Producing a POD curve with emulated signal response data

T Koskinen, I Virkkunen, S Papula, T Sarikka and J Haapalainen

Ultrasonic testing is an important tool for assessing the structural integrity of pressurised components in nuclear power plants during in-service inspection. Consequently, the reliability of inspection is of high significance. In particular, it is important to determine the largest crack that could conceivably be missed during in-service inspection. This information is utilised in order to choose the most effective method for different situations. Probability of detection (POD) curves are used to quantify the effectiveness of the inspection. However, these POD curves require a lot of data points in order to provide reliable estimates of the lower limit performance. Traditionally, obtaining these curves has been relatively expensive and this is why different simulation tools have been used to reduce costs and the number of physical test-pieces needed. In the current study, a novel method based on limited test-pieces is explored to produce a POD curve from the measured data. The present data contains just three artificial cracks made with thermal fatigue. This is insufficient to produce a POD curve. However, the data was sufficient to evaluate the novel approach and to gain preliminary estimates of the expected POD and the most important influential parameters. In this study, the idea is to emulate the amplitude response from the measured crack in a way that represents an amplitude response from a certain crack size. This amplitude data is converted to a B-scan image from which the inspector will evaluate whether there is a crack or not. Then, a POD curve is generated from the achieved hit/miss data.

1. Introduction

Probability of detection (POD) describes the ability of a non-destructive testing (NDT) technique to detect a specific crack size. This information is needed for comparative analysis of different NDT methods and crack sizes. Furthermore, POD is used to plan inspection intervals[1].

In order to estimate a POD curve, either a quantitative signal or a binary hit/miss response from the inspection is required. The detectable targets need to have measurable characteristics, such as size or chemical composition. In most cases, the height of the crack \( a \) is chosen as a measured characteristic. However, in most situations, the signal response and other parameters affecting the detection of the crack (such as signal-to-noise ratio (SNR)) are not necessarily exactly the same for every crack of the exact same size. This is due to the fact that there are parameters other than the size alone affecting the detection of a crack. These parameters are often related to how the ultrasound scatters or reflects from the surface of a crack. Thus, the orientation of the crack relative to the propagating ultrasonic beam and the surface roughness of the crack are important factors for crack detection. Also, a human factor may affect the end result of the inspection[2,3].

POD\((a)\) is a function that represents a proportion of all cracks of size \( a \) for a specific NDT method, thus it determines a POD curve as a function of crack size. Each individual crack of size \( a \) has its own detection probability, leading to a probability density function of PODs, which is used to define a mean POD value for a certain crack size. An example of a POD curve and an associated probability density function can be seen in Figure 1[1,3].

Without any statistical methods, a hit/miss POD curve would require an enormous number of cracks. This is constructed more efficiently by assuming a mathematical relationship between POD and size. In this case, the size must correlate with the POD, even though it is known that the detection is not only related to the size of the crack. Also, it must be taken into account that the model response does not vary continuously and is not unbounded. This is the reason why ordinary linear regression cannot be used. This can be worked around by using generalised linear models (GLMs) and transforming the probability to a polynomial function, which is linear in the explanatory variables. Unlike in traditional regression, the variance of the probability function is not constant and a maximum likelihood method is used to estimate GLM parameters. In other words, maximum likelihood is used to determine the parameters that are the most probable for certain test data[3].
1.1 Estimating probability of detection

When estimating the probability of detection as a function of crack size using hit/miss data, a parametric POD model is used. In this model, a maximum likelihood analysis is made for the acquired hit/miss data. GLM is used for this analysis and link functions are used to link the binary response to the function of \( x \), \( g(y) = f(X) \) to ‘generalise’ this model. These link functions are the logit link, the probit link, the log-log and the complementary log-log (cloglog). Logit and probit links are similar to each other, however, their largest differences come from the extremes. The probit link has a Gaussian probability density and suffers from higher sensitivity to improbable outcomes, so it may produce unreliable solutions when the data is highly random. Mainly, the function is S-shaped, clearly describing the relationship between POD and crack size. Thus, the logit link is used more often over the probit link, unless there are great advantages in using the probit link. The log-log and the complementary log-log functions differ from the logit and probit link functions by skewing to the right and left, respectively, while the logit and probit links are symmetrical\[4,5].

In most cases, the choice between probit and logit with \( a \) or log(\( a \)) is decided by comparing the deviance of the four possible combinations. The combination with the lowest deviance indicates the best fit for the model. Figure 2 demonstrates the overall eight possibilities, with log-log and complementary log-log links included\[5].

When deriving a POD curve, confidence bounds are used to estimate the plausibility and repeatability of the results. Cheng and Iles developed a procedure to determine these confidence bounds by calculating the variance-covariance matrix of the estimates and reflecting the sensitivity of the experiment to both the number and size of the cracks. The most common confidence bound used in PODs is 95%, which means that the given result is repeatable 95% of the time. For example, \( a_{\text{hit}} \) would mean the crack size has a probability of detection of 90% and \( a_{\text{miss}} \) means the crack size has a confidence level of 95%. This means that a crack with a size of \( a \) will be detected with a probability of 90% and, if this POD exercise is repeated, 95% of the results will fall inside this confidence limit\[4,6].

Another way of determining the confidence bound is by using the maximum likelihood estimate (MLE). This is based on the idea of choosing the parameters that represent the data the most, thus explaining the outcome better and maximising the likelihood. When choosing values that differ slightly from the maximum likelihood parameters, the likelihood diminishes, as expected. However, if the values are moved too far it results in values that are unlikely to describe the overall data. Therefore, the point is to stay close enough to the MLE parameters. Fortunately, the asymptotic behaviour of the ratio of the logs of the new values enables the determination of close values. This ratio is called the log-likelihood ratio\[4,6].

The log-likelihood ratio surface can be seen in Figure 3. The values \( \mu \) and \( \sigma \) are model parameters that are moved from their MLE position, which is marked with a large + in the Figure. The line consisting of alternating dashes and dots represents the 95% confidence bound, meaning the true \( \mu \) and \( \sigma \) pair is expected to be found 95% of the time within the area. In the Figure, the Cheng and Iles approximation (Cheng and Iles, 1983) is also demonstrated. However, it is a poor approximation for small sample sizes. When the sample size rises, the resulting contours should contract towards the MLE\[4,6].

Despite the statistical models, a considerable number of data points is still needed in order to achieve a reliable POD curve. The data should also concentrate on the most important transition region. Figure 4 shows a typical distribution of detected and missed cracks. As can be seen from the Figure, there is a large group of cracks between the smallest crack detected \( a_{\text{min}} \) and the largest crack missed \( a_{\text{max}} \).

To lower the number of different cracks needed, a model-assisted POD (MAPOD) is used. In this situation, the crack and the experimental set-up are simulated with finite element modelling (FEM) or another simulation method. The idea is to establish a relationship between the physical characteristics and the detection of the crack. A MAPOD approach is best suited to NDT situations with a signal response \( \Delta \). It is possible to obtain a hit/miss result using MAPOD as well, however, this is usually carried out using the signal response parameters. Thus, analysis in terms of signal response is a much more common method in MAPOD than the hit/miss approach\[4,6].

The ideal use for a POD curve is to estimate the size of the cracks it is feasible to search for. When the crack growth rate is taken into account, an economic inspection interval can be calculated. The ideal situation would be that a crack
could be easily detected, ie the size would be large enough for a probable detection. However, this should be carried out before the crack size reaches a critical size and endangers the structural integrity. The POD curves need to be reliable and accurate in order to provide a sound basis for these kinds of decision\[7\].

It is imperative to use these statistical methods to evaluate the acquired inspection data. Also, these methods are used in order to reduce the number of real samples by approximating the collected data in the most efficient way. However, the sample size remains relatively large due to the fact that the more samples there are the more reliable the statistical methods are.

2. Experimental procedure

A single weld test-piece containing three cracks was used in this study. The test-piece was a welded austenitic stainless steel 321 pipe representing an actual geometry found in a nuclear power plant primary circuit. The weld was 30 mm wide and the pipe thickness was 40 mm. The outer diameter of the pipe was around 400 mm. The weld inspected was a U-groove weld with a 3 mm root gap. A schematic of the pipe and the weld can be seen in Figure 5.

Table 1. Crack characteristics. The values given were validated with destructive testing

<table>
<thead>
<tr>
<th>Crack ID</th>
<th>Size (length × height) (mm)</th>
<th>Average opening (µm)</th>
<th>Median opening (µm)</th>
</tr>
</thead>
<tbody>
<tr>
<td>118BFB2761</td>
<td>23 × 8.5</td>
<td>119.1</td>
<td>123.5</td>
</tr>
<tr>
<td>118BFB2762</td>
<td>24 × 4.0</td>
<td>64.8</td>
<td>65.3</td>
</tr>
<tr>
<td>158BGB3709</td>
<td>6.2 × 1.6</td>
<td>*</td>
<td>*</td>
</tr>
</tbody>
</table>

*Due to the small crack opening, the opening could not be reliably measured

Figure 5. Schematic of the test-piece

The liquid penetrant testing indications from the artificial cracks can be seen in Figures 7, 8 and 9. As seen in the Figures, the shape of the cracks was linear with no branching. The smallest crack in Figure 9 had such a small crack opening it could not be reliably measured. This indicates that the crack might be harder to detect due to the small crack opening.

2.2 Scanning procedure

The probe used was a 1.5 MHz transmit-receive shear (TRS) wave probe with a wedge curved to the shape of the outer diameter of the pipe and water was used as a coupling agent. The weld centreline was selected as the origin in the index direction. The scan axis was
around the circumference of the pipe and the index axis was set lengthwise along the pipe. The scan area was set between 34 mm from the weld centreline to 54 mm from the weld centreline, from both sides of the weld. The index resolution was 20 mm, thus the number of scan lines was two. The focal laws were set for 40° to 70° sectorial scanning with a 1° step. The scanning procedure can be seen in Figure 10 and the overall area can be seen in Figure 11.

A Zetec manual pipe scanner was used to record the probe location in the X- and Y-coordinates. The pipe was set horizontally, resting on a support in a position where the inspected cracks were in the upper part of the pipe. This set-up ensured proper contact between the probe and the specimen. The scanning set-up can be seen in further detail in Figure 12.

The microstructure of the austenitic stainless steel weld caused significant noise during the ultrasonic inspection. Despite this, all three of the cracks could be detected with the TRS method. The signal-to-noise ratio for the smallest crack was below two, indicating another probable crack was detected in an unlikely location, possibly due to the noise.

A Zetec manual pipe scanner was used to record the probe location in the X- and Y-coordinates. The pipe was set horizontally, resting on a support in a position where the inspected cracks were in the upper part of the pipe. This set-up ensured proper contact between the probe and the specimen. The scanning set-up can be seen in further detail in Figure 12.

Figure 7. 118FB2761 – Linear thermal fatigue crack located in the HAZ with 23 mm length and 8 mm height. This crack had the largest crack opening and height out of the three cracks, indicating the strongest amplitude response.

Figure 8. 118FB2762 – Linear thermal fatigue crack located in the HAZ with 24 mm length and 4 mm height. Of the three cracks, this crack had an intermediate crack opening and height, indicating that the amplitude response lies between the strongest and the weakest.

Figure 9. 158BGB3709 – Linear thermal fatigue crack located in the HAZ with 6.2 mm length and 1.6 mm depth. This is the smallest crack of the three and thus the most unlikely to be detected.

Figure 10. Scanning procedure. The weld was scanned from both sides of the weld.

Figure 11. Scan area. The scanning was only performed near the area of the test specimen where the cracks resided.

Figure 12. The scanning set-up. The test-piece can be seen on the left side. An extension on the right side was glued to the test-piece for the pipe roller attachment.
the detection of this crack would not realistically be expected in blind inspection (not knowing where the cracks reside). Since the method used generated a lot of ultrasonic data and considering this test was thought of as a proof of concept, only the 45° angle of the closest scan line was used for further analysis. This was chosen because the 45° angle is one of the most common angles used for weld inspection and showed clear signals from the cracks.

2.3 Modification of amplitude response from artificial cracks

The scanned data was acquired in the UVData file format, containing both the raw UT data and the necessary metadata to interpret the various signals included. The raw UT data was extracted from the file and the most interesting channels were selected for this initial study. The data contains a wealth of other data that will be exploited in further research. For this study, a single scan line with a 45° beam angle was selected and extracted for further analysis. Figure 13 shows a view along the weld from the raw data extracted with the crack signals marked. In addition to the crack signals, some locations without crack signals were marked for removal and reintroduction. This was carried out in order to exclude the possibility that the reintroduction itself causes some detectable changes that the inspectors learn to look for.

The crack signals were identified and extracted by comparison with a manually selected flawless area at the same flight time area. The extent of the crack signal was determined manually using expert judgement. Figure 14 shows a similar view along the weld but with the crack signals removed. This data file functions as the 'baseline' data into which extracted crack signals are later introduced.

Next, the extracted crack signals were reintroduced into the data. Ideally, the crack signals should represent the breadth and variety of the cracks that could be encountered in the actual inspection. In particular, the crack sizes of the natural cracks should cover the range of interest for the expected or required inspection task. With sufficient crack data, the crack signals could be introduced unmodified while still providing sufficient data for POD determination.

In this study, the number of real cracks was three and, while the sizes did cover the expected range of interest, the number of cracks was far too small to provide sufficient data for POD determination as it is. Thus, each real crack signal was altered to make additional crack sizes available.

It was taken as a premise that the signal amplitude is the most significant feature of the crack signal from a detection point of view. (A similar premise is used, for example, in the a versus d POD analysis routinely conducted.) In particular, the crack amplitude (signal) in relation to the noise of the weld would be the most significant feature determining crack detection. Nevertheless, the premise is somewhat questionable for UT inspection and is further discussed in Section 3.

With this premise, each crack with a given size was modified to give several different equivalent crack sizes with the following formulae:

\[
A_{\text{shown}} = A_{\text{true}} + F \cdot (A_{\text{true}} - A_{\text{noise}}) \quad (1)
\]

\[
a_{eq} = a_{\text{true}} + F \cdot (a_{\text{true}} - a_{\text{noise}}) \quad (2)
\]

where:
- \(A_{\text{shown}}\) is the amplitude of the signal shown to the inspector;
- \(A_{\text{true}}\) is the average noise amplitude of the reference area (manually selected);
- \(A_{\text{noise}}\) is the amplitude of the true signal, unaltered;
- \(F\) is the scale factor applied (0.0-1.0);
- \(a_{\text{eq}}\) is the equivalent crack depth;
- \(a_{\text{true}}\) is the crack size that corresponds to the average noise level; and
- \(a_{\text{noise}}\) is the true crack size.

These formulae were obtained by linear interpolation between the local noise level and the original acquired amplitude. For \(F = 0\), Equation (1) results in no change in data, whereas for \(F = 1\), the acquired flaw signal is inserted as it is. Similarly, the reported equivalent flaw size is computed by linear interpolation between theoretical crack size with amplitude at noise level (at \(F = 0\)) and the actual flaw size of the crack (at \(F = 1\)). The linear interpolation is a rather crude approximation and, in the future, additional data on actual flaws can be used to support models that are more sophisticated.

The amplitude in the acquired UT data varies with the probe-to-defect range and thus Equations (1) and (2) are understood to apply in proportion to the local (range-dependent) amplitude. Figure 15 shows data with scaled cracks of different size introduced into the file.

In order to gather inspection information, an online interface (used with a modern web browser) was written to present the inspector with a series of generated views along the weld, each with randomly introduced cracks (0-3 cracks per image) and randomly scaled. The interface allows the user to change the 'software gain' on the data and to indicate cracks by clicking on the image. The programme randomly generated 150 such images, presented them and then collected associated crack indications.
from the user. This data was then analysed using the conventional ASTM-E2862 methodology to estimate the POD curve and the associated confidence levels.

Figure 15. View along the weld with scaled cracks introduced to the data. The true depth of the crack is 4.0 mm and it is here scaled to the equivalent crack sizes of 1.7 mm, 2.5 mm, 3.2 mm and 4.0 mm. Crack areas are marked with red rectangles

3. Results and discussion

It was taken as a premise that the signal amplitude is the most significant feature of the crack signal from a detection point of view. For ultrasonic inspection, it can be argued that the approach taken is an oversimplification and that other features (such as the shape of the signal) affect crack detection. At the same time, simplification is clearly necessary to estimate POD with limited real cracks. In addition to the present approach, model-assisted POD estimation (using mathematical simulations) and various other approaches used to reduce the number of cracks needed are subject to the same question: whether the simplifications made in order to obtain a POD curve estimate are justifiable in terms of the errors they may induce. The allowable errors and simplifications depend heavily on the intended use of the results and thus the necessary justification also depends on the intended use.

To estimate the validity of this premise, the data was analysed to look for tell-tale signs of crack detection based on features other than the amplitude. If inspectors detected cracks based on features other than the amplitude, the POD for large original cracks (with prominent features) should be higher than the POD for small cracks scaled to equal equivalent crack size. To test this, the available data was combined from different inspectors and divided according to the original crack size. The results are shown in Figure 16. Overall, it seems that the premise is quite well justified for a rather wide range of equivalent crack sizes compared to the real size of the crack. However, the ‘middle-sized’ crack with an original size of 4.0 mm seems to be markedly more difficult to detect than the smaller (1.6 mm) and larger (8.6 mm) cracks. This is likely to be a result of inter-crack variation and not a size effect. The cracks with 1.6 mm and 8.6 mm original sizes show surprisingly consistent results. The limited evidence thus suggests that inter-crack variation is a more significant factor in probability of detection than the signal amplitude itself. Thus, the data supports the chosen approach.

The extraction and reinsertion was also performed in some data regions that did not include a crack. This was carried out in order to investigate the possibility that the extraction and reinsertion process itself introduces some features that the inspectors learn to detect. If this were the case, the inspectors would indicate cracks in these locations. However, these locations were not indicated at a rate higher than the overall false call rate and thus no evidence was found for problems caused by the extraction and reinsertion.

A typical result for an inspection can be seen in Table 2.

Feedback from the inspectors revealed that they would have liked to have had more control over the inspection. Most of the interviewed inspectors stated that they would have liked to see the A-scan in order to evaluate the crack more efficiently. Also, other information such as a C-scan would have been helpful. In addition, utilisation of more or different beam angles was suggested. They stated that this would have probably lowered the number of false calls or made the detection of some missed cracks possible.

The inspectors also brought up a point that the shape of the crack stayed roughly the same, which made detection of the crack easier in some situations. Also, in the cases where a pure noise signal resembled a crack and stayed roughly the same shape, the inspector might have interpreted it as a crack numerous times, thus increasing the number of false calls artificially.

4. Conclusions

The experiment showed promising results. Inspectors were able to find cracks and generate a plausible POD curve for the application, even with the lack of an A-scan image and with minimal information. Of course, this might have led to a large number of false calls in some situations. The results indicate that it is possible to emulate the amplitude response of a crack and produce an amplitude response image for an inspector to evaluate. This in turn can be used to generate hit/miss data with real inspector responses. The results were achieved with only three real cracks, which is obviously too few to achieve accurate results. However, this indicates that when using this kind of an approach to generate a POD curve, costs could be reduced. It also provides a possible comparison tool with which to
investigate different parameters in NDT inspection through a cheap POD curve.

Further research needs to be carried out with a larger number of real cracks and also with a larger inspector group. To achieve a more appropriate inspection situation, an S-scan with various different angles, A-scan images and C-scan images should also be implemented to provide the inspector with a more life-like situation, thus producing more reliable and plausible results. Also, it needs to be taken into account how the human eye gets used to similar shapes, how that affects the results of this kind of experiment and how these shapes and noise can be made to randomly change shape and still represent the original cracks accurately.

Acknowledgements
The authors wish to thank Esa Leskelä for his invaluable comments.

References