Koskinen, Tuomas; Virkkunen, Mikko

Hit/Miss POD With Model Assisted and Emulated Flaws

Published in:
12th European Conference on Non-Destructive Testing (ECNDT 2018)

Published: 01/08/2018

Document Version
Publisher's PDF, also known as Version of record

Hit/Miss POD With Model Assisted and Emulated Flaws

Tuomas Koskinen[1] and Iikka Virkkunen[2]
1 VTT Technical Research Centre of Finland ltd, Finland, tuomas.koskinen@vtt.fi
2 Aalto University, Finland, iikka.virkkunen@aalto.fi

Abstract

The determination of the reliability of an inspection is of high significance. In particular, it is important to determine what is the largest crack, which could conceivably be missed during the in-service inspection. This information is utilized in order to choose the most effective method for different situations. Probability of detection (POD) curves are used to quantify the inspection effectiveness. However, major obstruction for POD curves is the requirement for a lot of data points in order to give reliable estimates of the lower limit performance, rendering reliable POD curves highly expensive to produce.

In this study, POD curve is estimated using only few thermal fatigue cracks, which is insufficient to produce a POD curve alone. In the present 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 and in addition CIVA simulation is used to produce amplitude response data from similar simulated cracks. Both amplitude responses are then in turn converted to a B-scan image for inspectors to evaluate whether there is a crack or not. Then a POD curve is generated from the achieved hit/miss data. The idea is to decrease the amount of needed real flaws and also to determine if an inspector can tell the difference between real, emulated and a simulated flaw.

1. Introduction

Probability of detection (POD) describes the ability of an 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 â, or a binary hit/miss response from the inspection is required. The detectable targets need to have measurable characteristics such as size. 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) are not necessarily exactly the same for every crack of the exact same size. This is due to the fact that there are other parameters 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 there is a human factor that may affect the end result of the inspection. Since in ultrasonic testing, eddy current testing and model assisted POD (MAPOD) response is a quantitative value, typical approach is to use signal response â vs. a instead of hit/miss analysis. Whereas hit/miss is more often used when the technique used do not produce a quantitative signal

such as penetrant testing. In theory â vs a contains more information than the binary hit/miss data, since â carries information about signal strenght. This correlates directly to the amount of flaws needed for the reliable POD analysis. For â vs a the minimum requirement is 40 different flaws and for hit/miss this requirement is 60. However, in order to reach more precise result 120 or even more different flaws is required for hit/miss evaluation. (1,2)

Although binary hit/miss may seem more restricted and straightforward compared to â vs a, there is a lot more to it. When Virkkunen et al. compared the â vs a and hit/miss results from the same data it was evident that these methods provide significantly different results when inspector insight is considered. This is due to the fact that inspector can readily check the suspected flaw again until he reaches the conclusion whether the indication is a flaw or just noise. Inspector can also use his insight to learn how the signal differs or stands out from noise. In low noise cases amplitude response is a straightforward way to determine a found flaw, but for example in austenitic welds noise raises the difficulty of the inspection significantly, thus straightforward signal threshold gives an unreliable POD result.(3)

Therefore, the main hindrance of a POD curve is its requirement of vast amount of different flaws. This leads to large number of artificial flaws and test specimens and eventually high costs. MAPOD offers a way of cutting costs by modelling majority of the flaws. Although, when using MAPOD the model is always a simplification of the real scenario, how to find correct variation between the flaws, how to take account the noise from the target and selecting proper thresholds for â vs. a analysis. Lastly, it may be difficult to actually prove that the simulations are correct and represent the real case for the most part and as already stated above, high noise situations cause problems with â vs a and MAPOD cannot be used to calculate a reliable POD. Previous study where a possibility to alter signal amplitude in reference to flaw size and produce a hit/miss POD with just three original flaws gave promising results (4). In this paper simulated flaws are generated based on these previous flaws scanned with ultrasonic phased array and simulated flaw signal altered the same way. This gives the possibility to exploit the inspector insight when evaluating an ultrasonic image, thus combining simulated signal and hit/miss POD with human inspector.

1.1 Influence of flaw properties

To measure the size of cracks with ultrasound accurately has been proven difficult. Cracks have a variety of characteristics which include location, orientation, size, opening of a crack tip, residual stresses and fracture surface roughness among others. These have an effect on the propagation, reflection, diffraction, transmission, attenuation and diffusion of ultrasound which may hinder the detectability of a flaw. (5)

In natural flaws the surface is never exactly planar caused by irregularities, which is why the production of artificial flaws which resemble real flaws is difficult. The flaw surface roughness has a significant effect on the detectability of a flaw. When the surface is rough, it decreases the scattered amplitude. As a smooth surface reflects a strong coherent field, a rough surface destroys the summation of the waves as the phase varies with the flaw surface. (5)

The surface roughness of a flaw may influence positively to flaw detection in some situations. Due to a rough surface there are corners and smooth sections in numerous directions. In a case of misorientation, the angle is not that effective as for a smooth flaw surface. Unlike in a smooth flaw, where the pulse may be continuous the rough surface may cause scattering of the pulse. The rough surface of a flaw may also cause loss of diffracted pulses. Even though the flaw may be noticeable it is hard to determine its actual size. (5)

In cases where huge amount of different flaws with different variability is required, it is reasonable to use simulation to support the artificial flaws. However, certain amount of artificial or real flaws will always be required in order to verify the results from simulation. Generally cracks form with slight roughness, opening and deviation variations it is reasonable to model the flaw representing these properties as well.

Although the flaws in this paper do not yet represent the actual real flaws, the method is a step closer to a more realistic flaw and flaw signal output.

2. Experimental Procedure

Austenitic stainless steel pipe with thermal fatigue flaws were used as scanned flaws with 1.5 MHz phased array TRS technique. To save time, only 45° angle was chosen for simulation. The scanned flaws were located in the inner diameter of the pipe near HAZ.

CIVA simulation software was used to generate ultrasonic signal response data. The simulation file was modified with Matlab to generate various different flaws with different shapes and surface roughness. The flaws were generated from a mesh of triangles.

Multiple triangular flaws allow large variety of shapes and surface roughness's for the flaws. Semi-elliptical flaw was chosen due to its simplicity and variability to produce different sizes easily. The size of the three scanned thermal fatigue flaws varied from 14 mm to 24 mm in length and from 2.27 mm to 8.5 mm in height. These parameters were set for the simulated flaws with the addition of 0.05 to 0.2 mm variation in roughness. Figure 1 represents the simulation set-up.

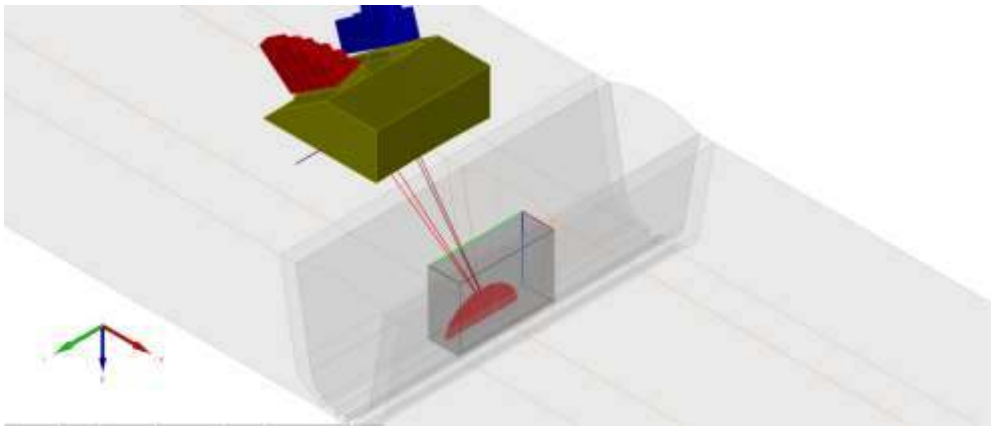

Figure 1. Simulation set-up with 8.5 mm height semi-elliptical flaw

B-scan image of the scanned flaws and the amplitude profile can be seen in figure 2. For the scanned flaws, maximum amplitude from the largest flaw h=8.5 mm was selected as a reference (0 dB). The smallest scanned flaw h=1.6 mm is almost undetected. This is due to the fact that the noise level is roughly -10 dB and the amplitude response is less than -6 dB from the reference. Due to anisotropy, preferred geometry or angle used in inspection, flaw with height of 4 mm gives larger amplitude response than the reference.

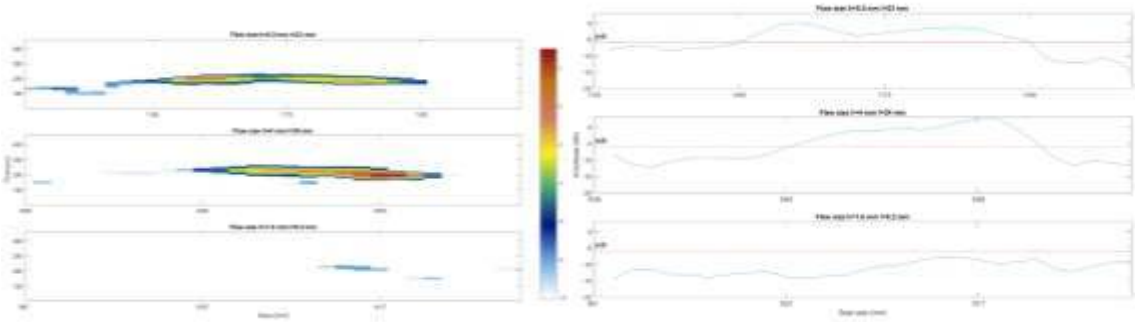

Figure 2. B-scan image of the thermal fatigue flaws

As figure 2 demonstrates with maximum amplitude profile of the scanned thermal fatigue flaws, the noise level within the austenitic weld is relatively high, around -10 dB from the maximum flaw signal.

B-scan image of the simulated flaws and the amplitude profile can be seen in figure 3. As in for the scanned flaws, the amplitude maximum of largest flaw of h=8.5 was selected as a reference. Similar to the scanned flaws, flaws with heights of 4 mm and 5 mm give higher amplitude response than the reference. In the simulation the flaw was set into the middle of the scanline. The highest surface roughness was in flaw 2.5x14.4 mm (average of 0.266 mm) and the lowest in flaws 4x24 mm and 8.5x23 mm (average of 0.1 mm).

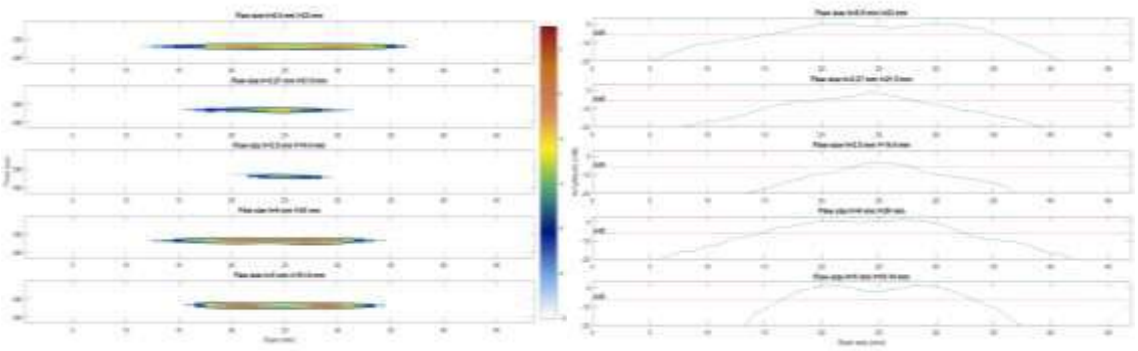

Figure 3. B-scan image of the thermal fatigue flaws

The simulated flaws seem clear and visible in figure above. This is due to the complete lack of noise from the austenitic weld. The noise was left out from the simulation, since the scanned data and where the simulated flaws were to be implemented was originally noisy already.

Flaw signal is the most significant feature from detection point of view and how it stands out from the noise, also the amplitude strength from the flaw is in correlation with the height of the flaw. Since there were only three original thermal fatigue flaws and five

simulated flaws, the flaw signal was altered to represent equivalent crack sizes different from the original flaws.

The equivalent crack sizes were obtained using the following formulae:

$$A_{shown} = A_{noise} + F(A_{true} - A_{noise}) \quad (1)$$
$$a_{eq} = a_{noise} + F(a_{true} - a_{noise}) \quad (2)$$

Where A_{shown} is the amplitude of the signal shown to the inspector, A_{noise} is the average noise level of the manually selected reference area F is the scale factor applied (0.0-1.0), A_{true} is the amplitude of the true signal. For equation 2, a_{eq} is the equivalent crack depth, a_{noise} is the crack size which corresponds to the average noise level and a_{true} equals to the true crack size. Using these formulae the original flaws are scaled down if F<1 and the F=0 stands no change in the data. Both formula are obtained by linear interpolation between the theoretical crack size with amplitude at noise level and actual flaw size from the crack. Although this is a crude approximation, the method can be improved by adding more realistic flaws.(4)

The modified flaw amplitudes were implemented in a web app for inspectors to determine if they can find a flaw from the image or not. The images showed modified thermal fatigue flaws and simulated flaws along each other and there could be more than one or no flaws at all in one image. The inspectors reviewed overall of 150 images for each session, allowing an inspector to evaluate whether he could find a flaw from the noise. At the end of the process, a hit/miss POD curve was shown. Afterwards this data was analysed more for this paper.

3. Results and Discussion

In the web app, the flaws merge fairly well within the original scan. On cases where the flaw represents a large flaw height the signal response is also quite high, whereas on smaller and lower height of flaws the signal response is lower. When inspectors were shortly interviewed after the test they stated that they were unable to tell on most cases which flaws were simulated, original or modified. One inspector stated that some patterns tended to repeat themselves. On situations where a wide flaw was made to represent a lower height flaw with signal representing almost noise level, which for example in â vs a analysis would have been a clear miss the pattern of a flaw could be distinguished from the noise. However, this is not yet clear weather does this represent a real situation where human eye is able to detect a flaw pattern from the noise or has this similar pattern repeated itself numerous times that the eye has learned to notice this kind of shape. The latter in this case would indicate that more of different shapes of flaws are needed and some way to implement noise partly over the flaw itself. Representation of this can be seen in figure below. This observation came from the inspector A, who also took part in simulating the flaws, his results were not taken in to the combination calculations in order to avoid contaminating the results. The test matrix can be seen in table 2.

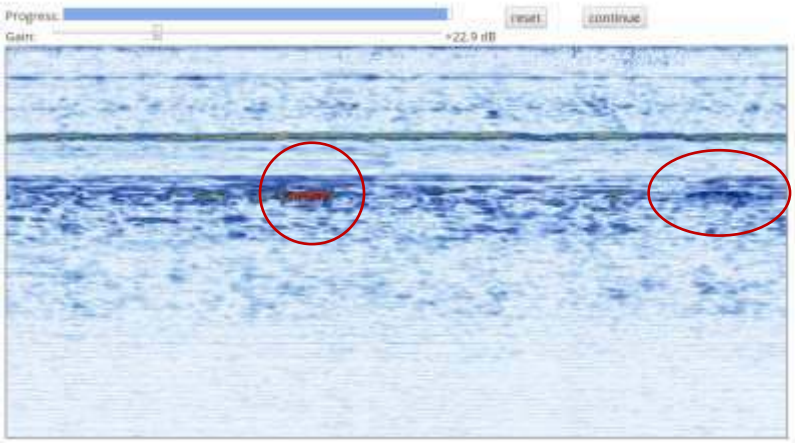

Figure 4. Shape of the flaw can be distinguished even though amplitude response is close to noise

Table 2. Inspectors and number of tries, cracks and images

Inspector	Number of tries	Number of cracks per inspector
~~A~~	~~6~~	~~~1080~~
B	2	~360
C	1	~180
D	1	~180
Overall*	4	~720

***Inspector A left out from combined POD calculations due to risk of "contamination".**

The results were plotted with as hit/miss POD vs. crack size in amplitude and POD vs. crack size in mm. In POD vs. crack size in amplitude the response was directly amplitude response and relation to crack size in mm was not taken into account. Even when POD vs. amplitude was plotted with hit/miss POD differences between the real flaws and simulated flaws could be noticed. Figure 5 shows PODs from two different inspection tries and separate PODs from simulated flaws and thermal fatigue flaws.

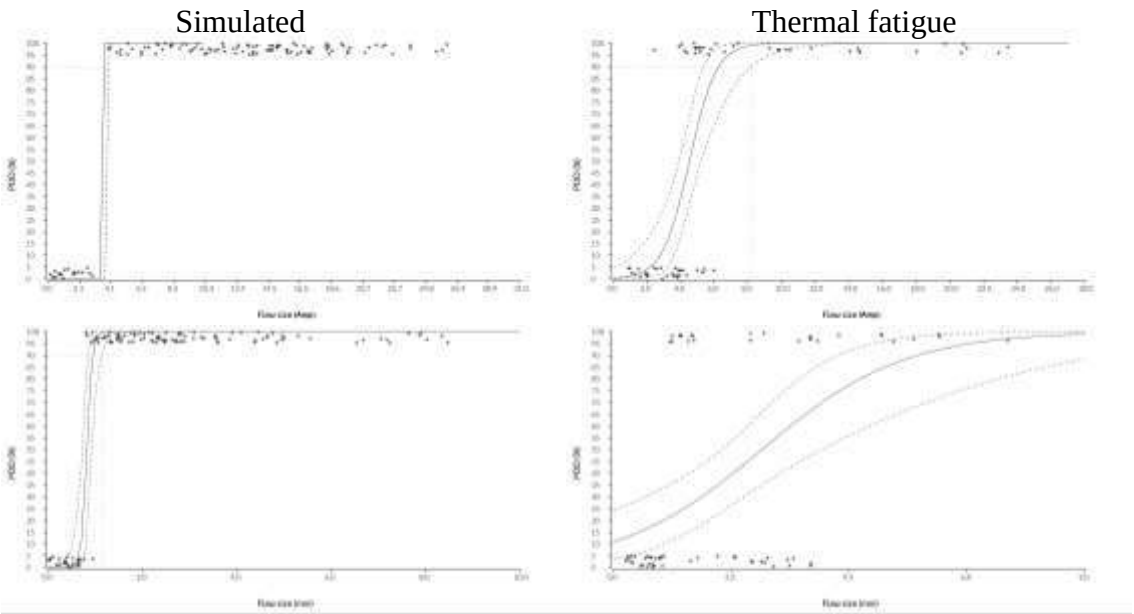

Figure 5. POD vs. amplitude curve (above) POD vs. crack length (below) from a single inspection try. Simulated flaws separated on the left and thermal fatigue flaws on the right.

In figure 5 PODs above are from one inspection try of inspector A, the POD calculated is plotted against amplitude response only. Simulated flaws give a clear amplitude threshold wtih flaw no-flaw situation, whereas thermal fatigue flaws give more variation. On other inspection tries, the threshold was not as absolute, but clearly the variation was less than for the reference flaws. Since thermal fatigue flaws and also simulated flaws tend to show variation regarding to amplitude and detection, the amplitude is not the only major factor influencing flaw detection. Thus, â vs. a analysis might give too narrow results.

In figure 5 PODs below are one inspection try from inspector C and also the crack length in mm has taken into account. The variation is much larger when crack length and amplitude correlation has been taken into account. The same effect applies for simulated flaws, however the effect is much smaller than for the thermal fatigue flaws.

In order to compare the simulated flaws and the thermal fatigue flaws, the inspection data gathered from inspectors B-D were combined and the two flaw types separated for their own hit/miss POD vs. crack size in mm in figure 6.

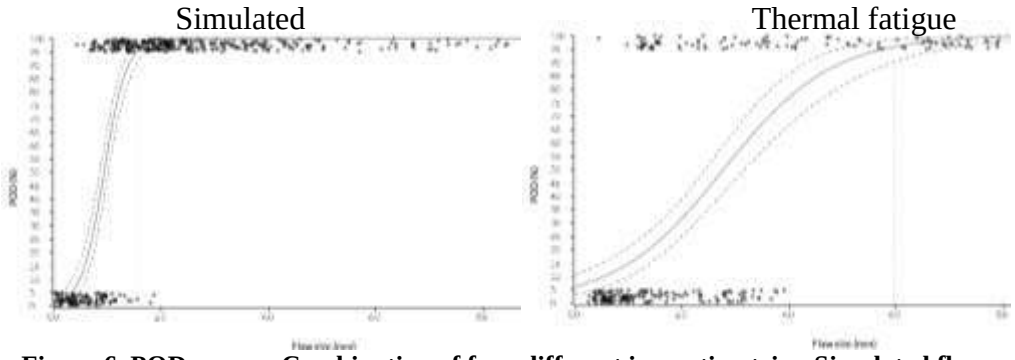

Figure 6. POD vs mm. Combination of four different inspection tries. Simulated flaws on the left and thermal fatigue results on the right.

Since there were more simulated flaws than real flaws, the confidence bound is larger on the real flaws as the same for figure 5. The results with simulated flaws gives a highly optimistic result compared to the real flaws. This might be because the simulated flaws do not represent the real flaws well enough.

4. Conclusions

During the inspection tries simulated flaws proved to blend well along with thermal fatigue flaws, and inspectors did not see any clear difference between different flaws. However, when separate hit/miss POD curves from the same inspection try was calculated. Simulated flaws gave highly optimistic result compared to the thermal fatigue flaws. Especially when the POD was plotted against crack size and amplitude correlation thermal fatigue flaws had larger variety than the simulated flaws.

Simulated flaws were found clearly more easily than then thermal fatigue flaws. There are multiple causes for the simulated flaws standing out more easily from the reference. Since the simulated flaw did not have any noise implemented to its signal like the

reference flaws, it might stand out as “clean”, thus standing out from the noise easily when the signal to noise ratio is near 1. Other reason might be that the simulated flaw does not represent the real flaw and inspection situation close enough. This seems also reasonable, since for this study the flaw was a rough approximation of the thermal fatigue flaws. In addition to surface roughness, factors affecting detectability are branching of the flaw, opening and macroscopic shape. Microstructure might have its own effect on detectability as well. Also the variation from POD vs. amplitude indicates that â vs. a analysis in noisy inspections might give unreliable results, since the detection has also other major influencing factors along with the amplitude response.

Since simulated flaws were more easy to detect compared to the thermal fatigue flaws and variation of amplitude response from flaws, it indicates that MAPOD might give overly optimistic results in high noise situations like the austenitic welds. It is also important to notice that these results were obtained with just three different thermal fatigue flaws and five simulated flaws. Therefore, a more thorough study with sufficient amount of thermal fatigue and more representative simulated flaws should be conducted.

Acknowledgements

This project was part of SAFIR2018, The Finnish Research Programme on Nuclear Power Plant Safety 2015 - 2018.
Contributions from Trueflaw Ltd. is greatly appreciated.

References

1. MIL-HDBK-1823A, 2009.
2. J.A. Ogilvy, Wave scattering from rough surfaces, Reports on Progress in Physics. Vol. 50 1987 , pp. 1553–1608.
3. I. Virkkunen, Comparison of â vs. a and hit/miss POD-estimation methods, 2018.
4. T. Koskinen, I. Virkkunen, S. Papula, T. Sarikka, J. Haapalainen, Producing a POD curve with emulated signal response data, Insight - Non-Destructive Testing and Condition Monitoring. Vol. 60 2018.
5. M. Kemppainen, I. Virkkunen, Crack characteristics and their importance to NDE, Journal of Nondestructive Evaluation. Vol. 30 2011 , pp. 143–157.