Simulation-Based Investigation of a Probability of Detection (POD) Model Using Phased Array Ultrasonic Testing (PAUT) Technique

N.B. When citing this work, cite the original published paper.

(article starts on next page)

Simulation-Based Investigation of a Probability of Detection (POD) Model Using Phased Array Ultrasonic Testing (PAUT) Technique

Xiangyu Lei[1] · Håkan Wirdelius[2] · Anders Rosell[3]

Abstract

Probability of detection (POD) as a metric for quantifying the capability of inspection procedures in nondestructive evaluation (NDE), has been applied and evolved in industries since 1970s. Progress had been noted when certain statistical functions were brought up to model POD behavior, including log-normal model (also referred as Probit model). This model had been concluded to be the best fit and therefore has been widely used in many studies, while the involved assumptions and conditions were not carefully addressed and explored. To make flexible POD datasets available for specific inspection procedures and reduce the number of expensive experiments needed, model-assisted POD (MAPOD) is an alternative. This paper addresses a pure simulation-based POD procedure of an inspection scenario involving phased array ultrasonic testing (PAUT) on lack-of-fusion defects in additive manufactured (AM) components. The mathematical simulations are performed by an ultrasonic testing (UT) simulation software, simSUNDT, developed at Chalmers University of Technology in Sweden. Resulted inspection datasets with the proposed data processing steps are evaluated in terms of the assumptions and conditions of log-normal POD model, with the purpose of discussing the POD model validity under different circumstances. Simulation-based POD curves are finally compared with several discrete POD values at some defect sizes, calculated through massive computations from physics-model based metamodel. Comparisons and observations confirm satisfactory application of log-normal POD model despite some violations in model hypotheses.

Keywords Nondestructive evaluation (NDE) · Phased array ultrasonic testing (PAUT) · MAPOD · Numerical simulation · simSUNDT

1 Introduction

Higher demands on the reliability of components used in the industry increase the needs for quality assessment in the manufacturing industries. It is therefore crucial of monitoring and controlling of manufacturing, inspecting safety–critical assets and assuring of structural integrity. However, consistent results are not always necessarily guaranteed through repeated inspections on a specific defect due to the probabilistic characterization of inspection capability. This makes it much essential to quantify the performance and reliability of a nondestructive evaluation (NDE) procedure, especially when risk-based inspection (RBI) methodologies and component life-cycle estimations are introduced for in-service property management. Probability of detection (POD), as a statistic metric initiated since 1970s from aero industry [1, 2], has been a tool used for assessing the reliability of NDE procedures. It helps describe the accuracy of an inspection and reveal how well an inspection procedure can detect vital defects. Inspection data is in this method transformed into a relation between the probability of defects detection and a characteristic parameter of the defects [3], usually the defect size [4], as it is crucial for structure integrity and estimation of component life. A constant POD value was proposed at the beginning for all defect types of the

✉ Xiangyu Lei
xlei@chalmers.se

Håkan Wirdelius
hakan.wirdelius@hv.se

Anders Rosell
anders.ar.rosell@gknaerospace.com

[1] Department of Industrial and Materials Science (IMS), Chalmers University of Technology, 41296 Göteborg, Sweden

[2] Division of Subtractive and Additive Manufacturing, University West, 46186 Trollhättan, Sweden

[3] GKN Aerospace Sweden AB, Flygmotorvägen 1, 46181 Trollhättan, Sweden

same size, while an analysis of a study showed discrepancies in repeated inspections on a certain crack and on different cracks of the same size [5]. In fact, there are many influential factors that may introduce variations when determining the POD values, including test methods, equipment and material conditions, defect features as well as human factors. It is therefore important to understand the process of obtaining a POD value and to question the validity of its application and limitations. The experimental data required to determine the POD value can be collected from for example field service records, artificially created defects in components and generic test blocks [6]. Experiments are repeated in many runs following a controlled condition specified by procedure to record several results. These results can be categorized in hit/miss type if binary conclusions are drawn based on certain detection criteria, or in signal response type if result amplitudes are of interest. The interpretation of POD value for a defect size is then the proportion of times the defect being detected (hit) among all trials. A POD curve is approximated with best fit through all POD values among a range of defect sizes using statistical analysis models, e.g., log-odds and log-normal model [7], also referred as Probit model [6]. Obviously, the experiment processes can provide practical measures of the capability of inspection instruments and the actual environments. However, these practical processes are always expensive and resource consuming, and even more challenging when introducing representative features to the artificial defects. It might also end up in poor statistics and the results could show scatters [8]. Besides, the experiments tend to have bias because the trials are conducted with defective test pieces, which could be different from the reality [9]. Parametric studies for investigating influential parameters in procedures are also unfeasible by practical trials.

With the development of physics-based nondestructive testing (NDT) model in recent decades, mathematical simulations are widely applied in many related applications, among which assisting or replacing part of the experimental works in POD estimations is included [8, 10]. There is also a European project PICASSO exploring the possibility of building simulation-supported POD curves [11]. This model-assisted probability of detection (MAPOD) approach, realized by either transfer function [12] or using full model [13], has the advantages of being resource- and time-saving. It also provides possibilities of parametric studies of performance and reliability. Models can also be utilized as training tools for inspectors and be part of the qualification and validation processes in NDT applications. Human factor, the most influential parameter within the manual experimental processes, can to a large extent be isolated, but can also be studied and approximated by modelling approach [14, 15]. Along with the formation of MAPOD Working Group at Center for Nondestructive Evaluation of IOWA State University in 2005 [13, 16], there are more groups of researchers devoted

into MAPOD research and development. The National NDT Center of UK is an example, now as part of ESR Technology [17]. There has also been a group of researchers in Netherlands developing a so-called POD-generator [18], in which 3 individual modules are contained and combined to simulate the entire POD estimation process from defect initiation and growth to the prediction of failure probabilities. A commercial NDT simulation software, CIVA, has also been developed for MAPOD purposes with corresponding module [19, 20]. Another ultrasonic testing (UT) simulation software, simSUNDT, developed at Chalmers University of Technology in Sweden, has undergone an investigation regarding POD estimation in terms of conventional UT inspection according to a testing procedure UT-01 for pipes [21]. Its mathematic kernel, UTDefect [22–25], which performs the actual UT modeling and calculation, can be further applied independently in various parametric studies, such as sound field optimizations for phased array ultrasonic testing (PAUT) technique in terms of maximizing the received echo amplitude [26].

In this paper, the above-mentioned mathematic kernel for UT simulation, UTDefect, is used for generating UT inspection signal response data for pure simulation-based POD estimation. The studied inspection scenario is based on PAUT on lack-of-fusion defect, which is plausible in components produced by some additive manufacturing (AM) processes (e.g., laser metal deposition, LMD). The results are transformed into corresponding POD curve using the widely accepted log-normal POD model (Probit). The POD curve is then compared with a set of discrete POD value points obtained through a large number of metamodel-based simulations, to verify the validity of the POD model. In addition, a modified distance amplitude correction and time varied gain (DAC/TVG-mod) framework is considered afterwards to contribute to simulation results and POD curves for further investigation and comparison.

2 simSUNDT Software

The simSUNDT software consists of Windows®-based pre- and post-processor as well as a mathematic kernel, UTDefect, for the actual modeling and calculation, which has been experimentally validated [23–25, 27] to some extent by comparing with available experimental data. This mathematic kernel takes specified text file with input parameter values and generates output information in plain text format, so that external programs can call this kernel for automatic operations with certain cases. To model the probes and the interaction with defects (scattering), a series of integral transforms and integral equations are employed. Together with the possibility of calibration, i.e., towards a reference reflector including side-drilled hole (SDH) and flat-bottomed hole

(FBH), the software can simulate the entire testing procedure under different testing scenarios.

The model is fully three-dimensional, while the simulated component in the software at this stage is limited to be an infinite plate with finite or infinite thickness and bounded by the scanning surface. The probe is modelled as a boundary condition on the surface of an elastic half-space. This surface is traction free except the active part of the contact area beneath the probe. This enables the flexibility of probe simulation regarding its shape, wave type, element size, angles, etc. The receiver in the UT system is modelled by using a widely known and used reciprocity argument [28]. The available defect types are volumetric and crack-like defects, and the methods of solution towards the defect modeling is the T-matrix method [29], where the transition matrix includes all information regarding the defects.

For the phased array (PA) probe technique, each individual element follows the principle of probe modelling above. The individual boundary conditions are then translated into the main coordinate system, so that a PA wave front with certain beam angle is formulated by constructive phase interference. In order to enable larger beam angles, a model of a wedge is also included as an option. This PA probe model has been experimentally validated qualitatively [26] and quantitatively [30].

3 Simulation-Based Probability of Detection

POD data from physical experiments are usually retrieved through repeating many inspection trials for a set of defects under a controlled condition specified by corresponding procedure. The POD value for a defect size is interpreted as the proportion of times this defect is marked as detected among all trials on this defect. As said in Sect. 1, experimental trials can provide measures of the capability of used inspection instruments and reflect on inspection surroundings, but they are also expensive and challenging because large amounts of specimens containing a specific defect type are needed for computing valid statistical parameters of POD function. The number of defects recommended for hit/miss type of data is at least 60 and their sizes should be uniformly distributed between the minimum and maximum defect size of interest. At least 30 defects should be available for signal response type of data thanks to added information in response signal [7, 31]. This type of data is sometimes referred as "a" vs. "$\hat{a}$" data, where "a" stands for defect size and "$\hat{a}$" for quantified signal response, recorded in terms of a parameter. Yet, the statistics might still be poor from this limited number of available defects and shows scatters in the results.

Physics-based mathematical simulation models compared to physical experiments of inspection, have the superiority of reducing the massive cost and extensive operations. The number of defects is not a limitation by simulations, which in turn facilitates satisfaction of appropriate statistical sampling requirement in design of experiments (DOE).

Since observed from physical experiments of probabilistic nature of NDE processes, repeated inspections of a certain defect will not necessarily produce the same signal responses due to minor variations in some inspection parameters, e.g., experimental setup and calibration. Variations could also come from minor differences of some defect-features related inspection parameters despite that the defect sizes, the characteristic parameter of the defects, against which the POD curve will be plotted, are the same. One way of representing these variations in simulation models is to use the uncertainty propagation method [32], where the inspection parameters are specified within individual uncertainty ranges (parameter space) and corresponding distributions. A series of simulations are then performed in the parameter space and the resulted signal responses are transformed into a POD curve using log-normal model, which was concluded the best fit among other assessed models [6]. This model is based on a linear relation between defect size (a) and signal response ($\hat{a}$):

$$\ln(\hat{a}) = \beta_0 + \beta_1 \ln(a) + \delta \tag{1}$$

where β_0 and β_1 are the intercept and slope, and δ is a random error term addressing the difference between observed and estimated signal response. Berens [7] assumes that δ is normally distributed with mean value of zero and constant standard deviation σ_δ, which is independent on defect size: $\delta \sim N(0, \sigma_\delta)$.

The POD curve generated by log-normal model is finally expressed in a form:

$$POD(a) = \Phi\left(\frac{\ln(a) - \mu}{\sigma}\right) = \Phi\left(-\frac{\mu}{\sigma} + \frac{1}{\sigma}\ln(a)\right) \tag{2}$$

where Φ denotes standard normal distribution function, and:

$$\mu = \frac{\ln(\hat{a}_{dec}) - \beta_0}{\beta_1} \tag{3}$$

$$\sigma = \frac{\sigma_\delta}{\beta_1} \tag{4}$$

$\hat{a}_{dec}$ in Eq. (3) stands for a specified decision threshold for signal response. The defect is marked detected only if its signal response exceeds this threshold value.

Upon the POD curve is obtained, a large number of additional simulations are continued in this work for a defect size to determine its detection percentage, as a specific estimate of POD for this defect size. This process is repeated for several defect sizes to acquire multiple discrete POD value points, aiming at comparing the modelled POD curve with specific

Fig. 1 General location and parameter definitions of a circular crack
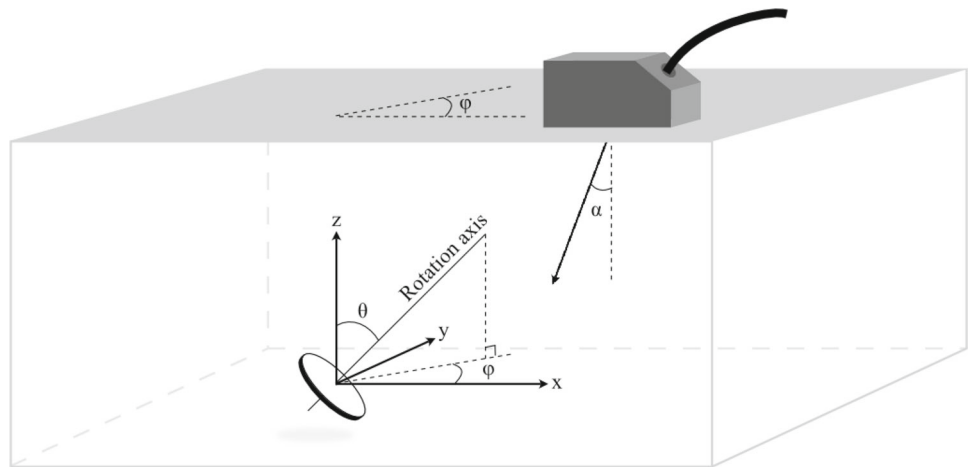
POD values and to evaluate the quality of the modelled lognormal POD curve. These massive amounts of simulations are realized through a metamodel, trained and built based on well selected simulation results, with the purpose of reducing runtime.

In addition, this work also involves a modified distance amplitude correction and time varied gain (DAC/TVG-mod) framework to contribute to simulation results as a postprocessing step. The POD curve and values are also added with these contributions in evaluation.

4 Studied Inspection Scenario for Evaluation

The inspection scenario studied in this paper is PAUT on lack-of-fusion defects in AM components, produced by e.g., laser metal deposition (LMD). The component is made of material with acoustic properties of 5573 m/s and 3150 m/s for longitudinal and transverse wave speed, respectively. To simplify the result obtaining process, factors such as material attenuations, noises etc. are excluded from this study. Within the scope and availability of the UT simulation software simSUNDT, penny-shaped circular crack [33] is chosen as representative of the actual defect in simulations. A general appearance and parameter definitions of a circular crack is shown in Fig. 1, where:

- the rotation axis of the circular crack passes through the center of the crack
- θ is the crack tilt angle (the angle between the crack rotation axis and the vertical z-axis perpendicular to scanning surface)
- ϕ is the crack skew angle (the angle between the probe scan line along x-axis and the projection of the crack rotation axis onto the scanning surface)

Table 1 Key information of linear PA probe simulated in this paper (LM-5MHz)

Specification	Value
Total elements	64
Total length (mm)	38.3
Elevation (mm)	10
Pitch (mm)	0.6
Kerf (mm)	0.1
Center frequency (MHz)	5
Relative bandwidth (%)	74

Figure 1 also shows the inspection setup considered in this work, i.e., contact test using PA probe under pulse-echo (one probe only) situation. Simulated UT instruments are the ones used in previous validations [26, 30]. This includes a PAUT device named Topaz64 [34] from Zetec as data acquisition hardware, and a longitudinal wave linear PA probe labelled LM-5MHz [35] from Zetec. Key information of this PA probe is summarized in Table 1.

To account for probabilistic characteristics of inspection capability, the inspection parameters in this studied case are formulated by uncertainty propagation as follows.

Using the PA probe with proper delay law, the generated sound beam inside the test component should have a nominal beam angle of 0° (α in Fig. 1) with uncertainty of $\pm 2°$. The nominal probe skew angle of 0° is set, which is along the bead direction of AM component, with uncertainty of $\pm 5°$ due to the play in physical situation. Test volume addressed includes defects with nominal depth of 25 mm with uncertainty of ± 5 mm. Focusing effect of the PA probe is used so that most wave energies could be reflected from the test volume. Each defect has a nominal tilt and skew angle of 0° but is specified with uncertainty of $\pm 10°$ and $\pm 90°$, respectively. The

Table 2 Summarization of preliminary inspection parameters with distribution and range

Inspection parameter	Distribution	Mean	Range
Defect size (diameter)	Uniform		[0.5:5] mm
Defect depth	Uniform	25	± 5 mm
Defect tilt angle (θ)	Normal	0	$\pm 10°$
Defect skew angle (ϕ)	Uniform	0	$\pm 90°$
Beam angle (α)	Normal	0	$\pm 2°$
Probe skew angle	Normal	0	$\pm 5°$
Focusing depth	Uniform	25	± 5 mm

diameters of circular cracks, taken as the defect size (characteristic defect parameter), are limited between 0.5 and 5 mm. Table 2 summarizes these preliminary inspection parameters as well as their distributed ranges.

The maximum echo amplitudes from all simulations are taken as the signal response. To assess if the defect is detected, a decision threshold ($\widehat{a}_{dec}$) of -6 dB drop from a calibration defect is set. The calibration defect in this work is selected as a SDH with 0.5 mm in diameter and 25 mm in depth.

5 Numerical Simulations

5.1 Sensitivity Analysis of Inspection Parameters

Practically, a number of inspection parameters are evaluated in terms of their impact on the resulted signal response amplitude, aiming at reducing the number of involved factors in analysis [21]. Identified preliminary inspection parameters in Table 2 thus undergo a sensitivity analysis. UT simulation kernel, UTDefect, is incorporated and launched through a simulation process automation software, modeFrontier, to generate and analyze a series of simulation results. For the purpose of sensitivity analysis, multi-level full factorial sequence of DOE is adopted in modeFrontier for all inspection parameters according to Table 2, except for the characteristic parameter (defect size). This so-called multi-level full factorial sequence of DOE helps sample the involved inspection parameters within each parameter range at multiple levels (values), and plans the experiments or simulations at all possible combinations of these values. It contributes to a comprehensive study of the effect of each parameter on the final response. When all simulations are completed, sensitivity analysis is performed in modeFrontier using Smoothing Spline ANOVA (SS-ANOVA) proprietary algorithm [36], which ends up in relative contribution index of each selected inspection parameter to the simulated maximum echo amplitude. Analysis shows that the inspection parameters can be ordered according to their relative contribution indexes, where larger index indicates higher impact of

inspection parameter to the echo amplitude. Note these contribution indexes in parenthesis indicate only relative impact of each parameter to the result under this analyzed situation:

Focusing depth (0.525) > Defect depth (0.469) > Defect tilt angle (0.004) > Defect skew angle (0.001) > Beam angle (1.9E$-$11) > Probe skew angle (4.7E$-$12).

Contribution index for beam angle and probe skew angle is almost 0 from the analysis. These two inspection parameters are therefore set constant at their nominal value of 0° in simulation parameter space of the study. Other inspection parameters are then treated as essential parameters with proper uncertainties, summarized in Table 3. Though as noted on focusing depth with the highest relative contribution index, it is however treated as constant and set at the nominal depth of test volume, i.e., 25 mm. This could be accomplished and assured in practice by performing corresponding calibration beforehand to exclude the impact of beam focusing effect. Defect sizes are evenly distributed within the range between 0.5 and 5 mm.

5.2 Simulation Scheme

As in physical experiments, operators tend to have a continuous exploration over the scanning surface to find where the potential signal response from the defect of interest could appear, then a local search proceeds around this limited region to find the ultimate maximum signal response. It is especially the case to search in a larger region when the defect features are not witting beforehand because these unknown features can change the position of the ultimate maximum signal response. For mathematical simulations, this overall idea and strategy can to a large extent help avoid inefficient and unintended simulation steps within a complete mesh grid with discrete computation points, with the purpose of capturing the potential maximum signal response, as operators behave in physical experiments. To realize this strategy, a so-called macro–micro scanning pattern is developed and implemented in simulation workflow. This pattern enables that a simulation firstly is performed with a coarse mesh setup with mesh size of s_{x1} and s_{y1} in x- and y-direction, respectively, over the complete scanning area to find a global maximum of signal response, i.e., echo amplitude $\widehat{a}_1$ in this work. Based on the location (x_1, y_1) of this global maximum, the next scanning region shrinks to $x \in [x_1 - s_{x1}, x_1 + s_{x1}]$ and $y \in [y_1 - s_{y1}, y_1 + s_{y1}]$, and new mesh size is decreased 5 times from previous size, i.e., $s_{x2} = s_{x1}/5$ and $s_{y2} = s_{y1}/5$. This process is repeated until the mesh size is decreased below 0.1 mm. The last mesh size is restricted to 0.1 mm because it already gives convergent results from UTDefect kernel. An illustration of the process is shown in Fig. 2.

Table 3 Summarization of important inspection parameters (essential parameters) in parameter space

Inspection parameter	Distribution	Mean	Standard deviation	Delta	Range
Defect depth	Uniform	25 mm		5 mm	[20:30] mm
Defect tilt angle (θ)	Normal	0°	2.67		[− 10:10]°
Defect skew angle (ϕ)	Uniform	0°		90°	[− 90:90]°
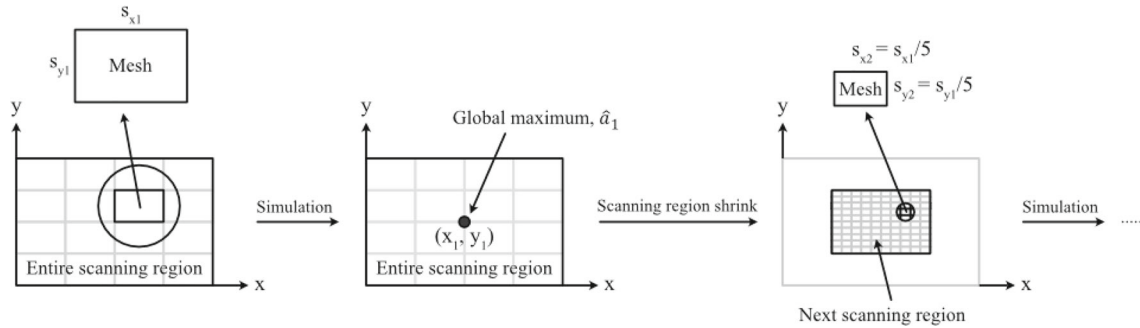
Fig. 2 Illustration of the macro–micro scanning pattern used in simulations
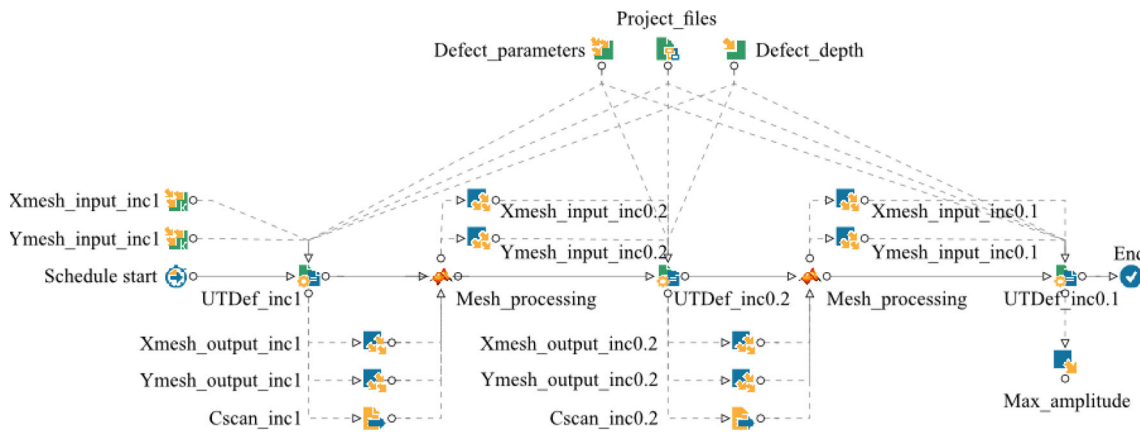
Fig. 3 Macro–micro scanning pattern in modeFrontier workflow used for efficient numerical simulation of current study cases

It is in the end implemented as an automation workflow in modeFrontier as shown in Fig. 3 for the specific study case in this paper, where $s_{x1} = s_{y1} = s_1 = 1mm$ (corresponds to "inc1" in the figure) and $s_{x2} = s_{y2} = s_1/5 = s_2 = 0.2mm$ ("inc0.2" in the figure). After a UTDefect simulation, a MATLAB script is used in the workflow for generating the new mesh sizes and updating the scanning area based on the resulted C-scan file. This helps save enormous simulation runtime. As an example of a defect with diameter of 5 mm, simulation with mesh size of 0.1 mm within the entire scanning region takes around 50 h for just one case, but is surprisingly reduced down to 45 min using the macro–micro scanning pattern. This makes heavy computations more feasible.

5.3 Simulation Base for Metamodel Training

To get statistically sufficient number of inspection results for estimating a POD value of a certain defect size, a method of response surface model (RSM), in other words metamodel, is to be used. The metamodel is an approximate mathematical representation of complex model responses, constructed or trained based on a selection of time-consuming simulation results from original mathematical model. Metamodel could then provide efficient and fast estimations of the model response instead of calculating through the original mathematical model.

To construct a valid metamodel for the current study, a multi-level full factorial DOE sequence is applied to gener-

ate a total of 625 inspection cases in modeFrontier, within the parameter space specified in Table 3. Simulation of these DOEs using UTDefect are sequentially and automatically launched by modeFrontier with application of the developed macro–micro scanning pattern. A metamodel with defect size, depth, tilt and skew angle as input parameters and maximum echo amplitude as output value is finally built using Duchon's Polyharmonic Spline based radial basis function [37]. This metamodel correlates well with results from UTDefect simulations in verification step, with mean absolute error of 0.5 dB and mean relative error of 0.6%. The output from this metamodel is called virtual simulations in the following context.

5.4 Modified Distance Amplitude Correction and Time Varied Gain (DAC/TVG) Function

Practically, DAC/TVG function is applied in inspections to compensate for material attenuation and beam spreading etc., so that the differences of signal amplitudes due to sound wave travel distance could be reduced to some extent. This function is additionally considered and evaluated in this work in a certain way, modified as followed, to examine its influence on final POD results.

A series of reference reflectors, usually FBHs, of the same size at different depth are commonly used in practice to generate DAC/TVG curve. A DAC curve plots variations of signal amplitudes of these reflectors as a function of their depth, and a corresponding TVG presents compensating gains for these reflectors such that all signal amplitudes are brought to the same height, e.g., normally 80% screen height of the instrument, see an illustration in Fig. 4.

Considering the current study case and simulation results from UTDefect, a modified DAC/TVG function framework is proposed, denoted as DAC/TVG-mod. Two reference blocks with the same material properties as the current study case are assumed. Reference block 1 contains 11 FBHs of size 0.5 mm as reference reflectors, situating at depth from 20 to 30 mm with increment of 1 mm, covering the entire test volume of the study case. Block 2 as an extra complement, is the same as Block 1 but the FBH sizes are 5 mm instead. For each reference block, simulations using the same inspection setup as the study case, i.e., 0 degree beam angle and probe skew angle with a fixed beam focusing depth of 25 mm, are performed on all contained FBHs. By the availability of defect types in simSUNDT, the FBHs are well-represented by open circular cracks in simulations, which are experimentally verified. Their maximum echo amplitudes are recorded as signal responses, where the largest response value from each block is then treated as the "standard level", A_s. Amplitude differences from other reflectors on the same block would be the corresponding gains at individual depth and the gain value
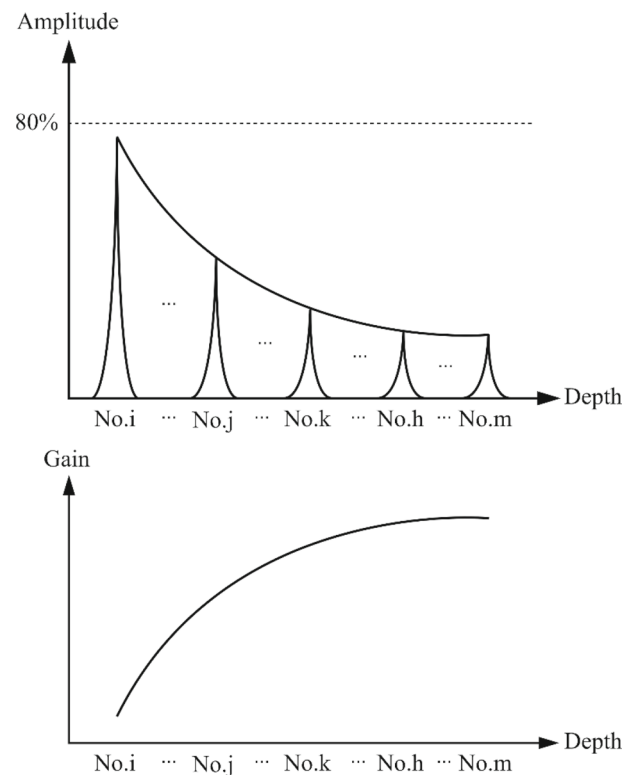

Fig. 4 General DAC curve (top) and corresponding TVG (bottom) through a series of reference reflectors

from any depth interval is retrieved through linear interpolation.

6 Results and Discussions

6.1 Modified DAC/TVG Function

As mentioned in Sect. 5.4, two reference blocks are proposed in this work with FBHs of different sizes. Simulations using UTDefect provide echo amplitudes for DAC curves, see Table 4 and Fig. 5. Note that the resulted amplitude from UTDefect is expressed in log scale in terms of decibel (dB), which has no practical meaning unless it is calibrated. The values with superscript "a" in Table 4 indicate the "standard level", A_s, for corresponding reference block.

Corresponding TVG are retrieved through amplitude differences to the "standard level", A_s, at each reflector depth, see Table 5 and Fig. 6. Required gain in this work for a defect depth is thus available through these results by interpolation.

6.2 Simulation Results Processing

With the help of the developed metamodel in this work, 20 inspection cases for each defect size are conducted, among which the essential parameters of the inspection cases are

Table 4 Maximum echo amplitudes (in dB) from two proposed reference blocks for DAC curve

Reflector depth (mm)	20	21	22	23	24	25	26	27	28	29	30
Block 1	51	52.5	54.7	60	63.7	64.7[a]	63.2	59.8	54.9	50.8	48.6
Block 2	81.6	82.3	83.4	84.8	88.2	89.8[a]	88	84.4	82.2	81.1	80.1

[a]The largest response amplitude from each block is taken as the "standard level", A_s
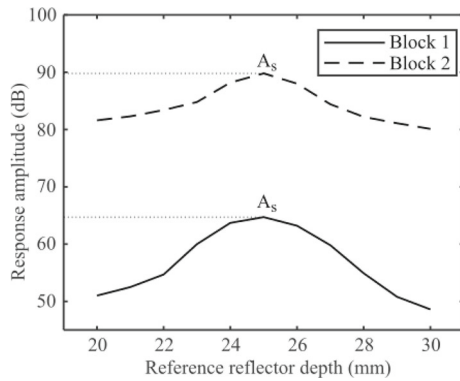
Fig. 5 DAC curve for two proposed reference blocks
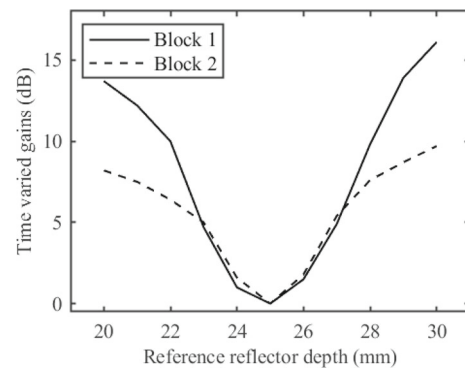
Fig. 6 TVG curve for two proposed reference reflectors

distributed according to Table 3. In total, 30 defect sizes are evenly distributed within the formulated defect size range, i.e., from 0.5 to 5 mm, which means 600 virtual simulations in total are computed.

Figure 7 plots the distribution of these calibrated echo amplitudes against the defect sizes. Based on report [38], extra attention should be focused on evaluating the performance of the default log–log transformation of defect size (a) and signal response ($\hat{a}$) in log-normal POD model. This is to ensure the linear relation between these two quantities, assumed by Berens [7] in Eq. (1). The report points out that log–log scale is not always the best transformation of this linear relation, but either linear or log scale of these quantities is to be considered for a combination that best fits the current dataset, to satisfy the basis of log-normal POD model. In this study, the calibrated echo amplitude results from simulations are already in unit of decibel (dB) as in log scale. It is for this reason that only defect sizes are taken in both log and linear scales in Fig. 7 for assessment. Additionally, to correctly apply the log-normal POD model, assumptions of normal distribution and constant standard deviation for the random

error term δ in Eq. (1) are to be ensured, as introduced in Sect. 3.

It is observed from Fig. 7 (left) that echo amplitude distribution under log scale of defect size can be better approximated using a straight line, and the assumptions of modelling error are also satisfied. Thus, using log-scale of defect size could ensure the validity of applying log-normal POD model for this dataset. It is also noticed that a data point from a large defect size (with circle) seems to stay out of the normal distributed region of other points. This will be addressed and discussed later.

Taking the DAC/TVG-mod function into consideration using the proposed two reference blocks, resulted echo amplitudes are hereby called TVG-compensated echo amplitudes in this paper. An example addressing 5 different defect sizes (0.5 mm, 1.6 mm, 2.8 mm, 3.9 mm and 5 mm) and 20 inspection cases per size is shown in Fig. 8 using TVG from two blocks. Note that these data from different blocks are plotted with a little shift in abscissa for clear visualization purpose only.

It is noted in Fig. 8 that reference Block 1 (with FBHs of size 0.5 mm) helps bring the calibrated echo amplitudes from defects of size 0.5 mm (Def. 1) to a very similar level (the

Table 5 Corresponding TVG (in dB) based on DAC and "standard level" for two proposed reference blocks

Reflector depth (mm)	20	21	22	23	24	25	26	27	28	29	30
Block 1	13.7	12.2	10	4.7	1	0	1.5	4.9	9.8	13.9	16.1
Block 2	8.2	7.5	6.4	5	1.6	0	1.8	5.4	7.6	8.7	9.7

Fig. 7 Distribution of defect echo amplitudes against defect sizes of 600 simulation cases (30 defect sizes with 20 inspections per size) in log (left) and linear (right) scale of defect size (abscissa). Decision threshold ($\hat{a}_{dec}$) is set at -6 dB after calibrating with a SDH
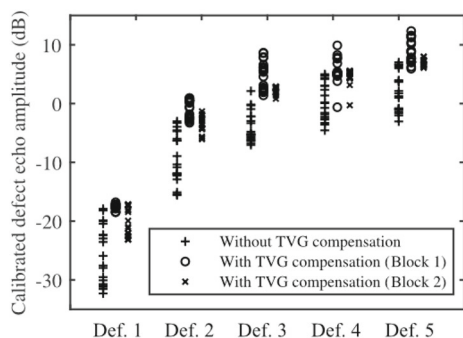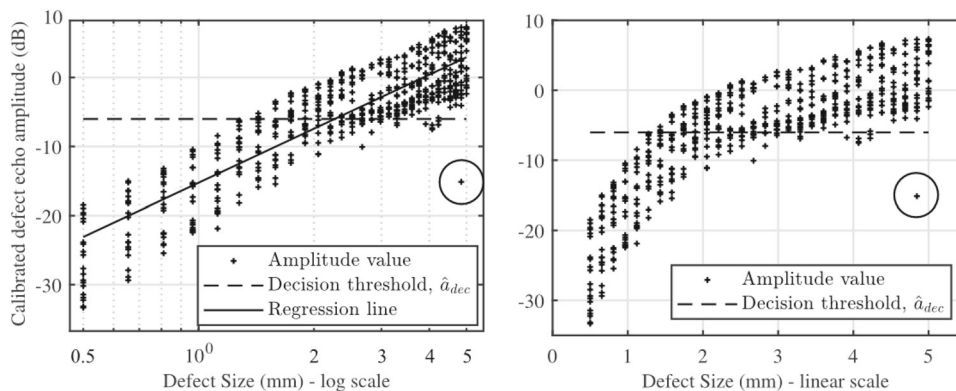

Fig. 8 An example dataset showing calibrated echo amplitudes on 5 different defect sizes (from Def. 1 to Def. 5 are 0.5 mm, 1.6 mm, 2.8 mm, 3.9 mm and 5 mm) with the impact from applying different TVG functions obtained by the two proposed reference blocks

circle signs of Def. 1 "shrink" very much from corresponding plus signs), as the purpose of reducing the impact of sound wave travel distance (i.e., defect depth in this work) using DAC/TVG-mod function. However, Block 1 helps little when the defect sizes deviate from 0.5 mm (the circle signs of Def. 5 do not "shrink" that much from corresponding plus signs). Reference block 2 (with FBHs of size 5 mm) reveals similar behavior as Block 1 on the same defect size (Def. 5), but it also reduces the impact of defect depth on much different defect sizes (Def. 1), which differ much to FBH sizes of 5 mm. This observation is further confirmed by sensitivity analysis, investigating the impact of defect depth to resulted echo amplitudes with or without using Block 1 and 2. The relative contribution index of defect depth from each scenario can be ordered in parenthesis as below, which indicates that DAC/TVG-mod function by Block 2 helps better in reducing the impact of defect depth to echo amplitudes:

No DAC/TVG-mod function (0.144) > With DAC/TVG-mod using Block 1 (0.052) > With DAC/TVG-mod using Block 2 (0.003)

Applying these DAC/TVG-mod functions to original dataset in Fig. 7 gives distribution of TVG-compensated echo

amplitudes, shown in Figs. 9 and 10 by using Block 1 and 2, respectively. Log scale of defect size (figures to the left) could still present better modelling behavior of linearity comparing to linear scale (figures to the right). This is consistent as concluded from Fig. 7, thus is employed in corresponding modelling process of POD curves. It is however noted from these log-scaled figures when it comes to other hypotheses of POD log-normal model, that the standard deviation of resulted data is not uniform about the estimated straight line, which represents the linear relation in Eq. (1). In other words, the standard deviation of random errors is dependent on defect sizes when DAC/TVG-mod function is involved. There are also some data points from large defect sizes fall outside of the normal distributed region along the regression line. In convenience of discussion for some of them, we number the data points No.1–5 in Fig. 10 (left), where data point No.1–4 fall above the decision threshold and point No.5 falls below the threshold. This point No.5 is also the circled data point indicated in Fig. 7. The combination of the inspection parameters of these resulted data points are listed in Table 6. It can be noticed that these data points come from the defects with large tilt and skew angle. However, it is inappropriate to conclude that a combination of large tilt and skew angle of defect will give a weak echo amplitude if this DAC/TVG-mod function is applied. In other words, if the DAC/TVG-mod function is applied, then the large tilt and skew angle of defect is the necessary and insufficient condition of having a weak echo amplitude. This is because a weak echo amplitude could be compensated with a high TVG depending on the defect depth. It is hereby a joint effect of these inspection parameters. These specific cases of No.1–5, especially the No.5 data point could however question the resulted POD curve when the log-normal POD model is fitted in upcoming work.

6.3 Estimation of Model Parameters and POD Curves

As seen from Eq. (2), log-normal POD model is controlled by parameters β_0, β_1 and σ_δ. For the dataset of calibrated

Fig. 9 Distribution of defect echo amplitudes against defect sizes of 600 simulation cases (30 defect sizes with 20 inspections per size) in log (left) and linear (right) scale of defect size (abscissa). DAC/TVG-mod function is applied using reference block 1 and decision threshold ($\hat{a}_{dec}$) is set at -6 dB after calibrating with a SDH
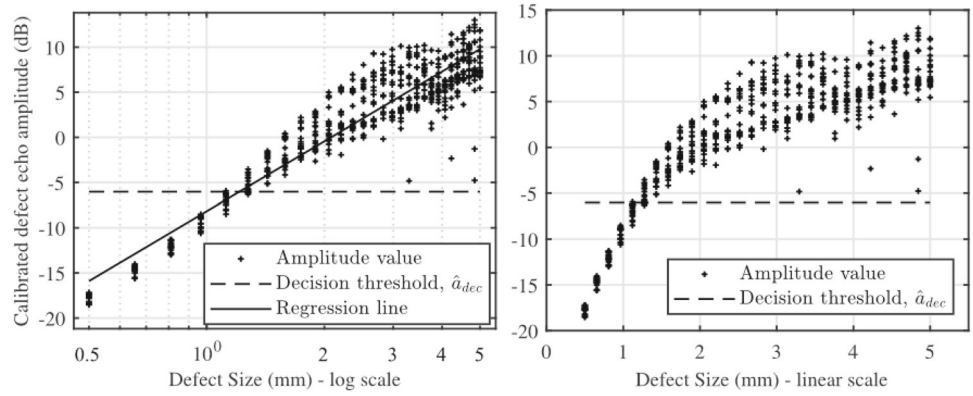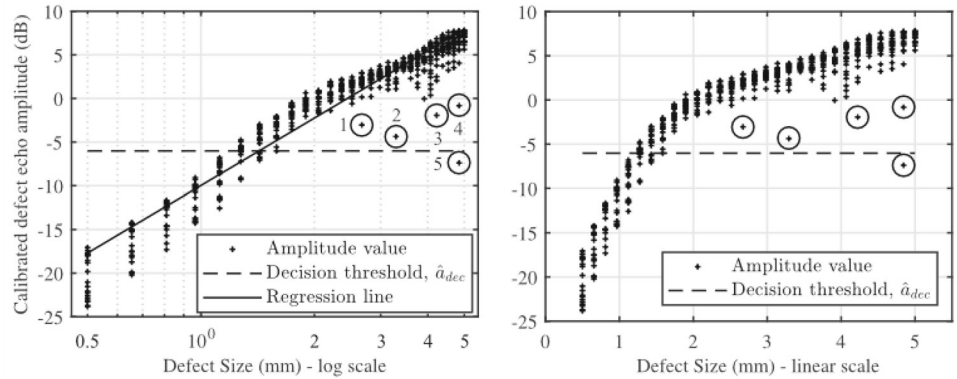
Fig. 10 Distribution of defect echo amplitudes against defect sizes of 600 simulation cases (30 defect sizes with 20 inspections per size) in log (left) and linear (right) scale of defect size (abscissa). DAC/TVG-mod function is applied using reference block 2 and decision threshold ($\hat{a}_{dec}$) is set at -6 dB after calibrating with a SDH

Table 6 Combination of inspection parameters for the five resulted data points in Fig. 10

No.	Defect size (mm)	Defect depth (mm)	Defect tilt angle (°)	Defect skew angle (°)
1	2.67	21.42	7.53	68.58
2	3.29	24.24	7.54	− 81.65
3	4.22	23.32	6.43	88.83
4	4.84	23.52	6.48	73.69
5	4.84	28.14	− 9.96	70.05

Table 7 Estimated POD (log-normal) model parameters for different datasets after calibration

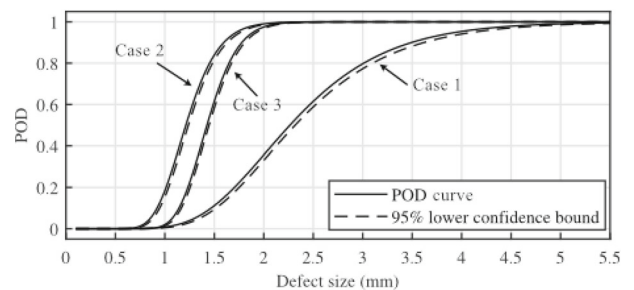|	β_0	β_1	σ_δ
No TVG compensation	− 15.3	11.3	3.8
TVG-compensated using reference block 1	− 8.2	11.2	2.4
TVG-compensated using reference block 2	− 9.9	11.1	1.9

Fig. 11 POD curves and confidence bound for three cases (with or without TVG compensation)

echo amplitudes without and with TVG-compensation using reference block 1 and 2 (figures to the left in Figs. 7, 9 and 10, respectively), the estimated model parameters are summarized in Table 7. POD functions for these datasets using parameters in Table 7 are then expressed as below in form of Eq. (2) and plotted in Fig. 11.

Case 1: Calibrated echo amplitude without TVG compensation:

$$POD(a) = \Phi(-2.428 + 2.95\ln(a))$$

Case 2: Calibrated and TVG-compensated (Block 1) echo amplitude:

$$POD(a) = \Phi(-0.916 + 4.67\ln(a))$$

Case 3: Calibrated and TVG-compensated (Block 2) echo amplitude:

$$POD(a) = \Phi(-2.069 + 5.78\ln(a))$$

Defect size of 90% POD with 95% confidence, $a_{90/95}$, is 3.6 mm, 1.6 mm and 1.8 mm for three respective cases, which clearly indicates from Fig. 11 that POD is improved when DAC/TVG-mod function is in use (case 2 and case 3) under the same decision threshold of -6 dB, i.e., smaller defect sizes could have a higher POD after TVG-compensation. This is because echo amplitudes from all defects are compensated with certain gains depending on their depth. The difference of resulted POD curves between case 2 and case 3 in Fig. 11 comes from the level of compensated gains by two reference blocks, as seen in Fig. 6. These gains help some echo amplitudes reach the detection threshold.

As concerned previously in Sect. 6.2, the resulted POD curves in Fig. 11 based on the estimated parameters in Table 7 could be doubtful due to non-uniform standard deviation about the estimated straight line and some data points outside of normal distributed region of most data points. This is to be further assessed and compared by estimating discrete POD value points at some defect sizes. A discrete POD value point for a defect size, according to POD definition, is obtained through performing a series of inspections on defects of this size and counting the proportion of times this defect size being detected among all trials. Taking the advantage of constructed metamodel, 5000 virtual simulations for a defect size can be rapidly accomplished. To account for a certain defect size range, 24 defect sizes are included in this investigation. The same detection criteria of -6 dB after calibration as in POD curve model is used for these virtual simulations. These POD value points for three investigated cases are plotted with original POD curves in Fig. 12.

In general, Fig. 12 shows good correlations in trend between POD curve and discrete POD value points. The POD curves based on log-normal POD model underestimate the POD for defect sizes larger than about 1.3 mm for all cases. Defect sizes smaller than this have limited POD as seen from discrete values, while POD curve could prescribe a higher probability.

Specifically, it should be pointed out that not all discrete POD value points in Fig. 12 converge to 1 above a certain value of defect size. For example, the points of defect size from 3 to 5 mm have values between 0.95 and 0.98 for case 1, between 0.9992 and 1 for the points of defect size above 1.5 mm for case 2, and between 0.9994 and 1 for the points
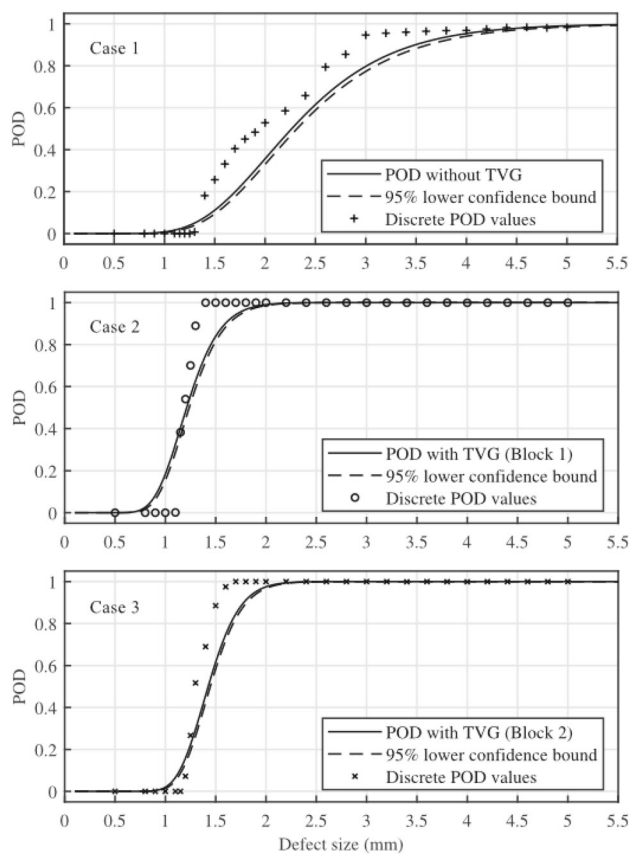

Fig. 12 Comparison between POD curve and corresponding discrete POD value points at some defect sizes

of defect size above 1.7 mm for case 3. Though these POD values are very close to and could be treated as 1, they however indicate that there are still a small number of cases with combination of inspection parameters that can give resulted echo amplitudes below the decision threshold level. Part of these small number of echo amplitudes could also fall outside of the normal distributed region of most other results, as concerned in Sect. 6.2. These small number of cases, as being examined, come from the combination of large tilt and skew angles of defect, similar to earlier discussions. It should yet be emphasized again that the large tilt and skew angle of defect is only the necessary but insufficient condition of having a weak echo amplitude, if the DAC/TVG-mod function is applied. Now even if there are some echo amplitudes fall outside of the normal distributed region, which violate the corresponding assumption of log-normal POD model, we should still note that the number of these cases are very limited among 5000 cases. The comparisons in Fig. 12 between the discrete POD value points and corresponding POD curve show no need for this concern of violation of model assumptions.

7 Conclusion

This paper addresses a process of generating datasets for quantifying NDE detection capacity in terms of POD, of certain ultrasonic inspection scenario based on numerical simulation model. A set of inspection parameters for this specific inspection scenario undergo a sensitivity analysis to determine essential parameters in parameter space. These essential parameters, which have greater impact on the inspection results, are specified with reasonable uncertainties and distributions based on uncertainty propagation method. A series of well-organized UT simulations are performed in the parameter space using a proposed efficient macro–micro simulation scheme, which saves numerous runtimes of simulations. By these simulated results, a metamodel is constructed with the aim of providing fast computations, instead of running the original model. This can help perform thousands of virtual simulations rapidly for estimating statistically valid POD values. In the end, 30 defect sizes with 20 inspection cases each have been computed within the parameter space and they are transformed into POD curve using log-normal POD model. A modified DAC/TVG function is also applied on this dataset and improvement of POD results are seen from the new curves. Despite the fact that some of Berens hypotheses of log-normal POD model are not perfectly satisfied in datasets with DAC/TVG-mod compensation, discrete POD values from some defect sizes could still prove the log–normal function as a satisfactory model to signal response type of inspection datasets.

Declarations

Conflict of interest The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

1. Rummel, W.D.: Probability of detection as a quantitative measure of non-destructive testing endto-End process capabilities. Mater. Eval. **56**, 29–35 (1998)
2. AGARD Lecture Series 190: A recommended methodology to quantify NDE/NDI based on aircraft engine experience (1993).
3. Forsyth, D.S., Fahr, A.: An evaluation of probability of detection statistics. Institute of Aerospace Research, 10–1 (1998)
4. Bato, M.R., Hor, A., Rautureau, A., Bes, C.: Experimental and numerical methodology to obtain the probability of detection in eddy current NDT method. NDT E Int. **114**, 102300 (2020). https://doi.org/10.1016/j.ndteint.2020.102300
5. Lewis, W.H., Dodd, B.D., Sproat, W.H., Hamilton, J.M.: Reliability of Non-destructive Inspection—Final Report. SA-ALC/MME 76-6-38-1. San Antonio Air Logistics Centre, Kelly Air Force Base, Texas (1978)
6. Berens, A.P., Hovey, P.W.: Evaluation of NDE reliability characterization. Report No. AFWAL-TR-81-4160 (1981)
7. Berens, A.P.: NDE reliability data analysis. ASM Handb **17**, 689–701 (1989). https://doi.org/10.1361/asmhba0002238
8. Wall, M., Burch, S.: Worth of Modelling for Assessing the Intrinsic Capability of NDT. 15th World Conf. Nondestruct. Test. pp. 15–21 (2000)
9. Wall, M.: Modelling of inspection reliability. IEE Colloq. Insp. Reliab. State-of-the Art, vol. 1996, IEE; pp. 7–7. https://doi.org/10.1049/ic:19961033 (1996)
10. Chapuis, B., Calmon, P., Jenson, F.: Best Practices for the Use of Simulation in POD Curves Estimation. Springer International Publishing, Cham (2018). https://doi.org/10.1007/978-3-319-62659-8
11. Schubert, F., Baron, H., Menges, J., Dorval, V., Gilles-Pascaud, C., Raillon-Picot, R., Barden, T.: Simulation-Supported POD for Ultrasonic Testing–Recommendations from the PICASSO Project. In 5th European-American Workshop on Reliability of NDE (2013)
12. Harding, C.A., Hugo, G.R., Bowles, S.J., Thompson, D.O., Chimenti, D.E.: Application of model-assisted pod using a transfer function approach. AIP Conf. Proc., AIP; p. 1792–9. https://doi.org/10.1063/1.3114176 (2009)
13. Thompson, R.B., Thompson, D.O., Chimenti, D.E.: A unified approach to the model-assisted determination of probability of detection. AIP Conf. Proc., vol. 975, AIP; p. 1685–92. https://doi.org/10.1063/1.2902639 (2008)
14. Wall, M., Wedgwood, F.A., Burch, S.: Modelling of NDT reliability and applying corrections for human factors. Proc. 7th Eur. Conf. on NDT, Copenhagen, Denmark (1998)
15. Spies, M., Rieder, H.: An approach to the question 'How to account for human error in MAPOD?'. In 12th ECNDT Conference, Gothenburg, Sweden (2018)
16. https://www.cnde.iastate.edu/MAPOD/. Accessed 22 December 2021
17. Burch, S.F., Stow, B.A., Wall, M.: Computer modelling for the prediction of the probability of detection of ultrasonic corrosion mapping. Insight Non-Destr. Test Cond. Monit. **47**, 761–764 (2005). https://doi.org/10.1784/insi.2005.47.12.761
18. Volker, A.W.F., Dijkstra, F.H., Terpstra, S., Heerings, H.A.M., Lont, M.A.: Modeling of NDE reliability: development of a POD generator. Proc. 16th World Conf. Nondestruct. Test., Montreal, Canada, August (2004)

19. Jensen, F., Mahaut, S., Calmon, P., Poidevin, C.: Simulation based POD evaluation of NDI techniques. 10th Eur. Conf. non-destructive Test., Moscow (2010).

20. Foucher, F., Fernandez, R., Leberre, S., Calmon, P.: New Tools in CIVA for Model Assisted Probability of Detection (MAPOD) to Support NDE Reliability Studies. NDE Aerosp. Mater. Struct. p. 32–43 (2018)

21. Wirdelius, H., Persson, G.: Simulation based validation of the detection capacity of an ultrasonic inspection procedure. Int. J. Fatigue **41**, 23–29 (2012). https://doi.org/10.1016/j.ijfatigue.2011.09.002

22. Boström, A., Wirdelius, H.: Ultrasonic probe modeling and non-destructive crack detection. J. Acoust. Soc. Am. **97**, 2836–2848 (1995). https://doi.org/10.1121/1.411850

23. Jansson, P.A.: Modeling of ultrasonic detection of a rectangular crack. AIP Conf. Proc., vol. 509, AIP; pp. 57–64. https://doi.org/10.1063/1.1306034 (2000)

24. Boström, A., Eriksson, A.: Scattering by two penny-shaped cracks with spring boundary conditions. Proc. R. Soc. Lond. Ser. A **443**, 183–201 (1993). https://doi.org/10.1098/rspa.1993.0139

25. Bövik, P., Boström, A.: A model of ultrasonic nondestructive testing for internal and subsurface cracks. J. Acoust. Soc. Am. **102**, 2723–2733 (1997). https://doi.org/10.1121/1.420326

26. Lei, X., Wirdelius, H., Rosell, A.: Experimental validation and application of a phased array ultrasonic testing model on sound field optimization. J. Mod. Phys. **12**, 391–407 (2021). https://doi.org/10.4236/jmp.2021.124028

27. Wirdelius, H.: Experimental validation of the UTDefect simulation software. Proc. 6th Int. Conf. NDE Relat. to Struct. Integr. Nucl. Press. Components, Budapest (2007)

28. Auld, B.A.: General electromechanical reciprocity relations applied to the calculation of elastic wave scattering coefficients. Wave Motion **1**, 3–10 (1979). https://doi.org/10.1016/0165-2125(79)90020-9

29. Wirdelius, H.: Probe model implementation in the null field approach to crack scattering. J. Nondestruct. Eval. **11**, 29–39 (1992). https://doi.org/10.1007/BF00566015

30. Lei, X., Wirdelius, H., Rosell, A.: Experimental validation of a phased array probe model in ultrasonic inspection. Ultrasonics **108**, 106217 (2020). https://doi.org/10.1016/j.ultras.2020.106217

31. USA Department of Defense Handbook, MIL-HDBK-1823, NDE system reliability assessment (2009).

32. Jenson, F., Iakovleva, E., Dominguez, N., Thompson, D.O., Chimenti, D.E.: Simulation supported pod: methodology and HFET validation case. AIP Conf. Proc. **1335**, 1573–1580 (2011). https://doi.org/10.1063/1.3592117

33. Boström, A., Wickham, G.: On the boundary conditions for ultrasonic transmission by partially closed cracks. J. Nondestruct. Eval. **10**(4), 139–149 (1991). https://doi.org/10.1007/BF00567096

34. https://www.zetec.com/products/ultrasound/instrumentation/topaz64/topaz64-smart-features/. Accessed 22 December 2021

35. https://www.zetec.com/wp-content/uploads/2016/11/2016-UT-Probe-Catalog-Content_R20161125.pdf. Accessed 22 December 2021

36. Gu, C.: Smoothing Spline ANOVA Models. Springer, New York (2013)

37. Chen, W., Fu, Z., Chen, C.: Recent Advances in Radial Basis Function Collocation Methods. Springer, Heidelberg (2014). https://doi.org/10.1007/978-3-642-39572-7

38. Gandossi, L., Annis, C.: Probability of detection curves: Statistical best-practices. ENIQ Rep. (2011). https://doi.org/10.2790/21826