e-Journal of Nondestructive Testing - ISSN 1435-4934 - www.ndt.net

Using modelling and metamodels for reliability studies in NDE

Fabrice FOUCHER[1], Mariana BURROWES GUIMARÃES[2], Edward GINZEL[3]

1 EXTENDE, 14 Avenue Carnot 91300 Massy, France, fabrice.foucher@extende.com
2 Federal University of Rio de Janeiro, Brazil, marianaburrowes@gmail.com
3 Materials Research institute, Waterloo, Canada, eginzel@mri.on.ca

Abstract

As in any other measurement process, Non-Destructive Evaluation (NDE) is subject to variability whose impact can be assessed to guarantee a given level of performance. Once NDE prevents catastrophic failures, deaths and environmental damage, identifying uncertainties and variability in NDE help to design more reliable inspections, therefore is a process that saves lives. This is the goal of a reliability study. Statistical indicators such as Probability of Detection (POD) curves give insights to allow building of mechanical designs with enough « secure margin » for structural integrity and to also define appropriate maintenance & inspection cycles. Simulation is very useful to support performance or reliability demonstrations that require a lot of data (such as POD studies and qualification campaigns), and where simulation can help by reducing the number of necessary mock-ups and experimental trials. In addition to physical models, the NDE simulation software CIVA now offers metamodelling techniques. Built from an initial set of physical simulations, such surrogate models give the user the possibility to generate a massive amount of data while combining and exploring multi parametric variations. This is particularly efficient in the context of reliability studies, when the best settings, track the worst-case scenario must be found or build POD curves. This paper illustrates the use of this metamodelling approach for the reliability study of a longitudinal weld AUT (automated ultrasonic testing) inspection. Real pipe mill inspection data are provided and compared to modelling and metamodelling results.

KEYWORDS: Reliability, POD, Ultrasound, Simulation, MAPOD

 https://doi.org/10.58286/28688

1. Introduction

In the context of NDE reliability studies, extensive parametric analyses are required to identify essential parameters that can affect the NDE performance. Such studies need a large amount of data which is often difficult and costly to obtain with a set of purely experimental results. Probability Of Detection methods, that links the probability to detect a detrimental flaw to its size, is one indicator used for NDE reliability evaluation including its quantification. The statistical validity of this approach is also dependent on a sufficient amount of data. Numerical simulation tools can be particularly useful at that stage thanks to their ability to generate large data sets at a relative low cost. It can also help to explore deeper and more precisely some parameters' variability that can be difficult to monitor in an experimental Design Of Experiment (DOE). In addition to classical "numerical" simulations, "Metamodels" are now available, which drastically increase the capacity to generate even larger sample sizes. For parametric and sensitivity analyses, or Model Assisted POD studies, such tools give access to results (such as Sobol Indices, beam of POD curves, non-parametric POD curves) that simply cannot be reached with experimental studies. This paper illustrates the use of this metamodelling approach for the reliability study of a longitudinal weld Automated Ultrasonic Testing (AUT) inspection. Real pipe mill inspection data are provided and compared to modelling results.

2. Modelling software for NDT

2.1 Overview of available models in CIVA platform

The CIVA platform is a well-established multi technique simulation and analysis software in NDT [1]. Its various modules give access to different NDT methods and techniques: Ultrasonic Testing (UT), Guided Waves Testing (GWT), Eddy Current Testing (ET), Radiographic Testing (RT) & Computed Tomography (CT), Thermographic Testing (TT) and is extended by Structural Health Monitoring applications based on guided ultrasonic waves.
The mathematical formulations used in the different modules often rely on semi-analytical models. This approach allows for solving a large range of applications while offering very competitive calculation times compared with purely numerical methods (FEA, etc.). For instance, most of the modelling configurations available in the UT module will rely on a geometrical ray approach to compute beam propagation (the so-called "pencil method"). The interaction with discontinuities involves several models depending on the context, some of them rely on semi-analytical or analytical formulations, the "Kirchhoff" or "GTD" (which stands for "Geometrical Theory of Diffraction") models can be mentioned but other ones have also been implemented to cover several configurations [2]. The current trend is to implement hybrid approaches where semi-analytical methods are used in conjunction with a transient Finite Element Method (FEM) [3].

2.2 About validation

To be effectively used as a source of quantitative justification for NDE reliability, one prerequisite is to rely on sound validation evidence. Models' validations in CIVA take place at different stages. Validation works are usually performed to establish the field of

validity of a new feature or model (comparison with experiments, with other models available in the platform, with literature, etc.). The development team also participated to the international UT and ET modelling benchmarks proposed annually for more than 10 years by the World Federation of NDE centres and published in the QNDE conference, which aimed at comparing different simulation codes to experimental data provided to all participants (see http://wfndec.org). Because validations are performed all along the development of new models or really targeted to an application, and as a lot of cases cannot be published, it is difficult to capitalize all these sets of works in an organized way clearly presented to the user. That is why a specific effort has been put on validation to provide evidence of the modelling results' validity in various situations, or to show the limits of semi-analytical models. These validation campaigns, funded by EXTENDE, have been performed during several years after 2010 and have been published on the EXTENDE website, an overview of these validation campaigns is provided in [4].
However, it is not possible to validate all potential configurations and therefore, it is advised to include as much as possible some relevant reference experimental measurements to evaluate models' accuracy in the frame of any reliability study.

2.3 Metamodeling approach in a few words

A metamodel is a surrogate model relying on "smart interpolators" and is built to replace a physical-based model. The first step consists in computing a data base of simulation results for a given range of a parameter variation. The metamodel is built from these reference data and, after having evaluated its accuracy, it enables a real-time exploration of the full range of inspection scenarios constituted by parameter variations. It becomes possible to achieve statistical analysis on data such as sensitivity and POD studies. For instance, Sobol indices can be computed from metamodel output to quantify the relative importance of influential parameters.
Various DOE methods can be selected to build the simulation data base used to initiate the metamodel. This can be a Full Factorial design (range of variation and number of values for each parameter explicitly defined); however, other drawing schemes such as Latin Hypercube Sampling "LHS" (which generates pseudo random sequences of parameters value, widely used to construct computer experiments), generally reaches a better metamodel accuracy with a much smaller number of computations. Also, several interpolators can be applied to build the metamodel from the database (Multilinear, Radial Basis function, Kriging, etc.). Interested readers can refer to the following paper for more detailed information on the metamodels currently implemented in the CIVA software [5].

3. Pipeline longitudinal weld inspection and model

3.1 Test piece and inspection system

The experimental inspection results considered for this paper are issued from the LNDC (Laboratory of Non-Destructive Testing, Corrosion and Welding) that is part of the Metallurgical and Material Engineering Department of the Federal University of Rio de Janeiro [6]. The test piece is a 12 m long pipeline portion made of X-65 alloy with a longitudinal weld performed by SAW technique (Submerged arc welding). A total of 99 artificial defects had been inserted in the weld area, varying in types, dimensions, positions and orientations. The table 1 lists the different flaws. The pipe has an 18'' outer diameter, 28 mm wall thickness and density of 7.8 g/cm^3. Sound velocity of X-65 steel for transverse waves was considered to be 3230 m/s. The weld exhibits a X-bevel profile with a main bevel angle of 95°.

Table 1. List of defects inserted in the pipe used in experimental AUT inspection.

		Sizes of Defects		
Types of Defects	Number of Defects	Heights	Lengths	Depths
Lack Of Fusion	9	0.35-2.10	1.5-12.0	0.5 - 24
Lack Of Penetration	14			
Cracks on HAZ	20			
Transverse cracks Type A	12			
Transverse cracks Type B	24			
Transverse cracks Longitudinal cracks	20			

The Automated UT (AUT) inspection system relies on a set of different mono-element ultrasonic probes combined to form a multi-channel tool, each channel having the role to inspect one zone in and around the weld area, as shown in Figure 1. Each channels pair has its own distance to the weld line and its own refraction angle. Once set around the weld, a mechanized scanner allows to cover the full length of the test piece.

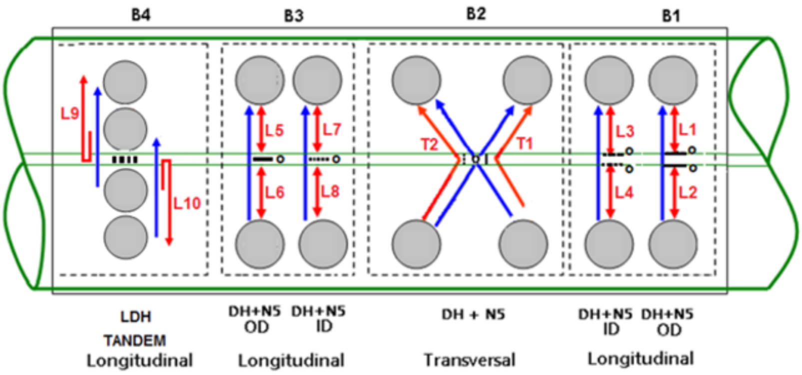


Figure 1. Overview of the different channels of the AUT inspection system

3.2 Building the nominal model

Before launching an extensive sensitivity or POD simulation study, the inspection “nominal” model shall be defined in the software. The Figure 2 illustrates the test piece (pipe, weld and Heat Affected Zones) and the L7 inspection channel set up in the CIVA software. The weld and HAZ is assumed homogeneous in this case, with the same acoustic properties as the pipe base material. This channel involves a 4 MHz conventional transducer mounted on a Plexiglas wedge with a wedge angle of about 46° that generates a 60° shear waves in steel. The index point is located at 60 mm from the weld center line such that this channel covers the lower volume of the weld and detects flaws in this area such as longitudinal cracks. The beam generated by the sensor is illustrated in the figure below and the coloured chart shows that a reasonable energy reaches a large part of the targeted zone. It can be seen that such information could be really useful to optimize transducer design, choice, or its location, in order to reach a certain beam, spread and angle in the target area.

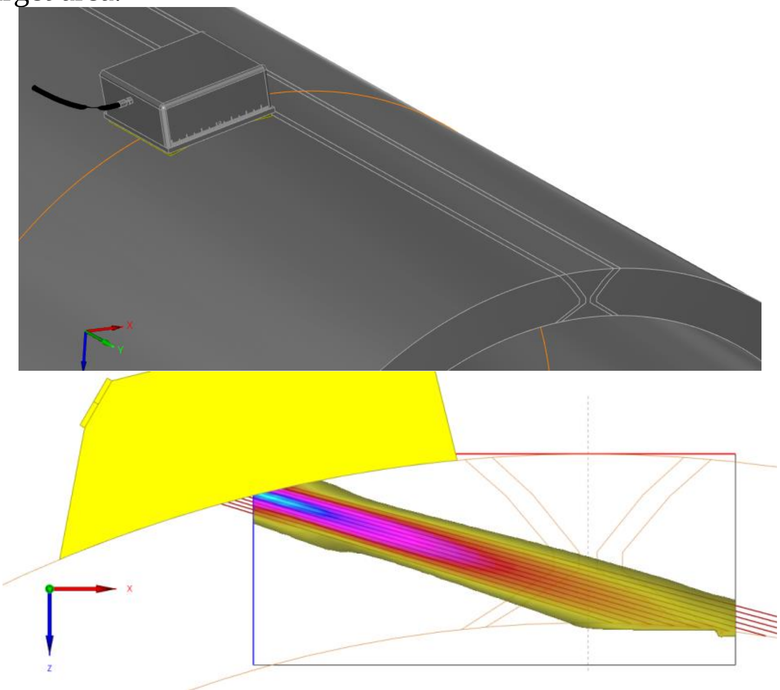


Figure 2. Model of the AUT system (L7 channel) in CIVA user interface and UT field (above-12dB)

Following the same process as real trials, the calibration reflector response shall also be simulated to adjust sensitivity levels to a normalized amplitude to establish decision thresholds in the model. Here, the reflector used for calibration are Through Drilled Holes (TDH) of 1.6 mm diameter. For this channel L7, a TDH located at the centre of the weld is used. Figure 3 illustrates the TDH response signal obtained by simulation. Then, the target flaw will be inserted, modelled here, by a rectangular planar flaw located in the target weld zone with a longitudinal orientation.

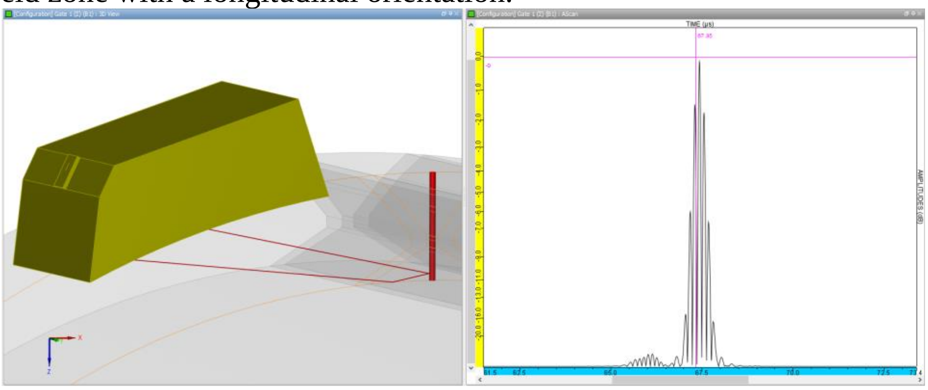


Figure 3. Response of the calibration reflector (Hole of 1.6mm diameter)

4. Weld Inspection POD study

4.1 Methodology

Determination of POD curves via a purely experimental approach requires large-scale experiments performed on representative test-blocks containing representative defects.

For instance, the Military Handbook 1823a [7] states that a minimum amount of 40 different defect locations shall exist in the trial mock-ups when a Signal response analysis is performed, and this minimum is 60 for a Hit-Miss analysis. Moreover, to be representative of a "real POD", this experiment shall "capture" the variability of the influential parameters in real inspections.

Recently, efforts have been made to fix a recognized methodology for the use of numerical simulation to determine POD curves ("MAPOD") and it is worth mentioning practical recommendations published in 2016 by International Institute of Welding [8]. The methodology, as described in this document, aims at using a numerical model which simulates the results of an inspection to reproduce the impact of the variability of influential parameters on the NDE response. The key idea consists in introducing variations in the model input parameters which lead to the variability on the output of the simulation. This variability is then analysed to calculate a POD curve. The estimation of a POD curve by simulation requires:

- To define a "nominal" configuration, that is all the parameters needed for simulating one inspection. From this nominal configuration are derived the configurations which will be computed by considering the variability of some inputted parameters. At this stage, it is important to carefully check input data to avoid any mistakes and this is not an easy task to have a clear and exhaustive list for all the relevant parameters. It is also important to adjust the accuracy of the model (mesh density, level of accuracy needed for the test piece geometry and materials compared to the "real case") to find a good compromise between results accuracy and computation times. Some relevant reference measurements can be very useful at this stage to make good choices.
- To define the characteristic parameter "a" (versus which the POD (a) is calculated),
- To define the "aleatory parameters" whose variability will be taken into account,
- To assign a statistical distribution to these parameters,
- To sample the statistical distributions of aleatory parameters and run the corresponding simulations,
- To compute the POD curve from the set of simulated cases.

In this study, the characteristic parameter will be the flaw height. Regarding the variable parameters, it is not obvious to define in advance the most relevant ones and to define their variability whereas this is the key aspect of a POD analysis. This is where the possibility to rely on parametric studies and metamodels is really interesting. Indeed, it will let us first select a list of candidate essential variables to assess their impact on our outcome. And it is not necessary anymore to postulate a priori a certain variability to run the simulations. Instead of this, the range of variation just need to be defined and then the user will be able to explore different scenarios with the combined variable parameters (i.e., different statistical distributions, different sampling) without launching new computations. After this first assessment, the variable list can be reduced for the final POD analysis.

4.2 Sensitivity Analysis

For the present research sensitivity analysis, first, a quite long list of 13 potential influential parameters has been established:

- Target Defect parameters (rectangular planar shape): flaw height, flaw length, flaw position in the weld defined by 2 parameters: ligament (vertical distance from the backwall) and axial position in the weld volume (which means an angular position in the case of a longitudinal weld), flaw orientations (with the 3 angles that define its orientation: tilt, skew and squint),
- Specimen parameters: Wall thickness, Pipe Outer Diameter, Shear Waves velocity,
- Inspection parameters: probe position versus the weld centre line (“StandOff”), Refraction angle, Squint angle.

To explore the impact of parameters on an outcome, different methodologies are available. Exploring them one by one individually while fixing the other parameters to a certain value can be a solution but it will ignore any inter-dependencies between these variables, the so-called “interactions”. The choice has been to combine them in multiparametric DOEs to explore interactions. However, as it is of course necessary to include several values of each parameter in the list, it is not possible to include all variables in the same computations as it can need a prohibitive number of simulations for 13 variables. In that way, the choice has been made to divide the study into “smaller” DOEs datasets combining 4 to 7 variables together for the sensitivity analysis and also to rely on LHS technique, more efficient than a full factorial one to fill the variable parameters “input space” and then build a metamodel from the simulated database.

Below is illustrated one of this multiparametric study where 4 parameters have been involved in one common LHS DOE made of 400 simulations:
- Flaw height: from 0.35 mm to 2.1mm,
- Probe Refraction Angle: from 57° to 63°,
- Probe Squint Angle: from -5° to +5°,
- Probe standoff: from -17.5° to -12.5° in angular position corresponding to a variation of 70mm to 50mm of the distance to the weld centreline.

On Figure 4, a parallel plot illustrates the simulated design of experiment. The 4 first columns show the sampling performed for the 4 variables while the last column shows the corresponding impact on the outcome which is in this case is the defect response ultrasonic signal maximum amplitude. This graph can be a first resource for a sensitivity analysis but in this case, the goal is to use the metamodel generated from this simulation database.

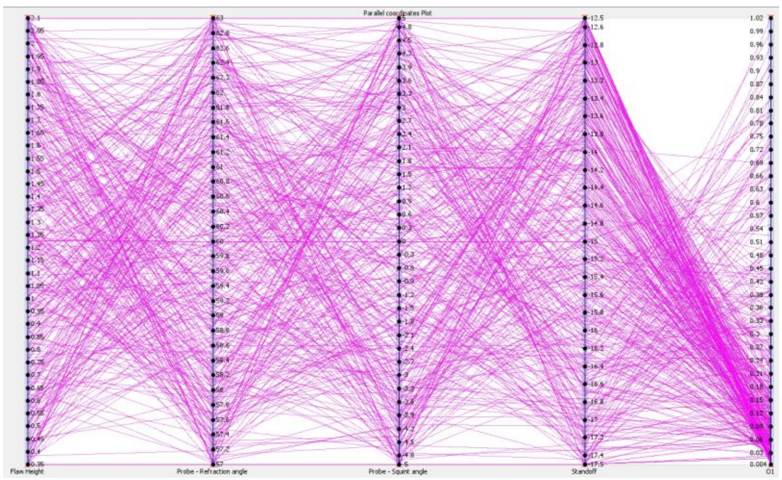


Figure 4. Parallel plot of the DOE performed on 4 variables with 400 simulations.

Before using a metamodel, it is necessary to check its accuracy. To do so, one can rely on cross validation techniques that successively compare the results obtained with metamodels built from different subsets of the simulation database to the remaining partition of this database. It results as “True vs Predicted” graphs that give an overall accuracy of the metamodel for a selected interpolator and can also help to track “where” (i.e., for which parameters’ values) the metamodel is less accurate. Figure 5 shows the accuracy obtained with the Kriging interpolator selected in this case. 99% of the evaluated cases have less than 5% discrepancy with the reference simulations, therefore the metamodel accuracy is estimated as really good here.

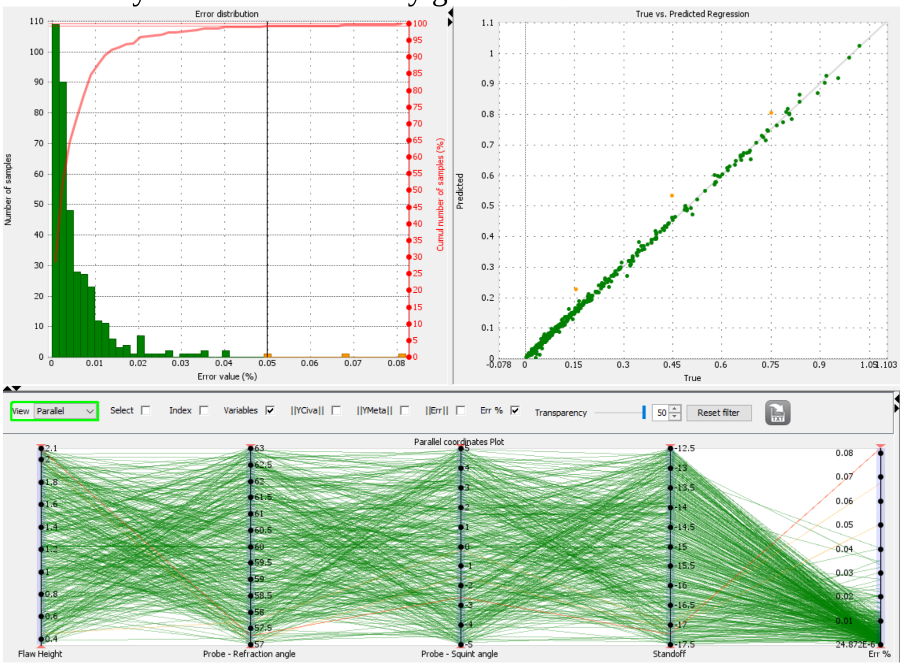


Figure 5. Cross validation technique to estimate metamodel accuracy.

Once validated, you can use the metamodel to analyse parameters impact for any parameter’s values combination or assuming any statistical distribution for parameters’ variabilities in the bounds of the variation range defined. It corresponds to an infinite number of potential inspection scenarios and going from one scenario analysis to another one is done in real-time as it just requires interrogating and/or resample the metamodel but does not need new simulations. This analysis can rely on different types of graphs: 1D plot, 2D response surfaces or parallel plots as seen on Figures 6, 7 and 8:

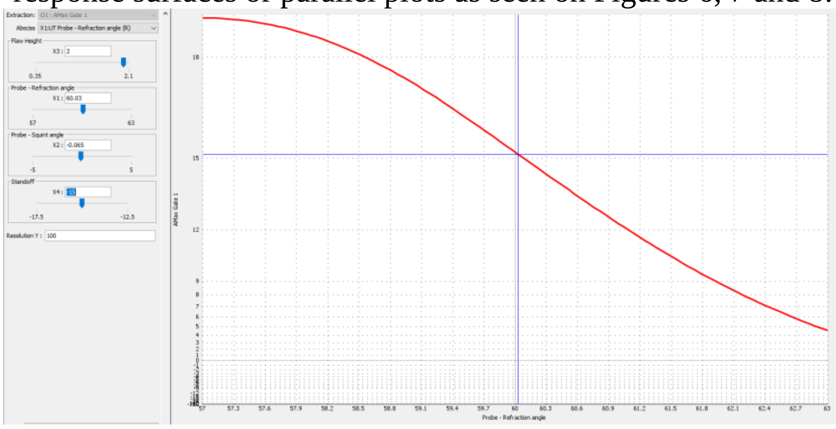


Figure 6. Impact of a refraction angle variation from 57° to 63° on signal amplitude (in dB calibrated versus TDH) for a flaw height of 2mm, probe squint of -2° and probe standoff of 60mm.

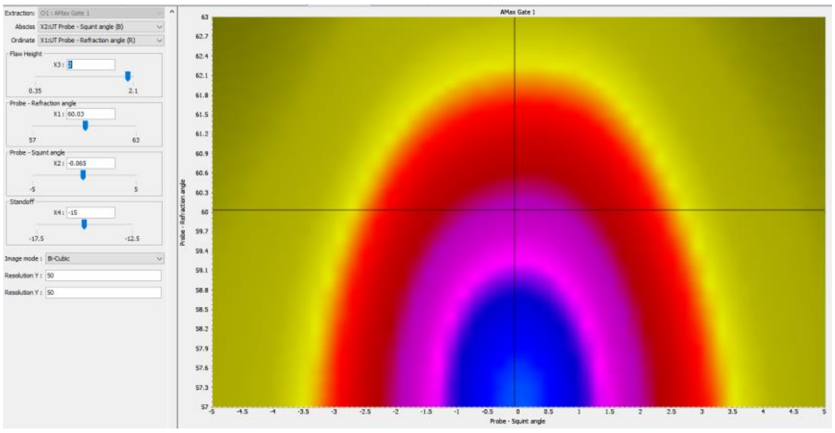


Figure 7. Impact of a refraction angle variation from 57° to 63° and a squint angle variation from -5° to +5° on the signal amplitude for a flaw height of 2mm and probe standoff of 60mm.

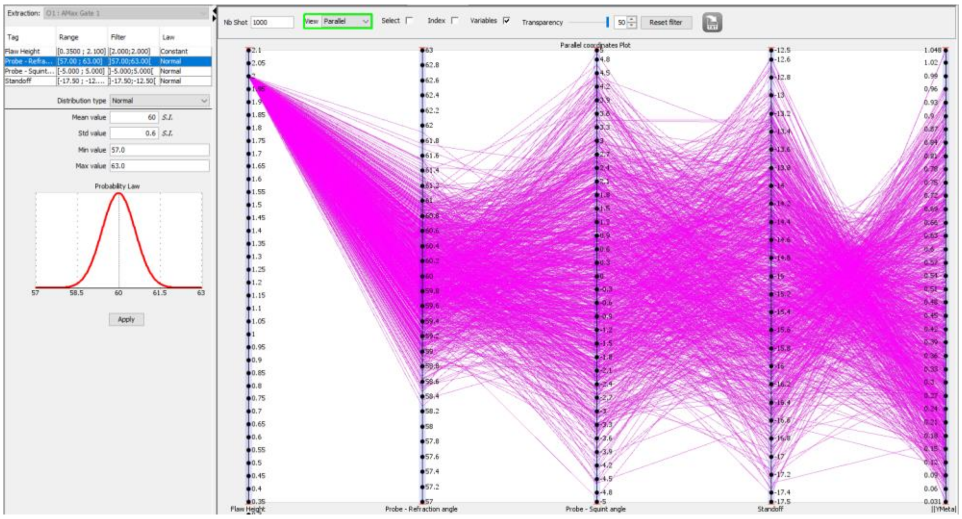
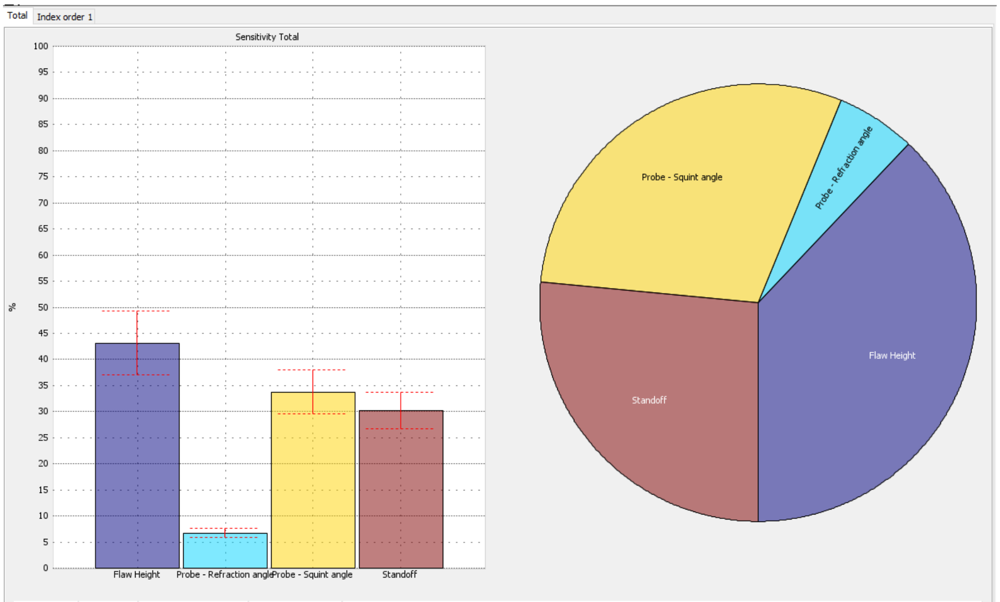


Figure 8. Parallel plot of 1000 inspection scenarios giving the signal amplitude variability (right column) assuming a constant flaw height of 2mm (1st column), and gaussian variability laws for refraction angle, squint angle and probe standoff (2nd to 4th column)

Another output of the sensitivity analysis is the ability to rank the variable impact thanks to the Sobol Indices accounting for parameters variability (defined by probability density functions). Concerning the present case, it gives the histogram on the Figure 9 when it is assumed a uniform law for the flaw height, and a normal law (with a certain mean and standard deviation) for the 3 other parameters. Flaw height, probe standoff and probe squint angles are the most influential ones in this case.

Based on the several metamodels performed to study the 13 parameters identified, a reduced list of influential parameters has been established for the final inspection configuration with only 6 parameters: Flaw height, Flaw positions (ligament and "angular" position), probe squint and refraction angles, probe standoff. A parametric study has been simulated combining these 6 parameters giving a new metamodel. Another sensitivity analysis can be again performed with this new subset of variables. But the other interest is that a POD analysis can be extracted very easily from such a metamodel in CIVA.



	Flaw Height	Probe - Refraction angle	Probe - Squint angle	Standoff
Sensitivity	43.10	6.70	33.67	30.15
Value %	37.94 %	5.90 %	29.63 %	26.53 %
Uncertainty	6.129	0.884	4.234	3.492
Uncertainty %	14.219 %	13.194 %	12.575 %	11.585 %

Figure 9. Sobol Indices ranking influential parameters (flaw height, probe squint and refraction angles, probe standoff) impact on the output signal depending on their variability.

4.3 POD Curves

Different ways exist to build a POD curve. The standard parametric Berens approach will be used here (a non-parametric model is also available in CIVA). A POD analysis starts by building a dataset where the defect responses obtained for different inspection situations are classified versus flaw "true" size (whatever form the defect response is recorded: signal amplitudes, Hit, or Miss). Using simulation allows to have no uncertainty on the flaw size, as this is one of the input parameters of the model. When using metamodels, building this dataset means just asking some samples from the available metamodel and therefore it can be resampled on demand for another scenario (i.e., input parameters variability) without new simulations if it remains in the bounds of the variation range initially defined for the simulation database that built this metamodel.

First, to be representative of the experimental data set available, 300 points are built for 4 different defect sizes [0.35 mm; 0.7 mm; 1.4 mm; 2.1 mm]. A normal law is assumed to describe the variability of inspection parameters (standoff, refraction angle, squint angle) around their mean values and a uniform law for the flaws position parameters (ligament and angular position). The decision threshold is established versus the TDH reference amplitude. A "signal response" POD curve model (or â vs a) has been selected because the underlying statistical hypotheses are reasonably fulfilled for this first data set. This simulated POD curve is compared to the experimental one (analysed in the mh1823POD software [9]) and a good agreement is obtained on the relevant indicator a50, a90 (flaw size for which the probability of detection is 50%, respectively 90%) and a90/95 (including the 95% confidence band) which gives credit to the model, as show on Table 2 and Figure 10.

Table 2. POD indicators

Type of Data	POD indicators		
	a50 (mm)	a90 (mm)	a90/95 (mm)
Simulated data (300 points with 4 defect sizes)	1.248	1.88	2.011
Experimental data (300 points with 4 defect sizes	1.366	1.892	1.961

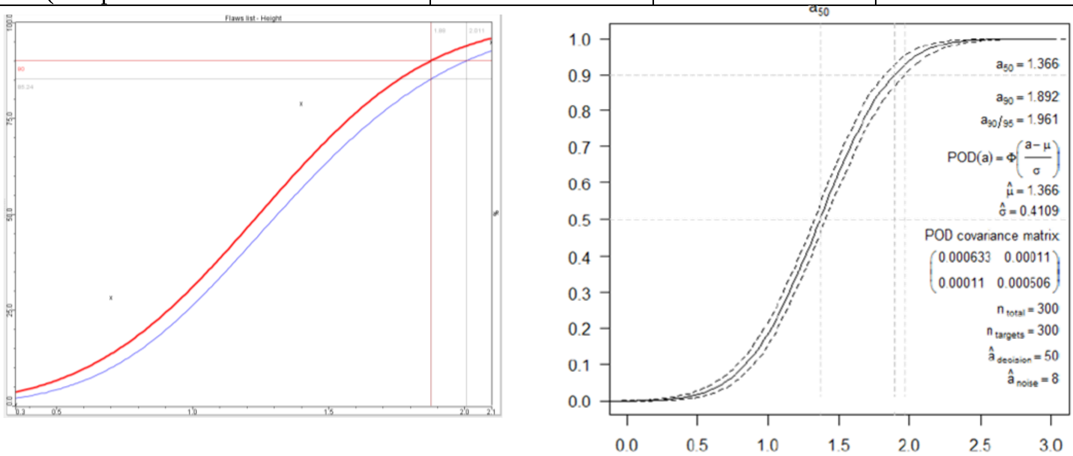


Figure 10. Signal Response POD curves based on a metamodel data set of 300 points (on the left) and experimental POD curves (on the right).

Even if the total number of points looked sufficient to build a POD curve, the experimental data set still suffers from a relative lack of flaw size distribution since only 4 sizes exist and actually only 3 are in the "interesting" part of the POD curve (the rising slope) such that it can still be estimated that a lack of sampling affects the whole accuracy of the POD curve. Now that metamodel shows to be reliable, it is easy and immediate to generate a new sampling with more flaw sizes. This will give more confidence to this curve and will also reduce the confidence band width. Figure 11 shows the POD curve with 36 different flaw sizes regularly distributed (in the same interval) and with 75 tests per flaw size leading to a data set of 2700 points … at no cost. The indicator a90/95 stays close to 2 mm with this new curve, but a more adapted sampling gives more confidence to this result. A Hit-Miss parametric POD curve is here selected since the Signal Response hypotheses are not fulfilled anymore, as often encountered.

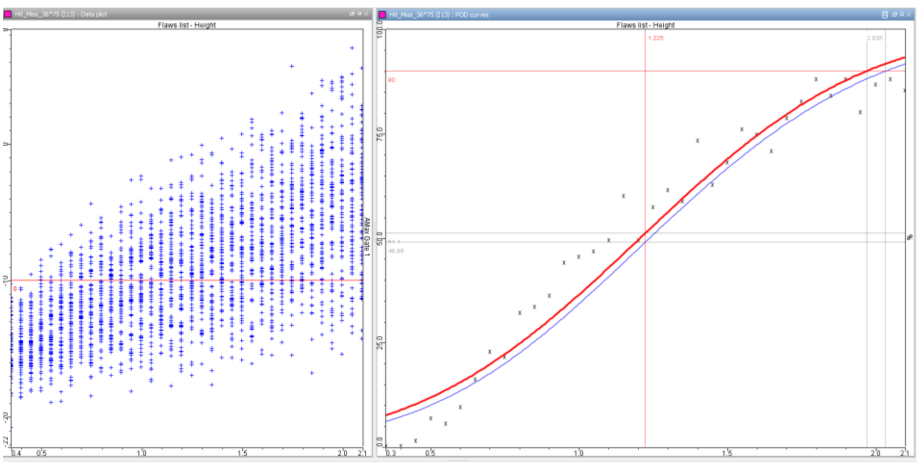


Figure 11. Data points (on the left) and Hit Miss POD curve based on a metamodel data set of 2700 points.

4.4 POD Curves for other scenarios

As the major potential error done in the MAPOD approach comes from the difficulty to describe the variability of the uncertain parameters, metamodels resampling feature allows to try different plausible scenarios to evaluate how sensitive the POD curve is to this uncertainty. It also allows to test inspection conditions that could optimize the flaw detectability on the figure 12, a new POD curve is built assuming a better inspection system control resulting to a standard deviation divided by 2 for all inspection parameters: probe standoff, squint and refraction angles variability. The a90/95 value then decreases from 2mm to 1.6mm.

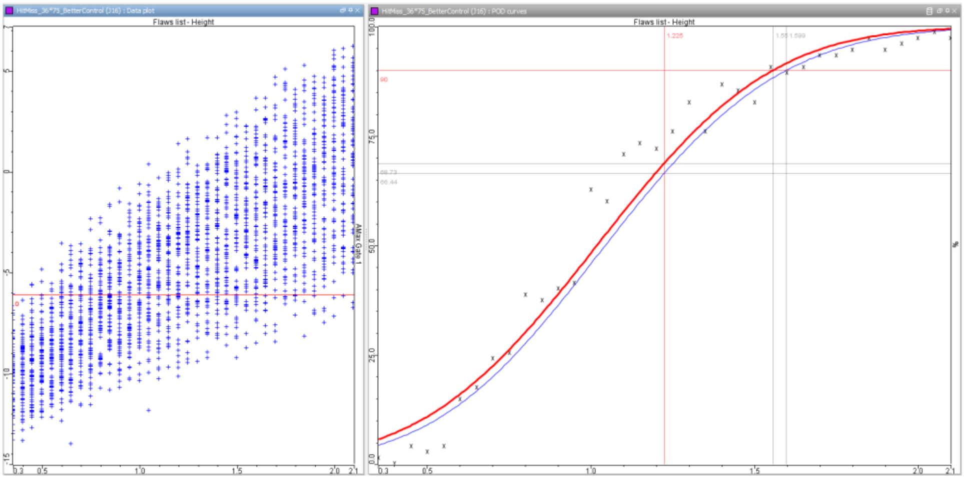


Figure 12. Data points and Hit Miss POD curve with less variability on the inspection parameters.

In the case of the multichannel AUT system, it is possible to include in the same analysis, the observations coming from the evaluated channel as well as its "adjacent" channels as they may also detect the flaws in the targeted zone. Here, L3 channel data has been added to the L7 one, and the maximum obtained between both channels (for the considered defects distribution in L7 zone) is kept assessing the detectability. In Figure 13, the red crosses correspond to the situation where L3 detects flaws with a higher amplitude than the L7 channel. Even if this might not be a sign of a well-suited inspection set-up in this case, a lot of flaws are actually better detected by L3, and a90/95 decreases from about 2mm to 1.3mm!

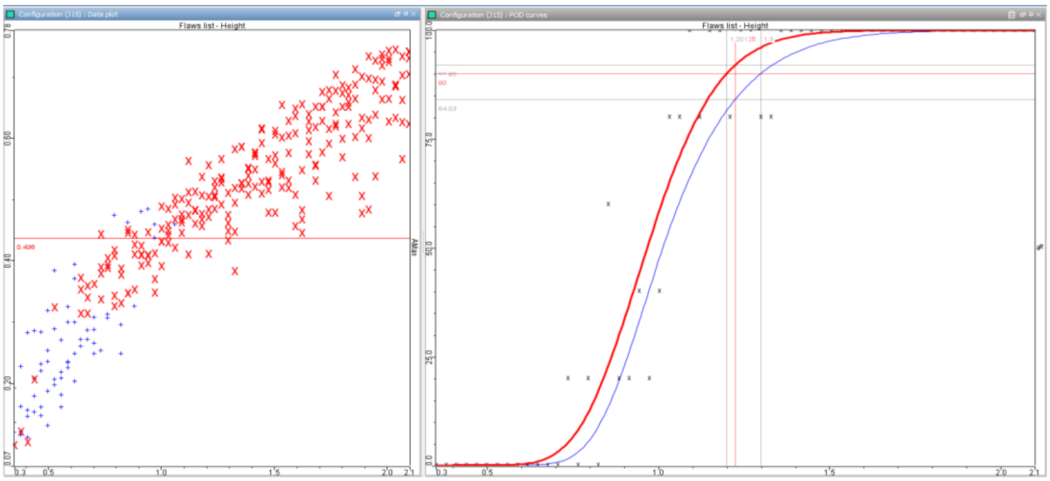


Figure 13. POD curve with L7 and L3 channels response

5. Conclusions

Modelling can greatly help reliability studies because it provides an efficient way to generate and explore many inspection and defect scenarios at low cost compared to a purely experimental approach based on mock-up tests. The availability of metamodels further enforces the ability to generate large enough sample sizes for statistical confidence and analyse deep degradation factors or optimization solutions. To allow quantitative use of such solutions, validation references and a controlled methodology must be considered. This paper has illustrated such type of reliability study for an AUT Inspection of longitudinal weld on carbon steel pipes with the simulation software CIVA.

References and footnotes

[1] https://www.extende.com

[2] S. Mahaut, S. Chatillon, M. Darmon, N. Leymarie and R. Raillon, “An overview of UT beam propagation and flaw scattering models in CIVA”, (2009), QNDE.

[3] A. Imperiale, E. Demaldent, “A macro-element strategy based upon spectral finite elements and mortar elements for transient wave propagation modeling. Application to ultrasonic testing of laminate composite materials”, International Journal for Numerical Methods in Engineering, 964–990 (2019).

[4] F. Foucher et al, “An overwiew of validation campaigns of the CIVA simulation software”, proceedings of ECNDT 2018

[5] R. Miorelli et al, Database generation and exploitation for efficient and intensive simulation studies, AIP Conference Proceedings, Volume 1706 (2016)

[6] M. Burrowes, G. R. Pereira, L. M. Tavares, Validation and Calibration of Computer-simulated Probability of Detection Curves from Ultrasonic Inspection, Insight - Non-Destructive Testing and Condition Monitoring, Volume 61, Number 4, April 2019

[7] MIL-HDBK-1823A, C. Annis and US Department of Defense, 2009

[8] B. Chapuis, P. Calmon, F. Jenson, 2016, Best practices for the use of Simulation in POD Curves Estimation, IIW Collection, https://doi.org/10.1007/978-3-319-62659-8

[9] https://statistical-engineering.com