18th World Conference on Nondestructive Testing, 16-20 April 2012, Durban, South Africa

Probability of Detection (POD) determination using ultrasound phased array for considering NDT in probabilistic damage assessments

Jochen H. KURZ[1], Anne JÜNGERT[2], Sandra DUGAN[2], Gerd DOBMANN[1]

[1] Fraunhofer Institute for Nondestructive Testing (IZFP), Campus E3 1, 66123 Saarbrücken, Germany,
Phone: +49 681 9302 3880, Fax: +49 681 9302 11 3880,
email: jochen.kurz@izfp.fraunhofer.de, gerd.dobmann@izfp.fraunhofer.de
[2] Materials Testing Institute University of Stuttgart, Pfaffenwaldring 32, 70569 Stuttgart, Germany,
Phone: +49 711 685 66787, Fax: +49 711 685 62761,
email: anne.juengert@mpa.uni-stuttgart.de, Sandra.dugan@mpa.uni-stuttgart.de

Abstract
NDT is an essential part during construction and maintenance of nuclear power plants. In Germany, type and extent of inspection are specified in Safety Standards of the Nuclear Safety Standards Commission (KTA). Only certified inspections are allowed in the nuclear industry. The qualification of NDT is carried out in form of performance demonstrations of the inspection teams and the equipment, witnessed by an authorized inspector. The results of these tests are mainly statements regarding the detection capabilities of certain artificial flaws. In other countries, e.g. the U.S., additional blind tests on test blocks with hidden and unknown flaws may be required, in which a certain percentage of these flaws has to be detected. The knowledge of the probability of detection (POD) curves of specific flaws in specific testing conditions using defined inspection methods is often not present. This contribution will present the results of a research project designed for POD determination of ultrasound phased array inspections of real cracks. The continuative objective of this project is to generate quantitative POD results and to integrate these results in a probabilistic damage assessment concept. Different POD applications will be shown. The distribution of the crack sizes of the specimens, the inspection planning and the designed inspection instruction will be discussed, and results of the ultrasound inspections will be presented. Furthermore, additional considerations for POD determination of phased array inspections will be discussed. In the context of the results, the remaining uncertainty of the inspections has to be taken into consideration for failure analysis.

Keywords: Probability of detection (POD), ultrasound phased array, Ultrasound sampling phased array, cracks, nuclear safety program

1. Introduction

Non-destructive testing (NDT) is an essential part during construction and maintenance of nuclear power plants. In Germany, type and extent of inspection are specified in Safety Standards of the Nuclear Safety Standards Commission (KTA). Only certified inspections are allowed in the nuclear industry. The qualification of NDT is carried out in form of performance demonstrations of the inspection teams and the equipment, witnessed by an authorized inspector. The results of these tests are mainly statements regarding the detection capabilities of certain artificial flaws. In other countries, e.g. the U.S., additional blind tests on test blocks with hidden and unknown flaws may be required, in which a certain percentage of these flaws has to be detected. The determination of the probability of detection (POD) is one possibility to quantify the probability to detect a specific flaw with an NDT method. Publications dealing with POD applications can be found in different fields (medical sciences, biology, materials science, physics, engineering sciences, telecommunication etc.) [1-6]. From which of these fields the POD concept originated cannot be determined with certainty. However, some of the earliest studies of a variety of aspects concerning POD can be found in the development of radar technology during World War II. In the field of NDT, the number of publications dealing with the topic POD has increased significantly during the last few years (Fig. 1). This fact indicates that probability of detection concepts are of increasing importance when NDT methods are applied. In the U.S., new ASTM standards concerning NDT require expert consensus on the uncertainty of the applied method.

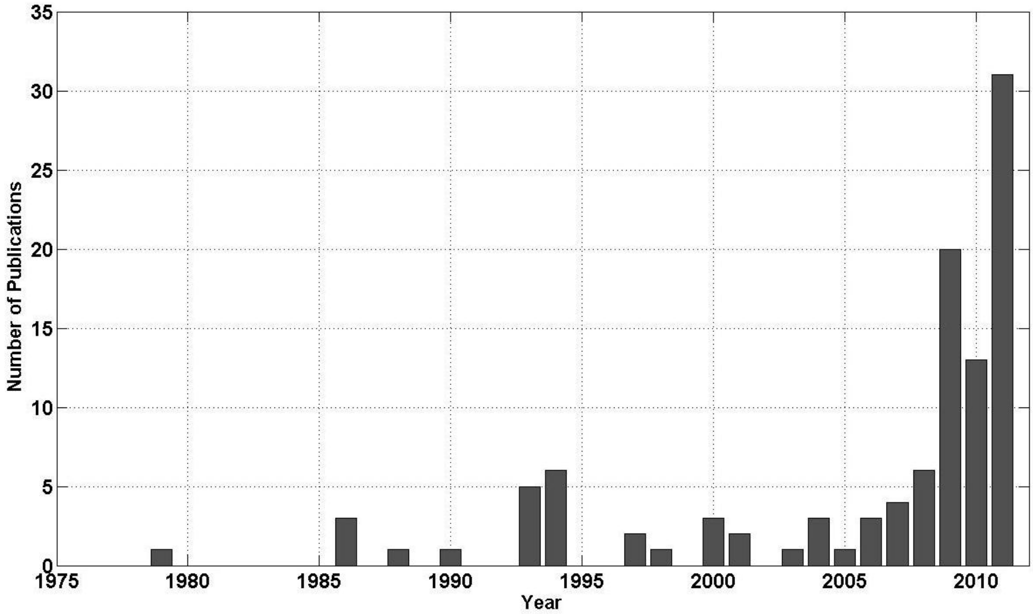

Figure 1. Number of publications on POD in the field of NDT.

The POD is strongly connected to the topic of risk assessment and probabilistic analyses in the assessment of the integrity of components. The probability of failure (POF) is calculated based on statistical data for relevant parameters. One parameter that influences POF calculations is the probability of detection (POD) of a flaw of a certain size. Reliable assessment of the integrity of components is an important requirement for reliable and safe operation of nuclear power plants and other installations relevant to public safety.
In this paper, the detection of real and artificial flaws of different sizes and different types using ultrasound is considered. The POD provides the probability for the detection of a certain flaw size. This does not exclude the occurrence of an isolated event that a larger crack is not detected. The POD delivers the realistic, statistical assessment of the reliability for an NDT method. The knowledge of a POD does not increase or decrease the occurrence of such an exception. However, a realistic safety oriented design and assessment is possible since risks can be evaluated. The knowledge of the probability of detection of a certain flaw allows to assess the consequences of this flaw in a probabilistic manner.
The experimental POD determination, the design of the inspection tests and the procedure described in this paper were carried out within the scope of the research project "POD - Consideration of NDT inspections for probabilistic failure analysis", funded by the German Federal Ministry of Economy and Technology (BMWi) as part of the German reactor safety program. With the objective of POD determination, ultrasonic phased array [7], sampling phased array [8], and fixed angle ultrasound [7] inspections were carried out. Due to the increasing number of applications of the phased array technique, the focus was put on the phased array inspection. The following section describes the selected and manufactured test specimens, the inspection procedure developed for the phased array investigations, and the inspections carried out within the scope of the project.

2. Specimens, procedure and tests

Within the scope of the research project, the Material Testing Institute (MPA) at the University of Stuttgart is responsible for manufacturing and supplying suitable test specimens for the ultrasonic investigations which are the basis for the calculation of the POD curves. The MPA holds a wide variety of test specimens and has broad experience with the creation of test specimen with realistic flaws. In addition to the already existing test specimens, several new test specimens were manufactured. The selection of suitable test specimens is essential, because the flaws should be as realistic as possible and not limited to artificial flaws. Therefore mainly test specimens with realistic flaws were chosen. The test specimens are listed in table 1.

Table 1. Test specimens for the ultrasonic investigations

	Number of test specimen	Total number of flaws	Realistic flaws	artificial flaws
Austenitic test specimen	29	43	24	19
Ferritic test specimen	2	5	5	-
Dissimilar welds	1	15	2	13
Cladded specimen	4	> 65	> 59	6
Total	36	> 128	> 90	38

Different materials were chosen to cover different materials commonly used in nuclear facilities.
A picture of three austenitic stainless steel test specimens is shown in figure 2. The test specimens are bar-shaped with a “wall thickness” of 32 mm, and contain different types of welds (different weld geometry). Some of the specimens are only base metal without weld. Most of the flaws are realistic flaws induced by inter-granular stress corrosion cracking or fatigue cracking [9].

Figure 2. Austenitic test specimens with and without weld.

Some of the test specimens contain EDM notches and some of them are completely intact. The flaws were documented using different reference tests like surface crack testing, radiography and creation of microsections.
The ferritic test specimens are made from ferritic pipe sections with an outer diameter of 275 mm and a wall thickness of 14.5 mm, and contain different welding flaws such as porosity and lack of fusion. The flaws were documented using radiography. To determine the through-wall extension of the flaws, an additional CT will be necessary.
The test specimen with a dissimilar metal consists of a circumferential pipe weld connecting an austenitic stainless steel pipe with a ferritic low alloy steel pipe, with austenitic cladding on the inner surface of the ferritic pipe section [10]. The test block has an outerdiameter of 327 mm, a wall thickness of 29.5 mm, and a total length of 435 mm. A cross section of the weld is shown in figure 3. The test block contains several EDM notches in the weld and in the base metal, as well as one axial and one circumferential crack.

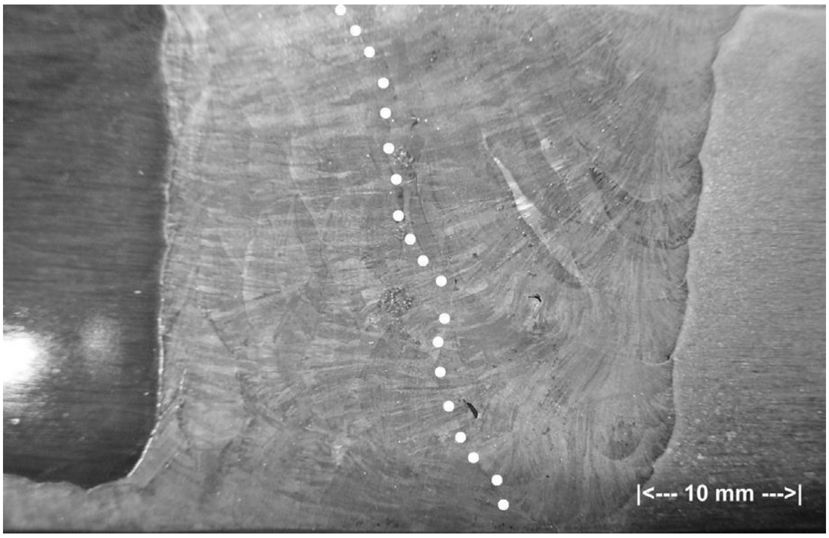

Figure 3. Cross section of the dissimilar weld.

The cladded test specimens consist of thick ferritic plates with an austenitic cladding, originally made to serve as pressure vessel test specimens, with wall thicknesses between 95 and 146 mm. The test specimens contain flaws that were created before, after and during the welding process [11]. A great number of flaws are natural unterclad cracks, which were created by applying a wrong heat treatment after the welding process. This type of crack forms as inter-granular in the heat affected zone. The reference tests for the plated test specimens are still going on. Due to the great wall thicknesses, the detection capability of radiography is limited.
Altogether there are more than 128 flaws in the test specimens; more than 90 are realistic flaws. The number of flaws therefore is sufficient for the POD estimation. The distribution of flaw sizes is shown in figure 4.
It can be seen that the distribution is asymmetrical with the maximum at a flaw size of 3 mm. It is expected that smaller flaws are harder to detect so this distribution is suitable for the POD-estimation.
The major part of the ultrasonic testing took place at the MPA’s NDT lab in spring 2011. Three teams of ultrasonic inspectors using two different types of measurement equipment examined the test specimens, using ultrasonic phased array and ultrasonic sampling phased array techniques. The details for the inspection are specified in a test procedure. The recording limits specified were based on the Safety Standards of the German Nuclear Safety

Standards Commission (KTA). The results of each team were compared to the actual flaws within the test specimens, known from reference tests. These results were then used as input for the POD determination.

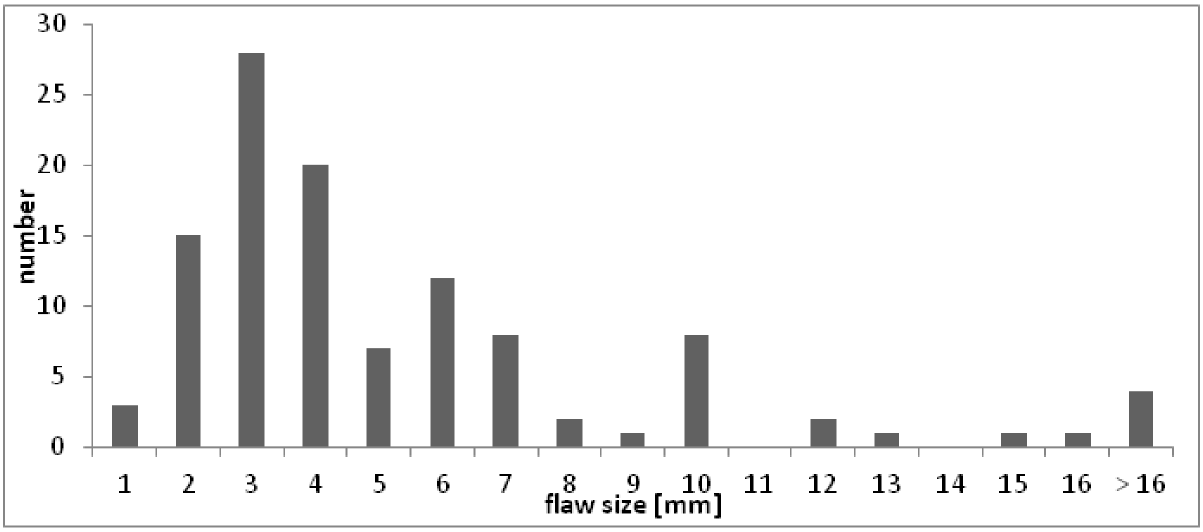

Figure 4. Distribution of flaw sizes.

3. POD determination

For an NDT inspection, a variety of factors can influence the inspection result, i.e. whether or not a flaw will be detected. In principle, two related approaches to a probabilistic framework for the inspection of reliability data can be used [12]. One is based on the analysis of binary data, i.e. whether or not a flaw was found (hit/miss) and the other one is based on the correlation of signal response and flaw size (â vs. a). Analyzing binary data requires maximum likelihood regression to get the cumulative distribution function representing the chosen POD model. However, as [12] points out, different results will be obtained when the two POD analysis methods are applied to the same data set.

The crucial decision in a hit/miss analysis is to define a hit/miss criterion. If a threshold value for the signal response is sufficient to characterize a hit or a miss event, the criterion can be easily defined. In our case, flaw size and location based on the evaluation of the registration length is the relevant information. Therefore, accuracy in the determination of size and location are the parameters which should be considered for POD determination in this case. Furthermore, the used maximum likelihood regression requires more data to achieve stable results [2, 13]. A hit/miss criterion was applied to the data, however, only in the form of a threshold criterion, i.e. defining a non-detected flaw as a miss without considering the accuracy in flaw size and location of the result. The results of the POD determination of the data analyzed as hit/miss data is shown in Fig. 5. The data was taken from the UT measurements at the 29 austenitic specimen acquired with a phased array transmitter and receiver (TR) probe.

The determined POD curve (Fig. 5) shows that a 90 % POD with a confidence interval of 95 % ($a_{90/95}$) for a flaw of size 15 mm was obtained. This result does not consider the accuracy in flaw size determination. Therefore, an additional “â vs. a” analysis was performed.

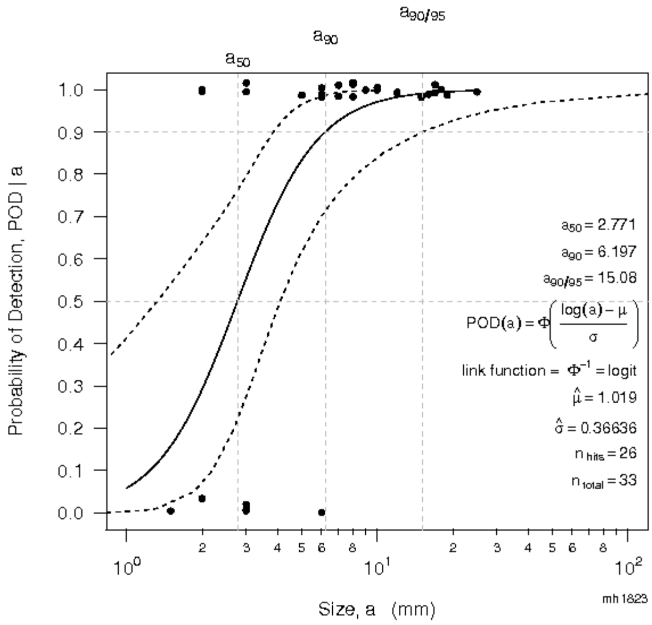

Figure 5. POD curve calculated from hit/miss data. For POD calculation the mh1823 software [14] was used.

POD determinations based on signal response data (â vs. a) have already been carried out for a variety of NDT methods [2, 3]. Therefore, the applicability of ultrasound data was already proved. In general, four conditions have to be fulfilled when measurement data are to be used for an "â vs. a" POD determination: a) linearity of the parameters, b) uniform variance, c) uncorrelated responses, and d) normal distribution of test flaw sizes. In our case, the determined flaw size represents the signal response data. Regarding an "â vs. a" POD analysis the four named conditions have to be valid for the data used. Fig. 6 (left) shows the linearity of the parameters and Fig. 6 (right) shows that the observations are uncorrelated. For this, data from ultrasonic inspection of 3 tilted notches of different size were taken (Fig. 6, right).

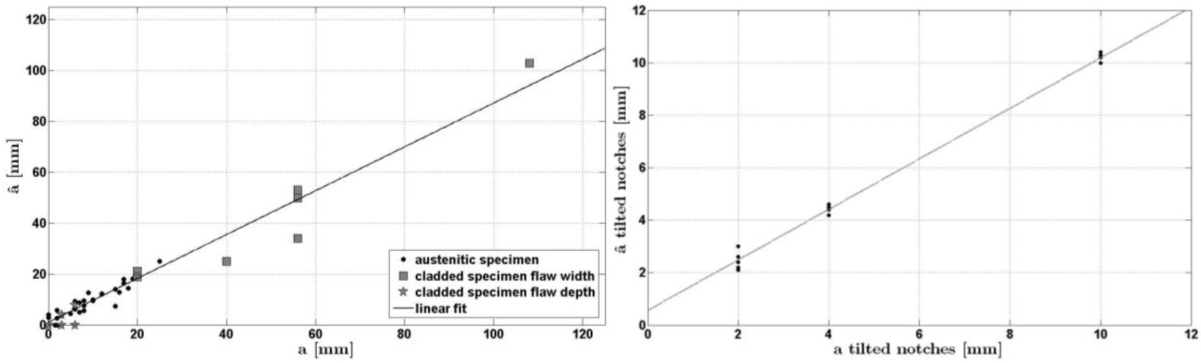

Figure 6. Left: linearity of the parameters. Right: uncorrelated signal response.

The uniform variance of the signal response was proven by determining the variance of a separate test where 3 different cracks were inspected 50 times and the data was analyzed. For these data the standard deviation was determined and the values were: 0.03, 0.03, and 0.04. The criterion that the data has a multivariate normal distribution is shown in form of a quantile-quantile distribution plot (Fig. 7). The linear relation shown in Fig. 7 indicates that the data values belong to a normal distribution with different mean value or different scattering as compared to the reference normal distribution [15]. Even if this plot alone is not a biunique proof of normal errors, it is a strong indication. Together with the other fulfilled conditions, it is now obvious that the collected data which is based on the analysis of the registering length can be used for an "â vs. a" POD analysis.

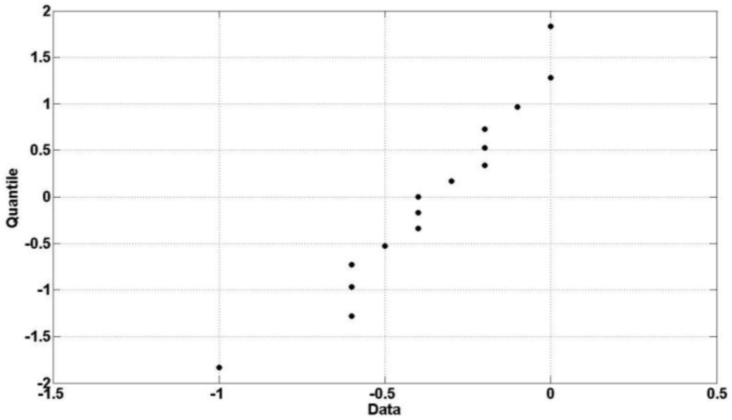

Figure 7. Quantile-quantile distribution plot of the data from Fig. 6 right.

A critical point in the POD analysis of signal response data is the definition of the decision threshold [2, 16]. In an analysis in which the amplitude value is used, the noise can be considered for selecting the decision threshold at any point above the noise level. In general, the decision threshold influences both the minimum size detected and the probability of false positive detections [2]. In our case, when the flaw size is the parameter used for a signal response based POD analysis, a criterion is required for the selection of a decision threshold based on the accuracy of the inspection. According to the approach of [16], a criterion was chosen which considers noise and the false positive detections in an indirect manner. The difference between true flaw size and determined flaw size was taken and the standard deviation was calculated. Fig. 8 shows the difference between true flaw size and determined flaw size taken from the measurements at the austenitic specimens. The calculated standard deviations for the two distributions are 3.2 mm and 2.5 mm. These values were taken as decision thresholds for the POD analysis. The POD curves calculated with the two decision thresholds are shown in Fig. 9.

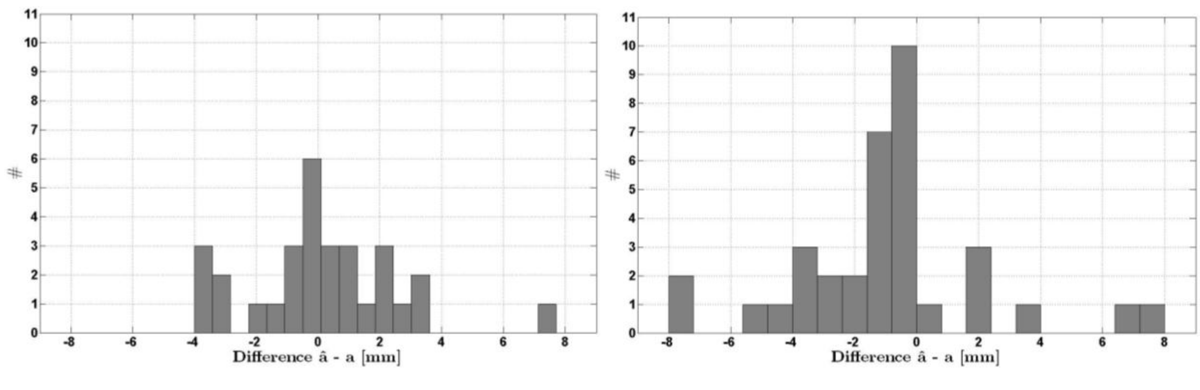

Figure 8. Differences between true flaw sizes and determined flaw sizes obtained from the measurements on the austenitic specimens. Left: RT phased array probe. Right: sampling phased array technique.

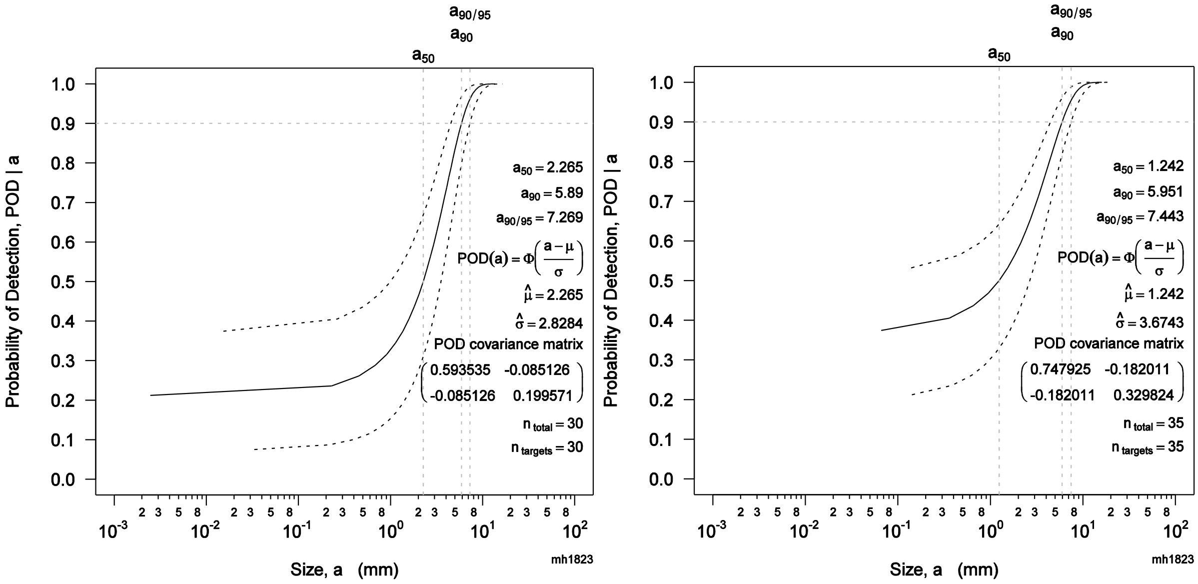

Figure 9. Left: POD curve from RT probe phased array data. Right: POD curve from SPA data. The POD calculation was carried out with the mh1823 software [14].

The decision threshold values for the phased array (RT probe) and the SPA data do not differ much. The deviations of true flaw size from determined flaw size represented in Fig. 8 show slightly different distributions of the values. The corresponding calculated POD curves (Fig. 9) show no significant differences in the $a_{90/95}$ values (7.2 mm vs. 7.4 mm). However, in the a_{50} value a difference of 1 mm is obtained.

4. Conclusions

The determination of the probability of detection (POD) curves is one possibility to quantify the probability to detect a specific flaw with an NDT method from a statistical distribution. NDT is an essential part during construction and maintenance of nuclear power plants. However, knowledge of the probability of detection (POD) curves of specific flaws in specific testing conditions using defined inspection methods is often not present.

For the described work on POD determination, test blocks with a large number of real flaws and artificial flaws were inspected by different teams with ultrasonic phased array probes and with the sampling phased array technique. Four cladded thick-walled ferritic test specimens, 29 austenitic specimens with and without welds, and one specimen containing a dissimilar weld were inspected by three different teams. Since no German standard is currently available for phased array inspections, the specifications for the tests were outlined in a detailed inspection procedure. The results of two teams obtained from the austenitic specimen were used for the POD determination presented in this paper. A direct use of the signal amplitude as a measure of flaw size is not possible when using focusing probes and reconstructed data. For such inspection techniques, the registration lengths of the flaws obtained from the data form the basis for the POD analysis. This is also has to be considered for other NDT devices in the future when more and more inspection results will be based on the analysis of images instead of using amplitude values. The results of the phased array and sampling phased array inspections provide flaw sizes and locations. A simple hit/miss POD analysis can be performed. However, a valid and objective criterion for defining hit and miss for the described type of results has not been determined yet. Regarding the “â vs. a” POD analysis, a decision threshold was defined when the flaw size is used as the signal response parameter. The chosen approach, which uses the standard deviation of the difference between true flaw size and determined flaw size, also allows to consider false positive detections. The results of the "â vs. a" POD analysis differ significantly from the simple hit/miss analysis which was performed. The $a_{90/95}$ values of the "â vs. a" analysis are 7.2 mm (phased array RT probe) and 7.4 mm (SPA). The slope of the POD curve obtained from the phased array data is greater and therefore the $a_{90/95}$ value is slightly smaller.

Acknowledgements

The financial support of the project „Einbeziehung der Aussagefähigkeit zerstörungsfreier Prüfungen in probabilistische Versagensanalysen", Förderkennzeichen 1501386 by the German Federal Ministry of Economy and Technology (BMWi) is gratefully acknowledged.

Supported by:

Federal Ministry
of Economics
and Technology

on the basis of a decision
by the German Bundestag

References

1. Schoefs, F., Clement, A., Nouy, A., 2009. Assessment of ROC curves for inspection of random fields. Structural Safety, Vol. 31, 409-419.
2. Annis, C., 2009. MIL-HDBK-1823A. Nondestructive Evaluation System Reliability Assessment. Department of Defense Handbook, Wright-Patterson AFB, USA.
3. Georgiou, G.A., 2006. Probability of Detection (PoD) curves. Derivation, application and limitations, Jacobi Consulting Limited, Research Report 454, 2006.
4. Kharin, V. V., Zwiers, F. W.: On the ROC Score of Probability Forecasts. Journal of Climate, Vol. 16, 4145-4150.
5. Sabelnikov, A., Zhukov, V., Kempf, R., 2006. Probability of real-time detection versus probability of infection for aerosolized biowarfare agents: A model study. Biosensors and Bioelectronics, Vol. 21 (11), 2070-2077.
6. Alvseike, O., Skjerve, E., 2000. Probability of detection of Salmonella using different analytical procedures, with emphasis on subspecies diarizonae serovar 61:k:1,5,(7) [S. IIIb 61:k:1,5,(7)]. International Journal of Food Microbiology, Vol. 58 (1-2), 49-58.
7. Moore, P. O. (Ed.), Workman, G. L. (Ed.), Kishoni, D. (Ed.), 2007. Ultrasonic Testing. Nondestructive Testing Handbook 7, 3rd ed. American Society for Nondestructive Testing (ASNT), (Nondestructive Testing Handbook; 7). - ISBN 978-1-57117-105-4.
8. Bulavinov, A., Joneit, D., Kröning. M., Bernus, L., Dalichow, M.H., Reddy, K.M., 2006. Sampling Phased Array A New Technique for Signal Processing and Ultrasonic Imaging. Insight. Vol. 48, No. 9, pp. 545-549.
9. Mletzko, U., Zickler, S., 2007. Verbesserte Bewertung der Aussagefähigkeit von zerstörungsfreien Prüfungen an Komponenten aus Nickelbasislegierungen oder austenitischen Werkstoffen. BMU Vorhaben SR 2501, Technischer Bericht 1.2, Stuttgart.
10. Mletzko, U., Zickler, S., 2007. Erste Bewertung der Aussagesicherheit von zerstörungsfreien Prüfungen an Mischschweißverbindungen auf Querrisse. BMU Vorhaben SR 2501, Technischer Bericht 1.1, Stuttgart.

11. Waidele, H. et. al., 2003. Bewertung der Aussagefähigkeit von Ultraschall- und Wirbelstromprüfung austenitischer Plattierungen von Reaktordruckbehältern. BMU Vorhaben SR2318, 2. Technischer Bericht 8892 01 001, Stuttgart.
12. Berens, A. P., 1994. NDE Reliability Data Analysis. In: ASM Handbook, Vol 17, Nondestructive Evaluation and Quality Control, 3rd Edition, ASM International USA, pp. 689-701.
13. Kurz, J. H., Jüngert, A., Dugan, S., Dobmann, G., Roos, E., 2011. Experimentelle POD Bestimmung mittels Ultraschall Phased Array zur Einbeziehung zerstörungsfreier Prüfungen in probabilistischen Versagensanalysen. In: Verbesserung der Prüfaussage für spezielle Prüfaufgaben, Seminar des FA Ultraschallprüfung. Berlin, Deutsche Gesellschaft für zerstörungsfreie Prüfung (DGZfP), (DGZfP-Berichtsbände 132-CD).
14. Annis, C., 2010. Statistical best-practices for building Probability of Detection (POD) models. R package mh1823, Version 2.5.4.4 http://StatisticalEngineering.com/mh1823/.
15 Rinne, H.1997. Taschenbuch der Statistik. 2nd Edition, Verlag Harri Deutsch, Frankfurt am Main, p. 650.
16. Feistkorn, S., 2011. Gütebewertung qualitativer Prüfaufgaben in der zerstörungsfreien Prüfung im Bauwesen am Beispiel des Impulsradars. PhD Thesis, TU Berlin.