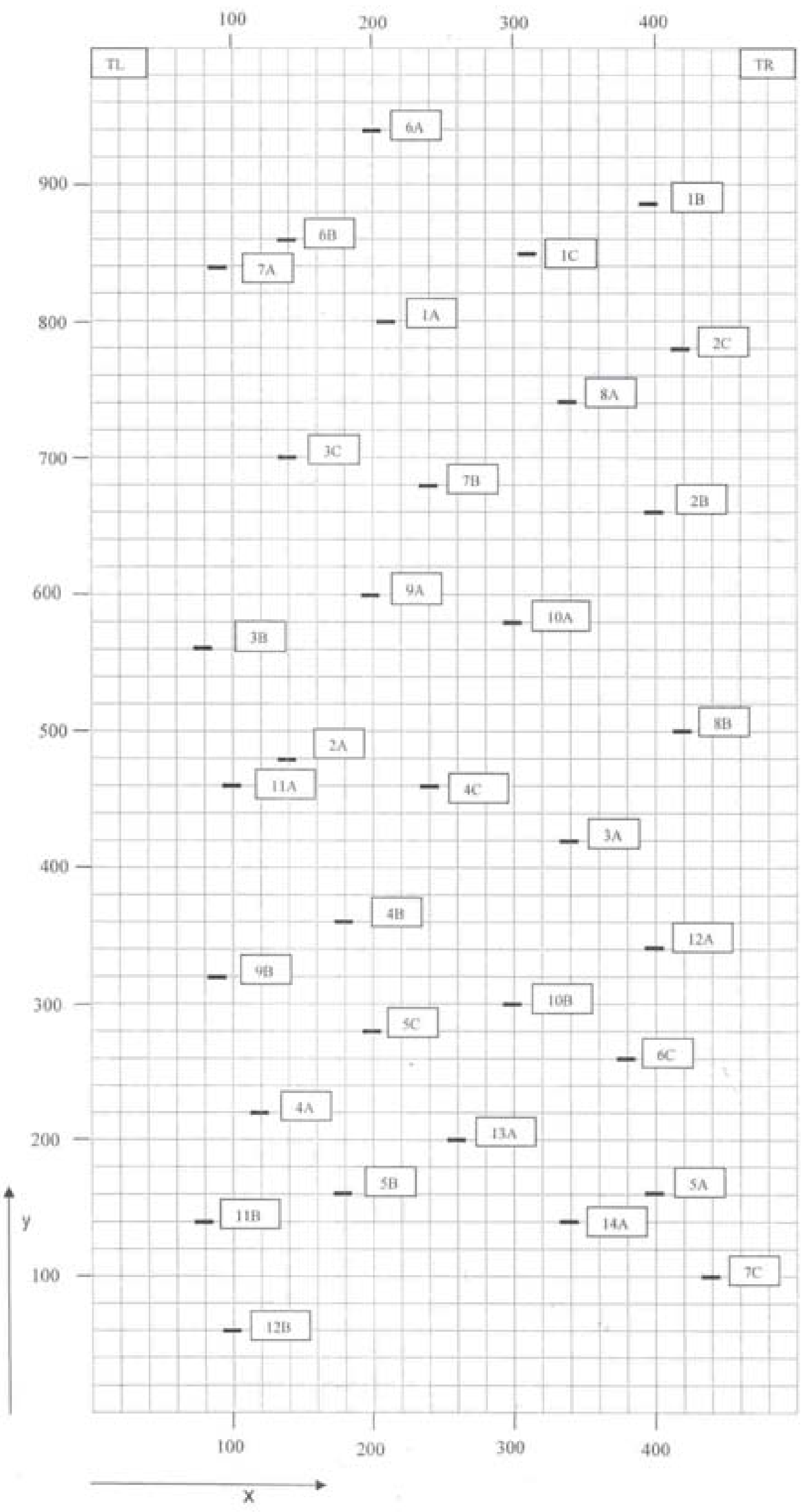

FIGURE 1 Testpiece 1

Slit	Depth (mm)	Length (mm)	Location slit centre(x, y)mm
1D	3	10	120, 880
2D	3	10	400, 720
3D	3	10	410, 500
4D	3	10	360, 400
5D	3	10	180, 160
1E	5	10	210, 800
2E	5	10	150, 700
3E	5	10	60, 480
4E	5	10	80, 180
5E	5	10	400, 160
6E	5	10	300, 110
7E	5	10	310, 220
8E	5	10	210, 240
9E	5	10	440, 320
10E	5	10	110, 360
11E	5	10	250, 460
12E	5	10	290, 620
13E	5	10	80, 780
14E	5	10	370, 820
15E	5	10	300, 900

Tolerances: Depth 0.1mm, Length 1mm, Location 5mm

TABLE 2 Defect details TP2

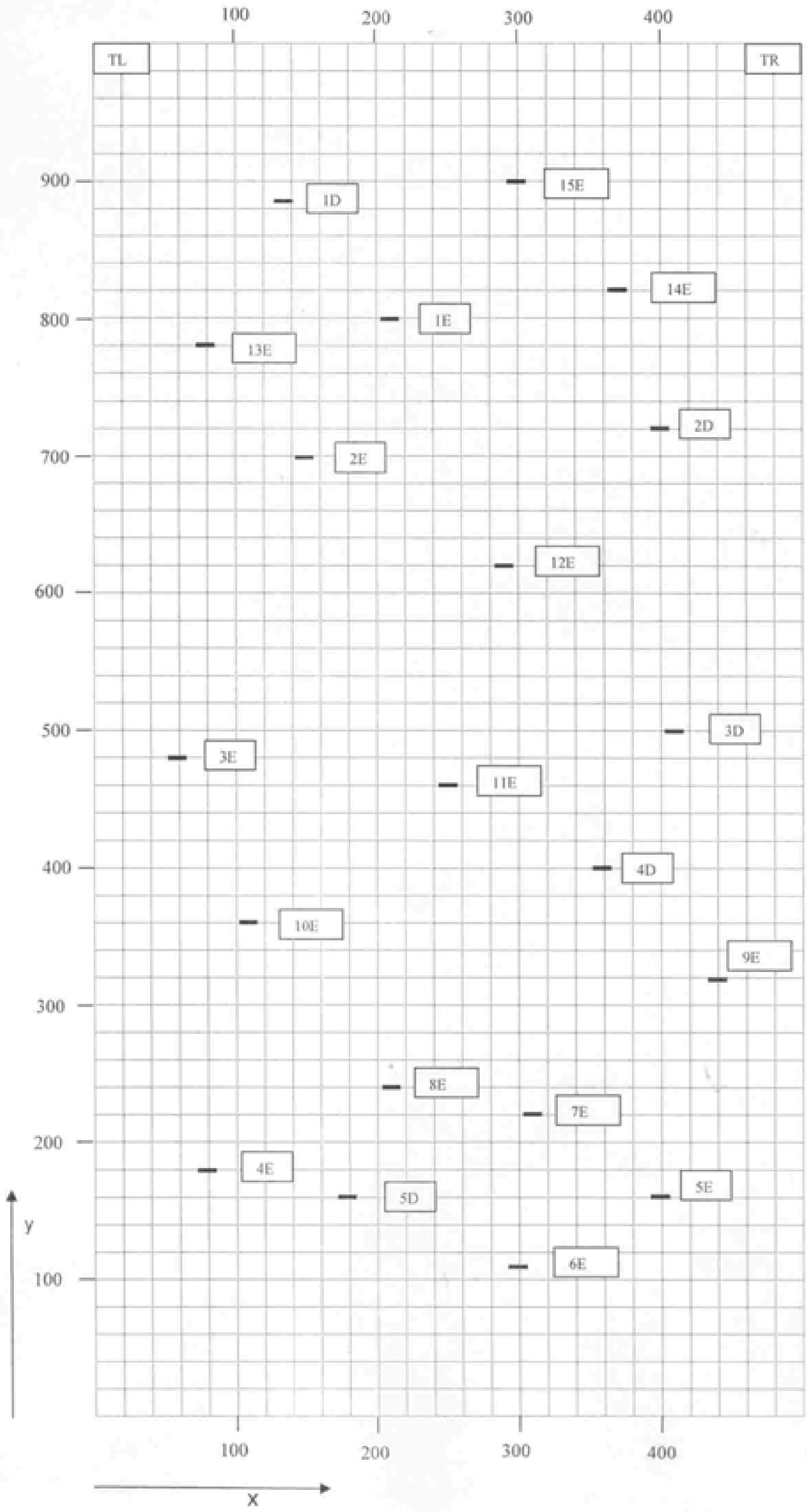

FIGURE 2 Testpiece 2

Slit	Depth (mm)	Length (mm)	Location slit centre(x, y)mm
15A	0.5	10	120, 880
16A	0.5	10	385, 820
17A	0.5	10	305, 720
18A	0.5	10	80, 560
19A	0.5	10	420, 480
20A	0.5	10	170, 150
21A	0.5	10	390, 150
13B	1	10	250, 850
14B	1	10	170, 800
15B	1	10	390, 560
16B	1	10	140, 420
17B	1	10	120, 215
8C	2	10	250, 620
9C	2	10	260, 540
10C	2	10	180, 360
11C	2	10	360, 330
16E	5	10	60, 800
17E	5	10	440, 680
18E	5	10	70, 650
19E	5	10	430, 245
20E	5	10	70, 105

Tolerances: Depth 0.1mm, Length 1mm, Location 5mm

TABLE 3 Defect details TP3

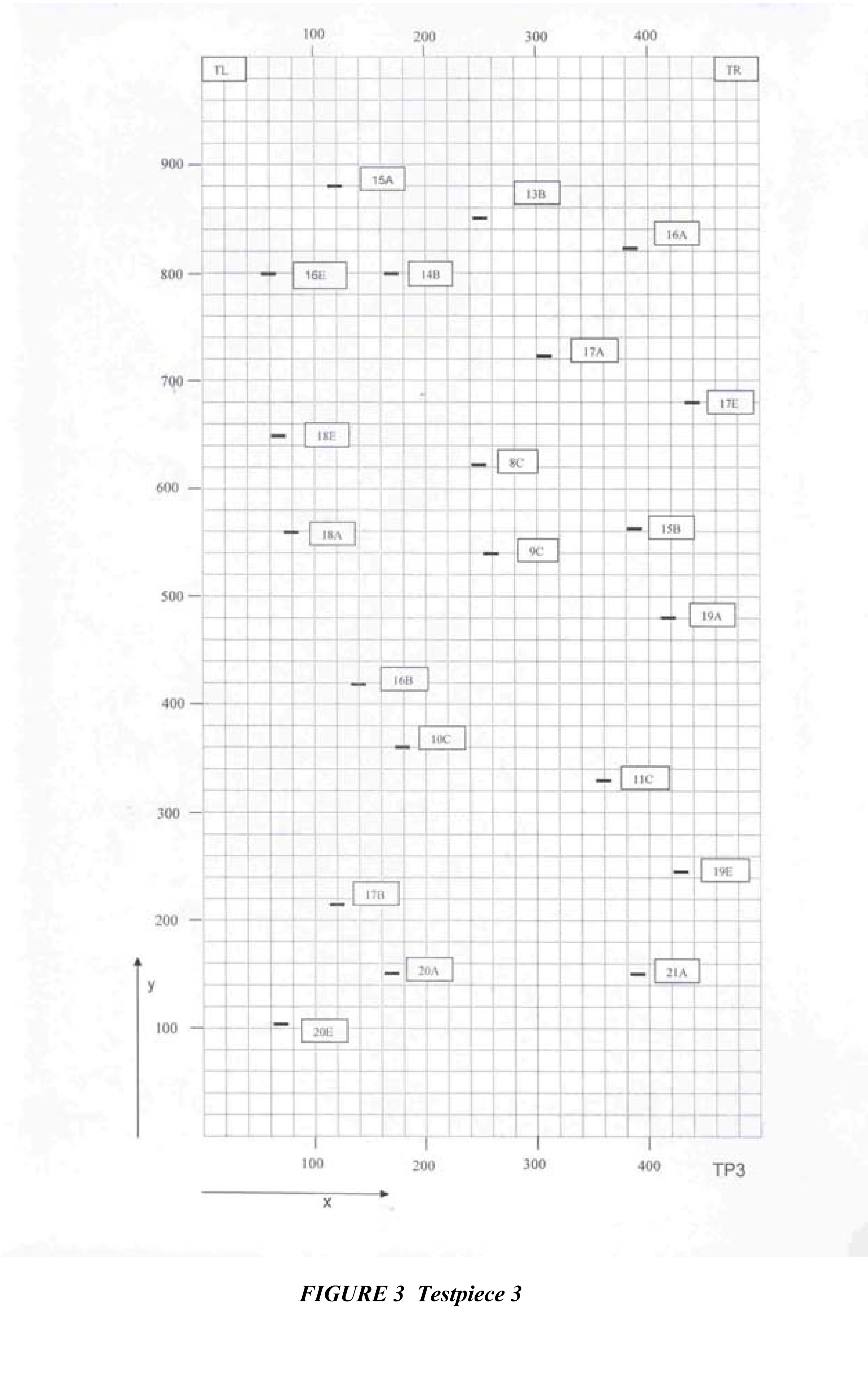

FIGURE 3 Testpiece 3

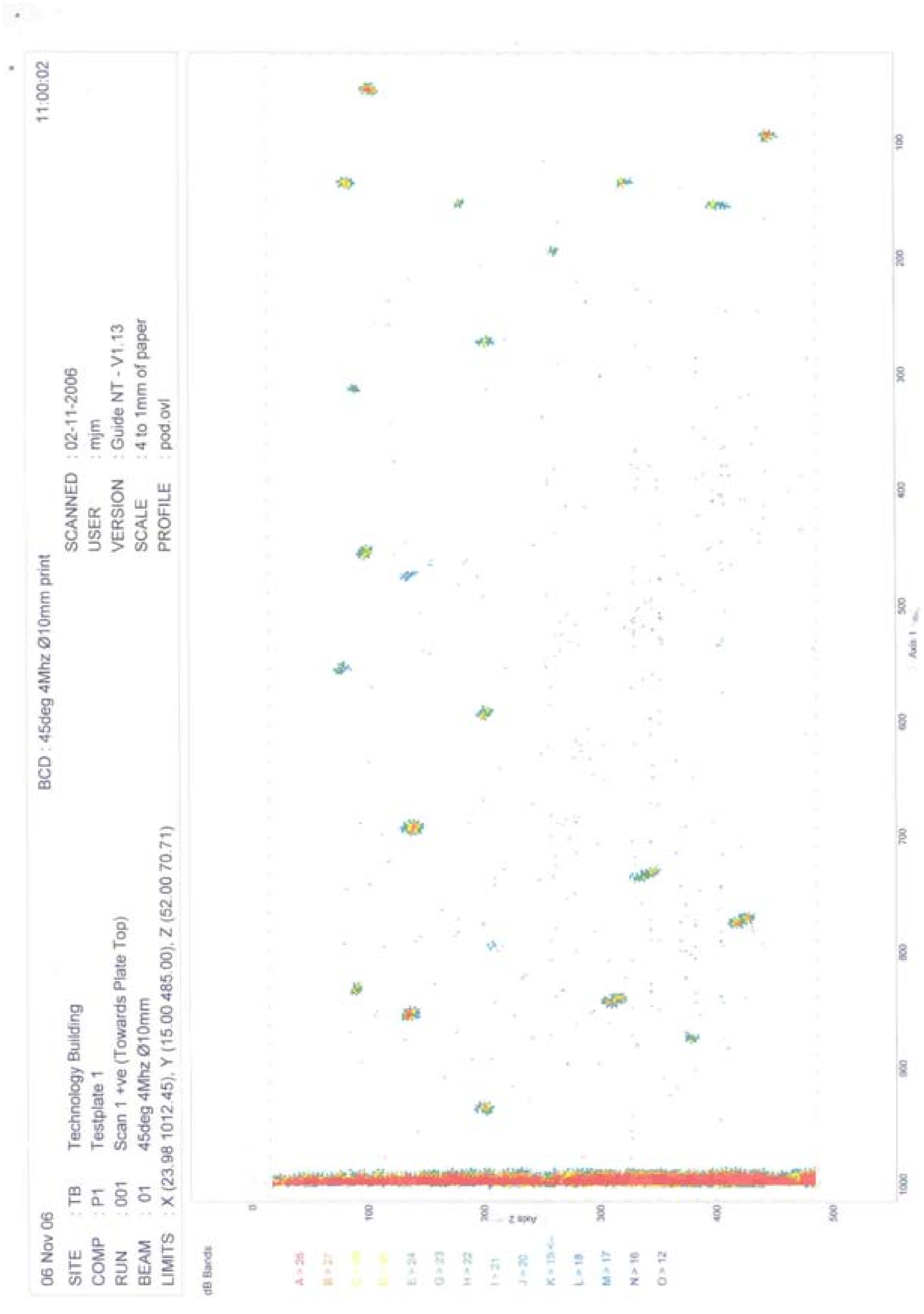

FIGURE 4 Example of Automated C-scan

TP1	M1	M2	M3	M4	M5	M6	M7	S/N meanA	S/N meanB	hits	misses
1A	-	-	-	-	9	-	-	1.3	9	1	6
2A	-	-	8	-	-	4	-	1.7	6	2	5
3A	-	-	-	-	-	-	-	-	-	0	7
4A	-	-	-	-	-	-	-	-	-	0	7
5A	-	10	-	12	12	9	-	6.4	10.8	4	3
6A	12	-	-	13	16	16	16	10.4	14.6	5	2
7A	10	7	9	6	-	-	-	4.6	8	4	3
8A	10	8	-	-	-	7	-	3.6	8.3	3	4
9A	8	-	-	9	8	9	14	6.9	9.6	5	2
10A	-	-	-	-	-	(8)	-	-	-	0	7
11A	10	10	9	5	13	8	-	7.9	9.2	6	1
12A	-	-	-	-	-	-	-	-	-	0	7
13A	8	6	7	10	9	10	9	8.4	8.4	7	0
14A	10	-	12	13	11	10	-	8.0	11.2	5	2
1B	-	-	10	-	-	10	-	2.9	10	2	5
2B	-	-	-	-	6	-	-	0.9	6	1	6
3B	-	9	10	-	10	11	-	5.7	10	4	3
4B	-	-	-	-	-	-	-	-	-	0	7
5B	-	-	10	-	-	10	-	2.9	10	2	5
6B	12	13	9	12	10	12	20	12.6	12.6	7	0
7B	-	-	-	-	-	-	-	-	-	0	7
8B	-	-	-	-	-	-	-	-	-	0	7
9B	6	-	8	8	-	5	-	3.9	6.8	4	3
10B	-	-	-	-	-	-	-	-	-	0	7
11B	10	12	11	12	14	11	17	12.4	12.4	7	0
12B	8	15	15	15	17	19	20	15.6	15.6	7	0
1C	8	12	10	-	11	11	14	9.4	11	6	1
2C	9	11	9	11	14	10	14	11.1	11.1	7	0
3C	14	9	-	-	12	10	18	9	12.6	5	2
4C	-	-	-	-	-	-	9	1.3	9	1	6
5C	7	-	7	-	10	6	8	5.4	7.6	5	2
6C	-	-	-	-	-	-	-	-	-	0	7
7C	12	13	12	17	14	9	17	13.4	13.4	7	0

S/N meanA takes S/N as zero if not detected
S/N meanB is averaged only for hits, i.e. misses don't reduce S/N
Defect 10A M6 result not counted since assumed to be false call - 27mm from defect and low amplitude.

TABLE 4 S/N Results manual UT

S/N mean (B)	Hits manual	Misses manual	**"POD" manual**
0-2.4	-	-	-
2.5-4.9	-	-	-
5.0-7.4	7	14	**33**
7.5-9.9	32	24	**57**
10.0-12.4	37	19	**66**
12.5-14.9	24	4	**86**
15.0-17.4	7	0	**100**
17.5-19.9	-	-	-

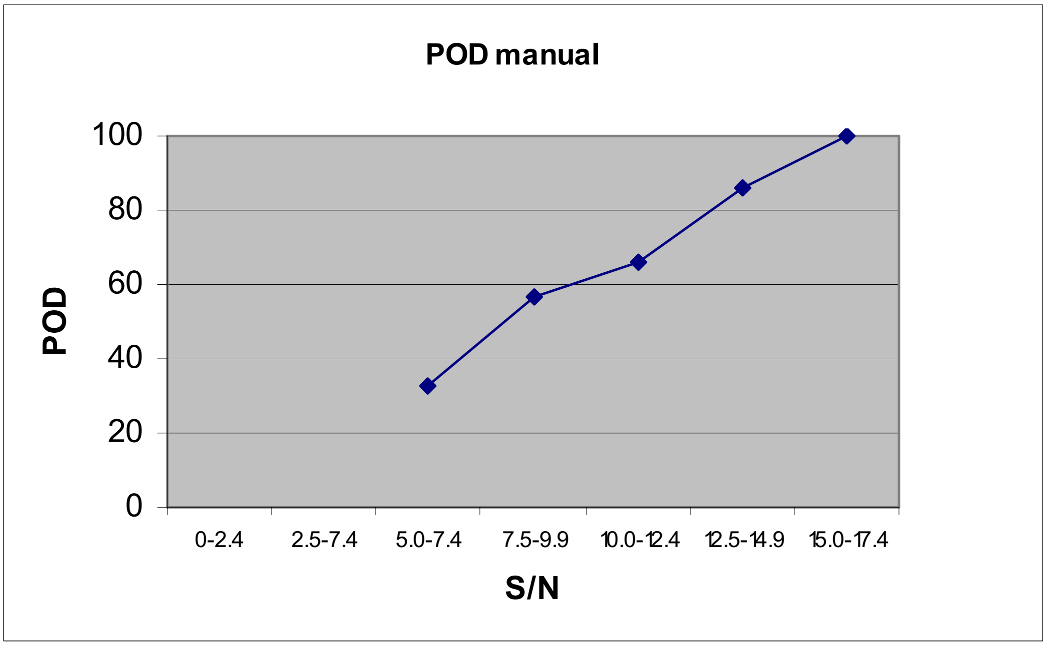

TABLE 5 Mean Signal to Noise Ratio versus "POD" (manual)

TP1	S/N meanB	M2 + M3 hit	M5 + M6 hit
1A	9	0	1
2	6	1	1
3	-	-	-
4	-	-	-
5	10.8	0	1
6	14.6	0	1
7	8	1	0
8	8.3	1	1
9	9.6	0	1
10	-	-	-
11	9.2	1	1
12	-	-	-
13	8.4	1	1
14A	11.2	1	1
1B	10	1	1
2	6	0	1
3	10	1	1
4	-	-	-
5	10	1	1
6	12.6	1	1
7	-	-	-
8	-	-	-
9	6.8	1	1
10	-	-	-
11	12.4	1	1
12B	15.6	1	1
1C	11	1	1
2	11.1	1	1
3	12.6	1	1
4	9	0	0
5	7.6	1	1
6	-	-	-
7C	13.4	1	1

S/N mean (B)	Hits	Misses	**"POD" Diverse x 2**
0-2.4	-		**-**
2.5-4.9	-		**-**
5.0-7.4	5	1	**83**
7.5-9.9	11	5	**69**
10.0-12.4	15	1	**94**
12.5-14.9	7	1	**88**
15.0-17.4	2	0	**100**
17.5-19.9			

TABLE 6b S/N versus "POD" for combined independent manual scans

TABLE 6a Results from combined independent manual scans

TP1	M1	M3	M4	M5	M6	M7	Mean	Δ	hits	misses
1A	-	-	-	71	-	-	71		0	6
2	-	71	-	-	72	-	71.5	1	0	6
3	-	-	-	-	-	-	-		0	6
4	-	-	-	-	-	-	-		0	6
5	-	-	61	69	65	-	65	8	2	4
6	66	-	63	65	64	64	64.4	3	5	1
7	68	69	66	-	-	-	67.7	3	1	5
8	67	-	-	-	68	-	67.5	1	1	5
9	68	-	63	72	67	66	65.2	9	3	3
10	-	-	-	-	-	-	-		0	6
11	65	71	73	69	68	-	69.2	8	1	5
12	-	-	-	-	-	-	-		0	6
13	68	71	66	73	69	71	69.7	7	1	5
14A	67	68	63	71	68	-	67.4	8	2	4
1B	-	68	-	-	67	-	67.5	1	1	5
2	-	-	-	74	-	-	74		0	6
3	-	67	-	72	67	-	68.7	5	2	4
4	-	-	-	-	-	-	-		0	6
5	-	71	-	-	70	-	70.5	1	0	6
6	65	69	60	69	62	60	64.2	9	4	2
7	-	-	-	-	-	-	-		0	6
8	-	-	-	-	-	-	-		0	6
9	-	71	65	-	71	-	69	6	1	5
10	-	-	-	-	-	-	-		0	6
11	65	67	62	71	67	63	65.8	9	5	1
12B	67	62	59	65	61	60	62.3	8	6	0
1C	67	69	-	71	67	66	68	5	3	3
2	66	67	64	69	65	66	66.2	5	5	1
3	62	-	-	69	65	62	64.5	7	3	3
4	-	-	-	-	-	71	71		0	6
5	68	70	-	71	69	70	69.6	2	0	6
6	-	-	-	-	-	-	-		0	6
7C	64	66	58	67	67	63	64.2	9	6	0

TP 1 Signal gain to 80% fsh

MO 2 excluded (different set)

Δ = spread

Arbitrary threshold 67dB

TABLE 7 Signal Amplitude Results (Manual)

Amplitude with respect to 67dB threshold	Amplitude (Gain to 80%fsh)	Hits manual	Misses manual	**"POR" manual**
7.5 – 5.1 below	74.5 – 72.1	0	6	**0**
5.0 – 2.6 below	72 – 69.6	1	35	**3%**
2.5 – 0.1. below	69.5 – 67.1	12	36	**25%**
0 – 2.4 above	67 – 64.6	15	9	**62%**
2.5 – 4.9 above	64.5 – 62.1	24	6	**80%**

TABLE 8 Manual Amplitude versus "POR"

TP1	S/N meanB	AO1 hit ↑	AO2 hit ↑	AO3 hit ↑
1A	9	0	1	0
2	6	1	1	1
3	-	0	0	0
4	-	0	0	0
5	10.8	1	1	1
6	14.6	1	1	1
7	8	1	1	1
8	8	1	1	1
9	9.6	1	1	1
10	-	0	0	0
11	9.2	1	1	1
12	-	0	0	0
13	8.4	1	1	1
14A	11.2	1	1	1
1B	10	1	1	1
2	6	0	0	0
3	10	1	1	1
4	-	0	0	0
5	10	1	1	1
6	12.6	1	1	1
7	-	0	0	0
8	-	0	0	0
9	6.8	1	1	1
10	-	0	0	0
11	12.4	1	1	1
12B	15.6	1	1	1
1C	11	1	1	1
2	11.1	1	1	1
3	12.6	1	1	1
4	9	0	0	0
5	7.6	1	1	1
6	-	0	0	0
7C	13.4	1	1	1

All S/N based on manual results

Table 9a Auto UT results

S/N mean (B)	Hits auto↑	Misses auto↑	**"POD" auto↑**
0-2.4	-	-	-
2.5-4.9	-	-	-
5.0-7.4	6	3	**67**
7.5-9.9	19	5	**79**
10.0-12.4	24	0	**100**
12.5-14.9	12	0	**100**
15.0-17.4	3	0	**100**
17.5-19.9	-	-	-

Table 9b S/N versus "POD" (auto)

www.ski.se

STATENS KÄRNKRAFTINSPEKTION
Swedish Nuclear Power Inspectorate

POST/POSTAL ADDRESS SE-106 58 Stockholm
BESÖK/OFFICE Klarabergsviadukten 90
TELEFON/TELEPHONE +46 (0)8 698 84 00
TELEFAX +46 (0)8 661 90 86
E-POST/E-MAIL ski@ski.se
WEBBPLATS/WEB SITE www.ski.se