

Model Assisted Probability of Detection Curves: New Statistical Tools and Progressive Methodology

Loïc Le Gratiet¹ · Bertrand Iooss¹ · Géraud Blatman² · Thomas Browne¹ · Sara Cordeiro³ · Benjamin Goursaud⁴

Received: 21 January 2016 / Accepted: 6 December 2016 / Published online: 19 December 2016
© Springer Science+Business Media New York 2016

Abstract The probability of detection curve is a standard tool in several industries to evaluate the performance of non destructive testing (NDT) procedures for the detection of harmful defects for the inspected structure. Due to new capabilities of NDT process numerical simulation, model assisted probability of detection (MAPOD) approaches have also been recently developed. In this paper, a generic and progressive MAPOD methodology is proposed. Limits and assumptions of the classical methods are enlightened, while new metamodel-based methods are proposed. They allow to access to relevant information based on sensitivity analysis of MAPOD inputs. Applications are performed on eddy current non destructive examination numerical data.

Keywords Non destructive testing · Eddy-current · Computer experiments · Uncertainty · Sensitivity analysis · Kriging

1 Introduction

In several industries, the probability of detection (POD) curve is a standard tool to evaluate the performance of non destructive testing (NDT) procedures [12, 15, 22]. The goal is to assess the quantification of inspection capability for the detection of harmful flaws for the inspected structure. For instance, for the French company of electricity (EDF), the potentialities of this tool are studied in the context of the eddy current non destructive examination in order to ensure integrity of steam generators tubes in nuclear power plants [20].

However, high costs of the implementation of experimental POD campaigns combined with continuous increase in the complexity of configuration make them sometimes unaffordable. To overcome this problem, it is possible to resort to numerical simulation of NDT process (see for example [25] for ultrasonics and [24] for eddy-current). This approach has been called MAPOD for “Model Assisted Probability of Detection” [30] (see also [22] for a survey and [8] for a synthetic overview).

The determination of this “numerical POD” is based on a four-step approach:

1. Identify the set of parameters that significantly affect the NDT signal;
2. Attribute a specific probability distribution to each of these parameters (for instance from expert judgment);
3. Propagate the input parameters uncertainties through the NDT numerical model;

✉ Bertrand Iooss
bertrand.iooss@edf.fr

Loïc Le Gratiet
loic.le-gratiet@edf.fr

Géraud Blatman
geraud.blatman@edf.fr

Thomas Browne
thomas.browne@edf.fr

Sara Cordeiro
sara.cordeiro@edf.fr

Benjamin Goursaud
benjamin.goursaud@edf.fr

¹ EDF R&D, EDF Lab Chatou, 6 Quai Watier, 78401 Chatou, France

² EDF R&D, EDF Lab les Renardières, Écuellles, France

³ EDF CEIDRE, Saint Denis, France

⁴ EDF R&D, EDF Lab Paris-Saclay, Palaiseau, France

- Build the POD curve from standard approaches like the so-called Berens method [4]. In POD studies, two main models are used: POD model for binary detection representation (using hit/miss data) and POD model for continuous response (using the values of the NDT signal). We focus in this work on POD model for continuous response, arguing that model-based data contain quantitative and precise information on the signal values that will be better exploited with this approach.

Of course, this process is conditioned by the quality of the NDT model, which has to accurately represent the inspection technique under evaluation. This quality can be judged by the so-called model verification and validation step [23], that is considered as a preliminary step of our four step process. Verification is testing the correctness of the software execution and the equation solvers inside the model, while validation evaluates the discrepancy between the model outputs and some measurements coming from representative (of the domain of interest) experiments. For the NDT system presented in this article, before the propagation of the input parameters uncertainties (step 3), the computer code (C3D) has been successfully verified and compared with experimental data on about 100 typical defects [20,29].

As it totally relies on a probabilistic modeling of uncertain physical variables and their propagation through a model, the MAPOD approach can be directly related to the uncertainty management methodology in numerical simulation (see [10] and [2] for a general point of view, and [13] for illustration in the NDT domain). This methodology proposes a generic framework of modeling, calibrating, propagating and prioritizing uncertainty sources through a numerical model (or computer code). Indeed, investigation of complex computer code experiments has remained an important challenge in all domain of science and technology, in order to make simulations as well as predictions, uncertainty analysis or sensitivity studies. In this framework, the numerical model G just writes

$$Y = G(X) = G(X_1, \dots, X_d), \quad (1)$$

with $X \in \mathbb{R}^d$ the random input vector of dimension d and $Y \in \mathbb{R}$ a scalar model output.

However, standard uncertainty treatment techniques require many model evaluations and a major algorithmic difficulty arises when the computer code under study is too time expensive to be directly used. For instance, it happens for NDT models based on complex geometry modeling and finite-element solvers. This problem has been identified in [8] who distinguishes “semi-analytical” codes (fast to evaluate but based on simplified physics) and “full numerical” ones (physically realistic but cpu-time expensive) which are the models of interest in our work. For cpu-time expensive models, one solution consists in replacing the numerical model

by a mathematical approximation, called a response surface or a metamodel. Several statistical tools based on numerical design of experiments, uncertainty propagation efficient algorithms and metamodeling concepts will then be useful [14]. They will be applied, in this paper, in the particular NDT case of a POD curve as a quantity of interest.

The physical system of interest, the numerical model parameterization and the design of numerical experiments are explained in the following section. The third section introduces four POD curves determination methods: the classical Berens method, a binomial-Berens method and two methods (polynomial chaos and kriging) based on the metamodeling of model outputs. In the fourth section, sensitivity analysis tools are developed by using the metamodel-based approaches. A conclusion synthesizes the work with a progressive strategy for the MAPOD process, in addition to some prospects.

2 The NDT System

Our application case, shown in Fig. 1, deals with the inspection by the SAX probe (an axial probe) of steam generator tubes to detect the wears, which are defects due to the rubbing of anti-vibration bars (AVB). This configuration has been studied with Code_Carmel3D (C3D) for several years and C3D has demonstrated its ability to accurately simulate the signature of a wear, with its influential parameters (mainly the AVB). Besides, C3D has been involved in several benchmarks between numerical tools and experimental data [21].

2.1 The Computer Cand Model Parameterization

The numerical simulations are performed by C3D, computer code derived from code_Carmel developed by EDF R&D and the L2EP laboratory (Lille, France). This code uses the finite element method to solve the problem. Hence, there is a large flexibility for the parameters that can be taken into account

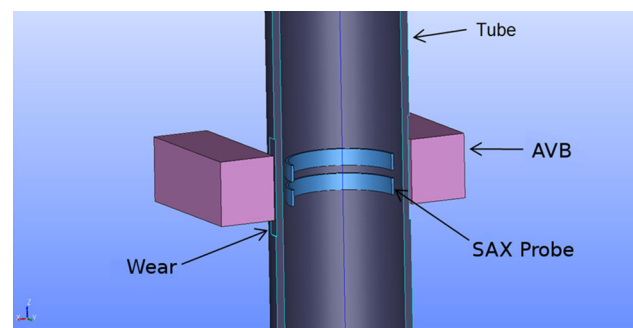


Fig. 1 Representation of the system under study (Tube with wear, AVB and SAX)

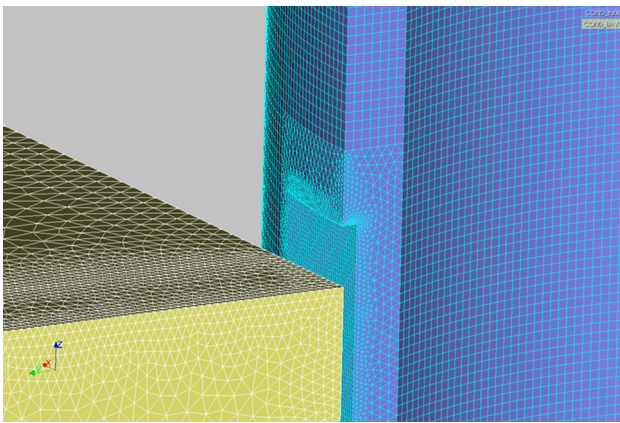


Fig. 2 Illustration of the mesh in the numerical model of NDT simulation

(cf. Fig. 2). The accuracy of the calculations can be ensured with a sufficiently refined mesh [29], using HPC capabilities if necessary.

The eddy-current non-destructive examinations are based on the change of the induction flux in the coils of the probe approaching a defect. When the tube is perfectly cylindrical, both coils of the probe get the same flux of induction. If there is a defect, the flux are distinct and hence the differential flux, which is the difference between the flux in each coil, is non-zero: it is a complex quantity whose real part is the X channel and the imaginary part is the Y channel. Hence, when plotting the differential flux for each position of the probe, one gets a curve in the impedance plane, called a Lissajous curve. The output parameters of a non-destructive examination are (as illustrated on Fig. 3):

- the amplitude (amp), which is the largest distance between two points of the Lissajous curve,
- the phase, which is the angle between the abscissa axis and the line linking two points giving the amplitude,
- the Y -projection ($ProjY$), which is the largest imaginary part of the difference between two points of the Lissajous curve.

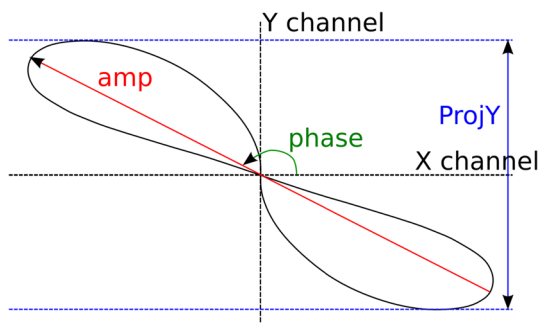


Fig. 3 Lissajous curve: output parameters of a NDT simulation for the SAX probe in differential mode (amp , $phase$ and $ProjY$)

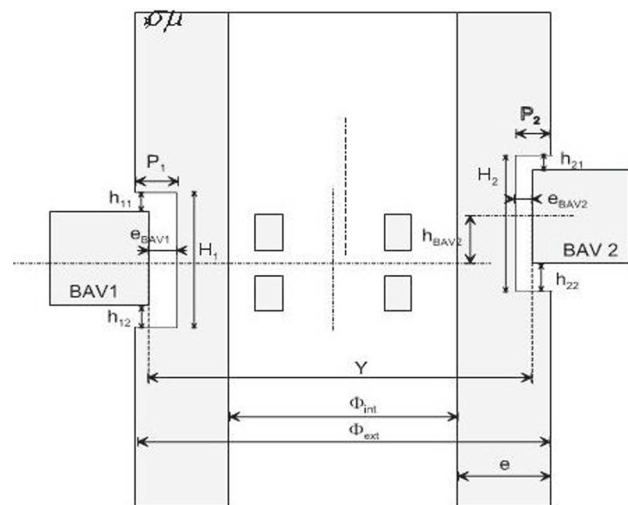


Fig. 4 Illustration of the considered inputs

2.2 Input Parameters and Associated Random Distributions Definition

By relying on both expert reports and data simulations, the set of the input parameters which can have an impact on the code outputs have been defined. Probabilistic models have also been proposed following deep discussions between NDT experts and statisticians. $\mathcal{N}(\cdot, \cdot)$ (resp. $\mathcal{U}[\cdot, \cdot]$) stands for Gaussian (resp. uniform) law. These parameters are the following (see Fig. 4):

- $E \sim \mathcal{N}(a_E, b_E)$: pipe thickness (mm) based on data got from 5000 pipes,
- $h_1 \sim \mathcal{U}[a_{h_1}, b_{h_1}]$: distance between the AVB and the top of the first wear (mm),
- $h_2 \sim \mathcal{U}[a_{h_2}, b_{h_2}]$: distance between the AVB and the bottom of the first wear (mm),
- $P_1 \sim \mathcal{U}[a_{P_1}, b_{P_1}]$: first flaw depth (mm),
- $P_2 \sim \mathcal{U}[a_{P_2}, b_{P_2}]$: second flaw depth (mm),
- $ebav_1 \sim \mathcal{U}[-P_1 + a_{ebav_1}, b_{ebav_1}]$: length of the gap between the AVB and the first flaw (mm),
- $ebav_2 \sim \mathcal{U}[-P_2 + a_{ebav_2}, b_{ebav_2}]$: length of the gap between the AVB and the second flaw (mm).

All these input parameters are synthesized in a single input random vector $(E, h_1, h_2, P_1, P_2, ebav_1, ebav_2)$.

As displayed in Fig. 4, we consider the occurrence of one flaw (or wear) on each side of the pipe due to AVB. To take this eventuality into account in the computations, 50% of the experiments are modeled with one flaw, and 50% with two flaws.

2.3 Definition of the Design of Numerical Experiments

In order to compute the output of interest with C3D, it is necessary to choose the points in the variation domain of the

inputs (called the input set). This dataset, called “design of experiments”, has to be defined at the very beginning of the study, which is to say before any numerical simulation. A classical method consists in building the design of experiments by randomly picking different points of the input set, obtaining a so-called Monte Carlo sample. However, a random sample can lead to a design which does not properly “fill-in” the input set [14]. A better idea would be to spread the numerical simulations all over the input set, in order to avoid some empty big subsets.

To this effect it is more relevant to choose the values according to a deterministic rule, such as a quasi-Monte Carlo method, for instance a Sobol’ sequence. Indeed, for a size of design N , it is proved that this design often happens to be more precise than the standard Monte Carlo method [14]. Given the available computing time (several hours per model run), a Sobol’ sequence of size 100 is created, and 100 model outputs are obtained after the computer code (G) runs.

3 Methods of POD Curves Estimation

In this section several methods (from the simplest relying on strong assumptions to the most complex) are presented and applied. The objective is to build the POD curve as a function of the main parameter of interest, related to the defect size. As there are two defects in the system, $a := \max(P_1, P_2)$ is chosen as the parameter of interest.

By using the computer code C3D, one focuses on the output $ProjY$ which is a projection of the simulated signal we would get after NDT process. The other inputs are seen as random variables, which makes $ProjY$ itself an other random variable. The model (1) writes now

$$ProjY = G(a, X), \tag{2}$$

with the random vector $X = (E, h_1, h_2, ebav_1, ebav_2)$. The effects of all the input parameters (a, X) are displayed in Fig. 5. The bold values are the correlation coefficients between the output $ProjY$ and the corresponding input parameter. Strong influences of P_1 and P_2 on $ProjY$ are detected, as it could be expected. i_{P_2} is the binary variable governing the presence of one flaw ($i_{P_2} = 1$) or two flaws ($i_{P_2} = 2$).

Given a threshold $s > 0$, a flaw is considered to be detected if $ProjY > s$. Therefore the one dimensional POD curve is denoted by:

$$\forall a > 0 \quad \text{POD}(a) = \mathbb{P}(G(a, X) > s \mid a). \tag{3}$$

Four different regression models of $ProjY$ are proposed in the following, in order to build an estimation of the POD

curve. Numerical simulations are computed for the $N = 100$ points of the design of experiments.

3.1 Data Linearization Step

All the POD methods consist in a (linear or non-linear) regression of the output $ProjY$. Then, a data linearization is useful to improve the adequacy of the models. This can be made by a Box–Cox transformation [6] of the output, which means that we now focus on:

$$y = \frac{ProjY^\lambda - 1}{\lambda}. \tag{4}$$

λ is determined by maximum likelihood as the real number that offers the finest linear regression of y regarding the parameter a (see Fig. 6). The same transformation has to be applied to the detection threshold s . In the following, we keep s for the notation of this threshold. It is important to note that this transformation is useful for all the different POD methods [13].

3.2 Berens Method [4]

The Berens model, based on y , is defined as

$$y(a) = \beta_0 + \beta_1 a + \epsilon, \tag{5}$$

with ϵ the model error such as $\epsilon \sim \mathcal{N}(0, \sigma_\epsilon^2)$. Maximum likelihood method provides the estimators $\hat{\beta}_0, \hat{\beta}_1$ and $\hat{\sigma}_\epsilon$. Hence the model implies the following result: $\forall a > 0, y(a) \sim \mathcal{N}(\hat{\beta}_0 + \hat{\beta}_1 a, \hat{\sigma}_\epsilon^2)$. On our data, we obtain $\hat{\beta}_0 = 2.52, \hat{\beta}_1 = 43.48$ and $\hat{\sigma}_\epsilon = 1.95$, which leads to the linear model represented in Fig. 7.

With the normality hypothesis, as displayed in Fig. 7, the values of the POD curve can be easily estimated, giving the POD curve of Fig. 8. By considering the error that is provided by the property of a maximum likelihood estimator in a case of a linear regression, we can use this uncertainty on both β_0 and β_1 to build non-asymptotic confidence intervals. Indeed, the Gaussian hypothesis on ϵ makes it possible to obtain the prediction law of β_0 and β_1 conditionally to σ_ϵ^2 :

$$\begin{pmatrix} \beta_0 \\ \beta_1 \end{pmatrix} \sim \mathcal{N}\left(\hat{\beta} = \begin{pmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \end{pmatrix}, \sigma_\epsilon^2 (\mathbf{X}^T \mathbf{X})^{-1}\right), \tag{6}$$

with \mathbf{X} the data input matrix:

$$\mathbf{X} = \begin{pmatrix} 1 & a_1 \\ 1 & a_2 \\ \vdots & \vdots \\ 1 & a_N \end{pmatrix}.$$

Fig. 5 *ProjY* with respect to the input parameters. On each plot, the *solid curve* is a local polynomial smoother and the *upper number* is the corresponding correlation coefficient between the input (in abscissa) and *ProjY* (in ordinate)

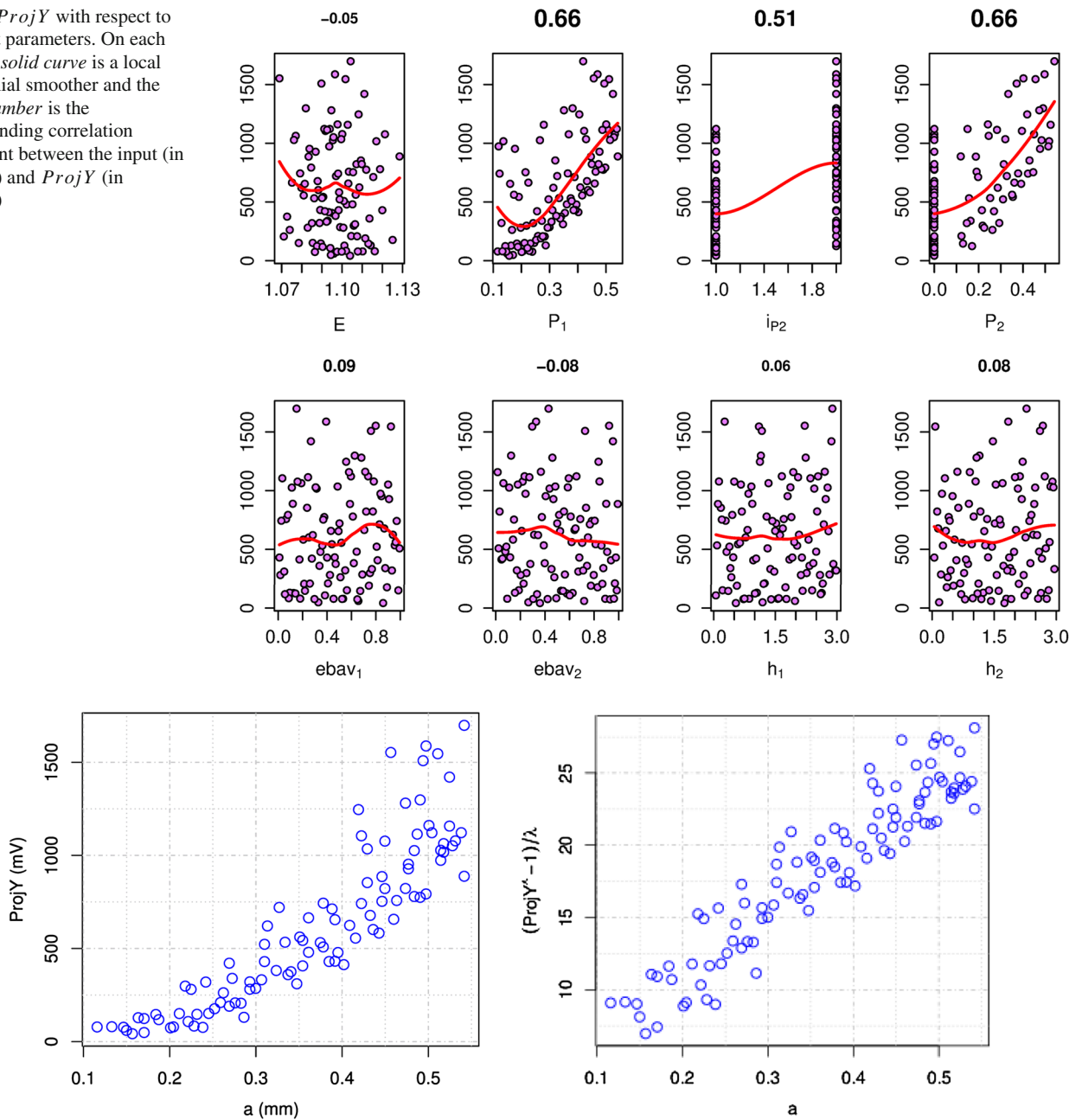


Fig. 6 Model response with respect to *a*. *Left*: Initial data (*ProjY* as response); *Right*: Linearized data (y_{ProjY} as response) by Box-Cox transformation with parameter $\lambda = 0.3$ of the response *ProjY*

Classical results on linear regression theory state that the variance σ_ϵ^2 follows a chi-2 distribution with $N - 2$ degrees of freedom:

$$\frac{(N - 2)\hat{\sigma}_\epsilon^2}{\sigma_\epsilon^2} \sim \chi_{N-2}^2, \tag{7}$$

where

$$\hat{\sigma}_\epsilon^2 = \frac{(y^N - \mathbf{X}\hat{\beta})^T (y^N - \mathbf{X}\hat{\beta})}{N - 2}, \tag{8}$$

with $y^N = (y(a_1), \dots, y(a_N))$ the data output sample. Then, we can obtain a sample $(\beta_0, \beta_1, \sigma_\epsilon^2)$ by simulating σ_ϵ^2 then (β_0, β_1) conditionally to σ_ϵ^2 . From this sample, we get a sample of $POD(a)$ via the formula:

$$1 - \Phi((s - \beta_0 - \beta_1 a)/\sigma_\epsilon), \tag{9}$$

where Φ is the standard Gaussian distribution. By simulating a large number of POD samples, we can deduce some confidence intervals. The 95%-confidence lower bound of the POD curve is illustrated in Fig. 8.

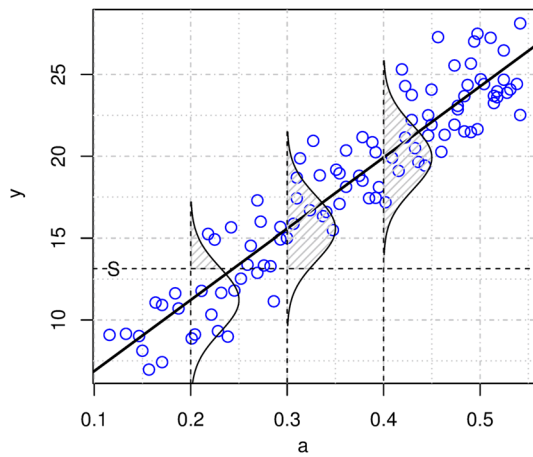


Fig. 7 Linear model illustration. The Gaussian predictive distributions for $a = \max(P_1, P_2) = 0.2, 0.3$ and 0.4 are given. The horizontal line represents the detection threshold s

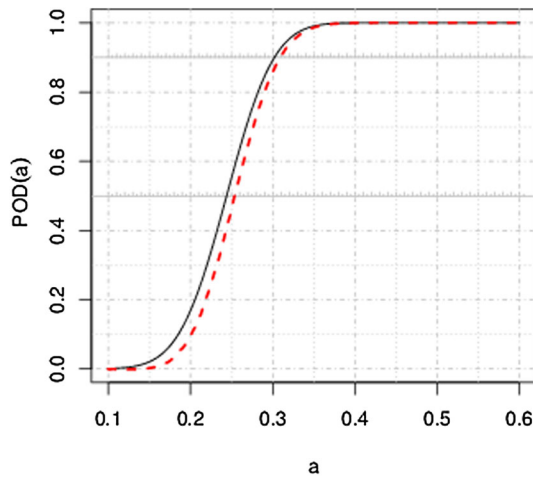


Fig. 8 Results of Berens method: POD curve estimation (solid curve) and POD lower curve (dashed curve) of the POD 95%-confidence interval

From the estimated POD of Fig. 8, we obtain $a_{90} \simeq 0.30$ mm for the defect size detectable with a 90%-probability. Taking into account the confidence interval, we obtain $a_{90/95} \simeq 0.31$ mm for the defect size which at least 90% of the cracks can be detected, established with 95% confidence. In other words, $a_{90/95}$ purports to be the size of the target having at least 90% probability to be detected in 95% of the POD experiments under nominally identical conditions.

In conclusion, we remind that the Berens method stands on two hypotheses that have to be validated:

- the linearity relation between y and a (after the Box–Cox transformation) that can be studied via classical linear regression residuals analysis [9]. On our data, we have for instance $R^2 = 88\%$ for the regression coefficient

of determination, indicator which denotes the explained variance of the linear regression;

- the Gaussian distribution, homoscedasticity and independence of the residuals that can be studied via many statistical tests (see for instance [31]). On our data, we have the following p-values: 0.62 for Kolmogorov–Smirnov test (Gaussian distribution), 0.10 for Anderson–Darling test (Gaussian distribution), 0.82 for Breusch–Pagan test (homoscedasticity) and 0.12 for Durbin–Watson test (non correlation). We conclude that, with a 90%-confidence, the homoscedasticity and non-correlation hypotheses of ϵ cannot be rejected, but the normality hypothesis of ϵ can be rejected.

3.3 Binomial-Berens Method

Here we keep the linear regression on y , which is: $\forall a > 0 \quad y = \hat{\beta}_0 + \hat{\beta}_1 a + \epsilon$ but we do not assume that ϵ is Gaussian anymore. However the errors are still assumed to be independent and identically distributed. We then consider that we have N of its realizations which we regroup in the following vector

$$\epsilon^N = y^N - \hat{\beta}_0 - \hat{\beta}_1 a^N. \tag{10}$$

Therefore we build its histogram and we add it to the prediction of the linear model as shown in Fig. 9. By using the i.i.d. property of ϵ , let us consider that we have N realizations of the random value $y(a)$ for $a > 0$. We propose to use them to estimate the probability for $y(a)$ to exceed the threshold s (see Fig. 9).

For each $a > 0$, let $N_s(a)$ be the number of realizations of the random variable $y(a)$ that are higher than s . That is to say:

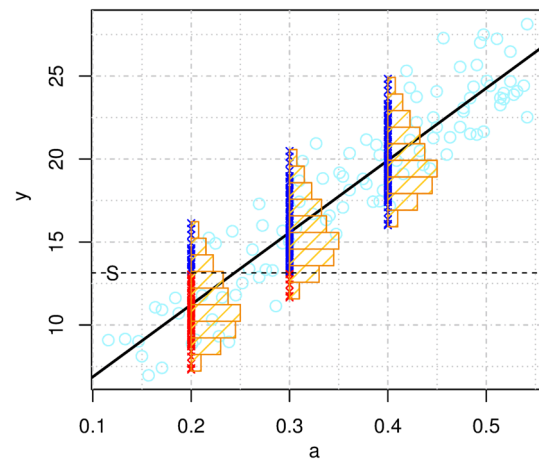


Fig. 9 Binomial-Berens method: Berens method without normal hypothesis. The Gaussian densities are replaced by the sample histogram. The horizontal line represents the detection threshold s

$$N_s(a) = \text{Card} \left(\left\{ (\epsilon_i)_{i \in \{1, \dots, N\}} \mid \hat{\beta}_0 + \hat{\beta}_1 a + \epsilon_i > s \right\} \right). \tag{11}$$

Therefore an estimation of $\text{POD}(a)$ is given by $\frac{N_s(a)}{N}$, with $N_s(a) \sim \mathcal{B}(N, \text{POD}(a))$, with \mathcal{B} the binomial probability law. The assumption on $N_s(a)$ distribution can then be used to build confidence intervals on the value of $\text{POD}(a)$, for $a > 0$.

Let us note that the Binomial-Berens method only requires to validate the linear relation between y and a . For the 90%-level defect, we obtain $a_{90} \simeq 0.30$ mm and $a_{90/95} \simeq 0.305$ mm. A slight difference with the classical Berens method is present for $a_{90/95}$.

3.4 Polynomial Chaos Method

As some criticism could be made at some point regarding the simplistic linear model of equation (5), let us build a meta-model [14] of the transformed output y . Now the influence of the other inputs (described in Sect. 2.2) are explicitly mentioned in the model whereas it used to be all included in ϵ . The model response of interest, e.g. the Y -projection, is represented as a ‘‘pure’’ function of X (i.e. without additional noise):

$$Y = G(a, X). \tag{12}$$

The so-called polynomial chaos (PC) method [5, 28] consists in approximating the response onto a specific basis made of orthonormal polynomials:

$$Y \approx \hat{Y} = \sum_{j=0}^{P-1} a_j \psi_j(a, X), \tag{13}$$

where the ψ_j ’s are the basis polynomials and the a_j ’s are deterministic coefficients which fully characterize the model response and which have to be estimated. The orthonormality property reads:

$$\mathbb{E} [\psi_i(a, X) \psi_j(a, X)] = 1 \text{ if } i = j \text{ else } 0. \tag{14}$$

The derivation of sensitivity indices (see Sect. 4.1) of the response is directly obtained by simple algebraic operations on the coefficients a_j . The latter are computed based on the experimental design and the associated model evaluations by least squares.

PC approximations are computed with several values for the total degree, and their accuracies are compared in terms of predictivity coefficient Q^2 , itself based on the leave-one-out error. The greatest accuracy is obtained with a linear approximation (i.e. with degree equal to one), with $Q^2 = 88\%$. This PC representation reads:

$$\begin{aligned} \hat{Y} \simeq & 27.9 - 0.5 \psi_1(E) + 11.4 \psi_2(a) + 0.7 \psi_3(eba v_1) \\ & + 0.3 \psi_3(eba v_1) + 0.4 \psi_4(h_1) + 1.0 \psi_5(h_2) \quad (\text{mV}). \end{aligned} \tag{15}$$

As in the Berens model in Sect. 3.2, it is assumed that the approximation error is a normal random variable ϵ with zero mean and standard deviation equal to σ_ϵ , that is:

$$Y \approx \hat{Y} + \epsilon, \quad \epsilon \sim \mathcal{N}(0, \sigma_\epsilon^2). \tag{16}$$

Thus the POD associated with a given defect size a can be approximated by:

$$\text{POD}(a) = \mathbb{P}(Y > s \mid a) = \mathbb{P}(\hat{Y}(a, X) + \epsilon > s). \tag{17}$$

For any value of a , this probability is estimated by Monte Carlo simulation of the random quantities X and ϵ (10^4 random values are drawn).

Note that this estimate relies upon the assumption that the chaos coefficients are perfectly calculated. However, their estimation is affected by uncertainty due to the approximation error ($1 - Q^2 = 12\%$ of unexplained variance of the Y -projection) and the limited number of available evaluations of C3D. As for the Berens model, standard theorems related to linear regression hold for the PC expansions and can be used to define the probability distribution of the chaos coefficients and the residual standard deviation σ_ϵ . Based on these results, 150 sets of both quantities are randomly generated and each realization is used to compute the POD (Eq. (17)). Hence, for any a , a sample of 150 values of $\text{POD}(a)$ is obtained. We computed its 5%-empirical quantile in order to construct the 95%-POD curve. The average and the 95%-POD curves are plotted in Fig. 10. The characteristic defect sizes (defined in the previous sections) are given by $a_{90} \simeq 0.30$ mm and $a_{90/95} \simeq 0.32$ mm.

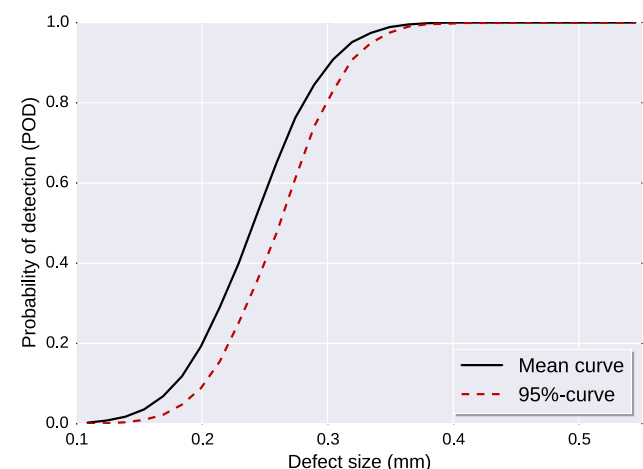


Fig. 10 Average and 95%-POD curves based on a PC approximation

It has to be noted that the chaos results are closed to the ones obtained by the Berens approach. Indeed, the PC representation (15) is similar to the Berens model (5) as all the coefficients except the mean value and the factor related to a are relatively insignificant in our application case. Furthermore, it is also supposed that the residuals are independent realizations of a normal random variable. As discussed previously, this assumption can be rejected by statistical tests. Another kind of metamodel, namely kriging, is based on the weaker and more realistic assumption of correlated normal residuals (the correlation between two model evaluations increases as the related inputs get closer). This is the scope of the next section.

3.5 Kriging Method

We turn now to a probabilistic metamodel technique, which is the Gaussian process regression [26], first proposed by [11] for POD estimation. Since the linear trend used in the Berens method was rather relevant, we keep it as the mean of the Gaussian process that we are about to use. The kriging model is defined as follows:

$$Y(a, X) = \beta_0 + \beta_1 a + Z(a, X), \tag{18}$$

where Z is a centered Gaussian process. We make the assumption that Z is second order stationary with variance σ^2 and covariance Matérn 5/2 parameterized by its lengthscale θ ($\theta \in \mathbb{R}^6$ in our application case). Thanks to the maximum likelihood method, we can estimate the values of the so far-unknown parameters: $\beta_0, \beta_1, \sigma^2$ and θ (see for instance [19] for more details).

Kriging provides an estimator of $Y(a, X)$ which is called the kriging predictor and written $\widehat{Y}_P(a, X)$. On our data, we compute the predictivity coefficient Q^2 in order to quantify the prediction capabilities of this metamodel [19]. We obtain $Q^2 = 90\%$.

In addition to the kriging predictor, the kriging variance $\sigma_Y^2(a, X)$ quantifies the uncertainty induced by estimating $Y_P(a, X)$ with $\widehat{Y}_P(a, X)$. Thus, we have the following predictive distribution:

$$\forall x \quad \left(Y(a, X) \mid y^N \right) \sim \mathcal{N} \left(\widehat{Y}_P(a, X), \sigma_Y^2(a, X) \right), \tag{19}$$

where $\widehat{Y}_P(a, X)$ (the kriging mean) and $\sigma_Y^2(a, X)$ (the kriging variance) can both be explicitly estimated.

Obtaining the POD curve consists in replacing $Y = G(a, X)$ by its kriging metamodel (19) in (3). Hence we can estimate the value of $\text{POD}(a)$, for $a > 0$ from:

$$\text{POD}(a) = \mathbb{P} \left((Y(a, X) \mid y^N) > s|a \right). \tag{20}$$

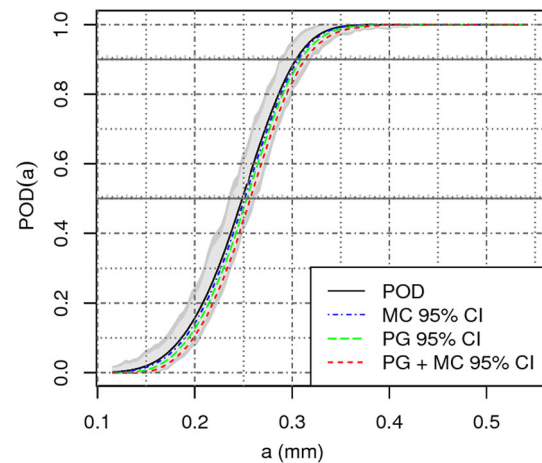


Fig. 11 Example of POD curves estimated with a kriging model

Two sources of uncertainty have to be taken into account in (20): the first coming from the parameter X and the second coming from the Gaussian distribution in (19). From (20), the following estimate for $\text{POD}(a)$ can be deduced:

$$\text{POD}(a) = \mathbb{E}_X \left[1 - \Phi \left(\frac{s - \widehat{Y}_P(a, X)}{\sigma_Y(a, X)} \right) \right]. \tag{21}$$

This expectation is estimated using a classical Monte Carlo integration procedure.

By using the uncertainty implied by the Gaussian distribution regressions, one can build new confidence intervals as explained in [16]. It is illustrated in Fig. 11. We visualize the confidence interval induced by the Monte Carlo (MC) estimation, the one induced by the kriging (PG) approximation and the total confidence interval (including both approximations: PG+MC). For the 90%-level defect, we obtain $a_{90} \simeq 0.305$ mm and $a_{90/95} \simeq 0.315$ mm.

The four methods discussed in this section have given somewhat similar results. This will be discussed in the conclusion of this paper, which also introduces a general and methodological point of view for the numerical POD determination.

4 Sensitivity Analysis on POD Curve

Sensitivity analysis allows to determine those parameters that mostly influence on model response. In particular, global sensitivity analysis methods (see [17] for a recent review) take into account the overall uncertainty ranges of the model input parameters. In this section, we propose new global sensitivity indices attached to the whole POD curve. We focus on the variance-based sensitivity indices, also called Sobol' indices, which are the most popular tools and were proved robust, interpretable and efficient.

4.1 Sobol’ Indices on Scalar Model Output

If all its inputs are independent and $\mathbb{E}(Y^2) < \infty$, the variance of the numerical model $Y = G(X_1, \dots, X_d)$ can be decomposed in the following sum:

$$\text{Var}(Y) = V = \sum_{i=1}^d V_i + \sum_{i < j} V_{ij} + \dots + V_{1\dots d}, \quad (22)$$

with $V_i = \text{Var}[\mathbb{E}(Y|X_i)]$, $V_{ij} = \text{Var}[\mathbb{E}(Y|X_i X_j)] - V_i - V_j$, etc. Then, $\forall i, j = 1 \dots d, i < j$, the Sobol’ indices of X_i write [27]:

$$S_i = \frac{V_i}{V}, S_{ij} = \frac{V_{ij}}{V}, \dots, \text{ and } T_i = S_i + S_{ij} + \dots \quad (23)$$

The first-order Sobol’ index S_i measures the individual effect of the input X_i on the variance of the output Y , while the total Sobol’ index T_i measures the X_i effect and all the interaction effects between X_i and the other inputs (as the second-order effect S_{ij}). T_i can be rewritten as $T_i = 1 - \frac{V_{-i}}{V}$ with $V_{-i} = \text{Var}[\mathbb{E}(Y|X_{-i})]$ and X_{-i} the vector of all inputs except X_i .

These indices are interpreted in terms of percentage of influence of the different inputs on the model output uncertainty (measured by its variance). They have been proven to be useful in many engineering studies involving numerical simulation models [10].

4.2 Sobol’ Indices on POD

In order to define similar sensitivity indices for the whole POD curve, we first define the following quantities:

$$\begin{aligned} \text{POD}_X(a) &= \mathbb{P}(Y > s | a, X), \\ \text{POD}_{X_i}(a) &= \mathbb{P}(Y > s | a, X_i), \\ \text{POD}_{X_{-i}}(a) &= \mathbb{P}(Y > s | a, X_{-i}), \\ D &= \mathbb{E}\|\text{POD}(a) - \text{POD}_X(a)\|^2 \end{aligned} \quad (24)$$

with $\|\cdot\|$ the euclidean norm. The POD Sobol’ indices are then defined by:

$$\begin{aligned} S_i^{\text{POD}} &= \frac{\mathbb{E}\|\text{POD}(a) - \text{POD}_{X_i}(a)\|^2}{D}, \\ T_i^{\text{POD}} &= \frac{\mathbb{E}\|\text{POD}_X(a) - \text{POD}_{X_{-i}}(a)\|^2}{D}. \end{aligned} \quad (25)$$

These POD Sobol’ indices are easily computed with the metamodels. In particular, the kriging metamodel allows one

to replace $\mathbb{P}(Y > s | a)$ by the expectation $\mathbb{E}_X \left[1 - \Phi \left(\frac{s - \widehat{Y}_P(a, X)}{\sigma_Y^2(a, X)} \right) \right]$ in the POD expressions of (24).

Figure 12 gives the sensitivity analysis results on our data. We find that the POD curve is mainly influenced by $ebav_1$ parameter, with smaller effects of $ebav_2$ and h_{12} parameters. As the first-order and total Sobol’ indices strongly differ, we know that the main contributions come from interactions between these three influent parameters. From an engineering point of view, working on the uncertainty reduction of $ebav_1$ is a priority in order to reduce the POD uncertainty.

4.3 Sobol’ Indices for a Specific Defect Size or Probability

The POD Sobol’ indices quantify the sensitivity of each input on the overall POD curve. However, we could be interested in the sensitivities on the detection probability at a specific

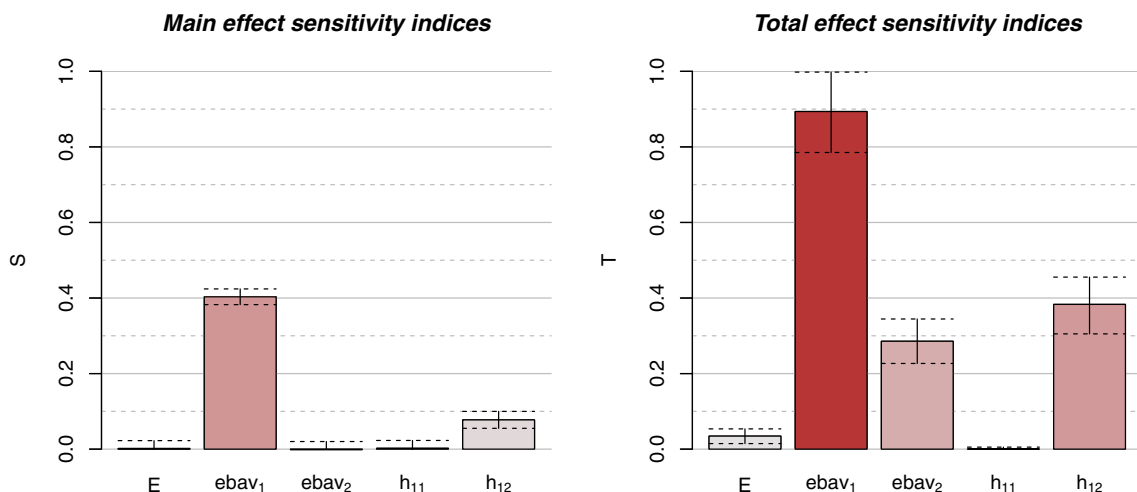


Fig. 12 POD first order (left) and POD total (right) Sobol’ indices

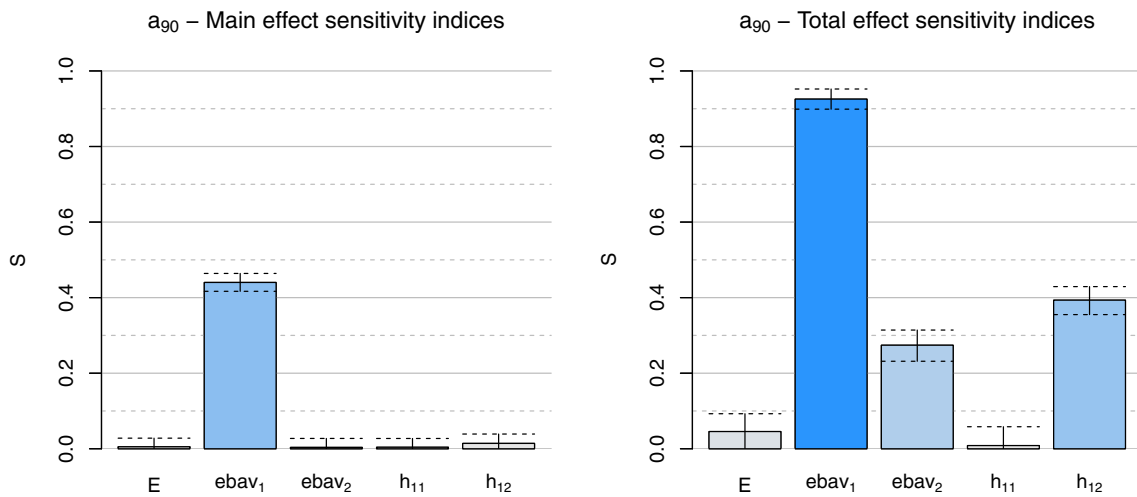


Fig. 13 First order (left) and total (right) Sobol’ indices on a₉₀

Fig. 14 General and progressive MAPOD methodology

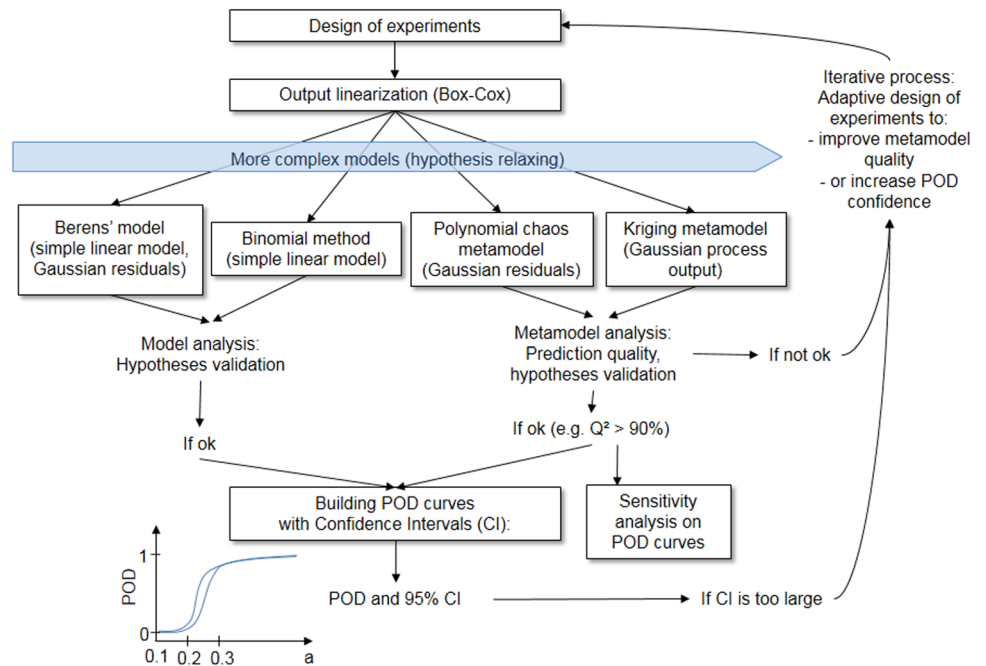


Table 1 Synthesis of results for detectable defect sizes (in mm) with the four methods of the POD methodology

	Berens	Binomial-berens	Polynomial chaos	Kriging
a ₉₀	0.30	0.30	0.30	0.305
a _{90/95}	0.31	0.305	0.32	0.315

defect size *a*. As it is a scalar value, this can be directly done by replacing *Y* by $POD_X(a)$ in all the equations of Sect. 4.1.

If we are now interested by the sensitivities on the defect size at a specific probability detection, we have to study the inverse function of the POD: $POD_X^{-1}(p)$ with *p* a given probability. Similarly to the previous case, the defect size Sobol’ indices can be obtained by replacing *Y* by $POD_X^{-1}(p)$ in all the equations of Sect. 4.1. Figure 13 displays these sensitivity indices on our data for *p* = 0.90. We conclude that *a*₉₀ is

mainly influenced by *ebav*₁ parameter, with smaller effects of *ebav*₂ and *h*₁₂ parameters. The influences are similar than those of the POD curve.

5 Conclusions

This paper has presented four different techniques for POD curves determination (flaw detection probability), valuable

over a wide range of NDT procedures. As part of this study, we focus on the examination under wear anti-vibration bars of steam generator tubes with simulations performed by the finite-element computer code C3D. The model parameterization and the design of numerical experiments have been firstly explained.

Based on these methods of POD curves (and associated confidence intervals) determination, a general methodology is proposed in Fig. 14. It consists in a progressive application of the following methods:

1. the Berens method, based on a linear regression model, and requiring normality assumption on regression residuals;
2. the Binomial-Berens method which relaxes the normality hypothesis;
3. the polynomial chaos metamodel which does not require the linearity assumption but requires normal metamodel residuals;
4. the kriging metamodel.

Other techniques, not discussed here, could be introduced in this scheme, as the quantile regression used in [13] to relax Berens' hypothesis on the residuals distribution, or bootstrap-based alternatives.

The results of these four techniques in terms of the estimation of the defect size detectable with a 90%-probability (a_{90}) and its 95%-lower bound ($a_{90/95}$) are synthesized in Table 1. While a_{90} is rather unchanged, we observe slight variations on $a_{90/95}$ between the different methods.

From the metamodel-based techniques, variance-based sensitivity analysis can also be performed in order to quantify the effect of each input on the POD curve. Other sensitivity analysis methods devoted to POD curves allow to quantify the effects of the modifications of each input distribution. For example, the Perturbation-Law based sensitivity Indices [18] have been used in [16]. Finally, an iterative process can be applied to choose new simulation points in order to improve the metamodels predictivity or to reduce the POD confidence interval (see Fig. 14). These metamodel-based sequential procedures have not been discussed in the present paper.

It is important to note that the obtained POD curves are based on a probabilistic modeling of system input parameters that has to be validated. Moreover, the initial simple model (1) does not fully represent the reality, and taking into account the numerical model uncertainty is an important task [1]. Additional noise as reproducibility noise and measurement errors have also to be added. Solutions for this problem, based on random POD models, are currently under study [7]. Another idea would be to use the Gaussian process framework to infer an error model between the model outputs and some measurements, as done in many validation

model problems [3]. Propagation of this model error, associated to parameter uncertainties, may allow to obtain some more realistic prediction intervals for the POD curve.

Acknowledgements Part of this work has been backed by French National Research Agency (ANR) through project ByPASS ANR-13-MONU-0011. All the calculations were performed by using the OpenTURNS software [2]. We are grateful to Léa Maurice for initial works on this subject, as Pierre-Emile Lhuillier, Pierre Thomas, François Billy, Pierre Calmon, Vincent Feuillard and Nabil Rachdi for helpful discussions. Thanks to Dominique Thai-Van who provided a first version of Fig. 14.

References

1. Aldrin, J., Knopp, J., Sabbagh, H.: Bayesian methods in probability of detection estimation and model-assisted probability of detection evaluation. In: AIP Conference Proceedings vol. 1511, pp. 1733–1744 (2013)
2. Baudin, M., Dutfoy, A., Iooss, B., Popelin, A.: Open TURNS: an industrial software for uncertainty quantification in simulation. In: Ghanem, R., Higdon, D., Owhadi, H. (eds.) Handbook of Uncertainty Quantification. Springer, Berlin (2017)
3. Bayarri, M., Berger, J., Paulo, R., Sacks, J., Cateo, J., Cavendish, J., Lin, C.H., Tu, J.: A framework for validation of computer models. *Technometrics* **49**, 138–154 (2007)
4. Berens, A.: NDE reliability data analysis. In: Higdon, D., Owhadi, H. (eds.) Metals Handbook, pp. 689–701. ASM International, Materials Park (1988)
5. Blatman, G., Sudret, B.: Adaptive sparse polynomial chaos expansion based on Least Angle Regression. *J. Comp. Phys.* **230**(6), 2345–2367 (2011)
6. Box, G., Cox, D.: An analysis of transformations. *J. R. Stat. Soc.* **26**, 211–252 (1964)
7. Browne, T., Fort, J.C.: Redéfinition de la POD comme fonction de répartition aléatoire. Actes des 47èmes Journées de Statistiques de la SFdS, Lille, France (2015)
8. Calmon, P.: Trends and stakes of NDT simulation. *J. Non Destruct. Eval.* **31**, 339–341 (2012)
9. Christensen, R.: Linear Models for Multivariate, Time Series and Spatial Data. Springer, New York (1990)
10. de Rocquigny, E., Devictor, N., Tarantola, S.: (eds) Uncertainty in industrial practice. Wiley, Chichester (2008)
11. Demeyer, S., Jenson, F., Dominguez, N.: Modélisation d'un code numérique par un processus gaussien - Application au calcul d'une courbe de probabilité de dépasser un seuil. Actes des 44èmes Journées de Statistiques de la SFdS, Bruxelles, Belgique (2012)
12. DoD.: Department of Defense Handbook - Nondestructive evaluation system reliability assessment. Technical Report MIL-HDBK-1823A, US Department of Defense (DoD), Washington, DC (2009)
13. Dominguez, N., Jenson, F., Feuillard, V., Willaume, P.: Simulation assisted POD of a phased array ultrasonic inspection in manufacturing. In: AIP Conference Proceedings, vol. 1430, pp. 1765–1772 (2012)
14. Fang, K.T., Li, R., Sudjianto, A.: Design and Modeling for Computer Experiments. Chapman & Hall/CRC, London (2006)
15. Gandosi, L., Annis, C.: Probability of detection curves: statistical best-practice. Technical Report 41, European Network for Inspection and Qualification (ENIQ), Luxembourg (2010)
16. Iooss, B., Le Gratiet, L.: Uncertainty and sensitivity analysis of functional risk curves based on Gaussian processes. Preprint paper http://www.gdr-mascotnum.fr/media/uai2016_iooss (2016)

17. Iooss, B., Lemaître, P.: A review on global sensitivity analysis methods. In: Dellino, G., Meloni, C. (eds.) *Uncertainty management in Simulation-Optimization of Complex Systems, Algorithms and Applications*. Springer, New York (2015)
18. Lemaître, P., Sergienko, E., Arnaud, A., Bousquet, N., Gamboa, F., Iooss, B.: Density modification based reliability sensitivity analysis. *J. Stat. Comput. Simul.* **85**, 1200–1223 (2015)
19. Marrel, A., Iooss, B., Van Dorpe, F., Volkova, E.: An efficient methodology for modeling complex computer codes with Gaussian processes. *Comput. Stat. Data Anal.* **52**, 4731–4744 (2008)
20. Maurice, L., Costan, V., Guillot, E., Thomas, P.: Eddy current NDE performance demonstrations using simulation tools. In: *AIP Conference Proceedings*, vol. 1511, pp. 464–471 (2013a)
21. Maurice, L., Costan, V., Thomas, P.: Axial probe eddy current inspection of steam generator tubes near anti-vibration bars: performance evaluation using finite element modeling. In: *Proceedings of JRC-NDE, Cannes*, pp. 638–644 (2013b)
22. Meyer, R., Crawford, S., Lareau, J., Anderson, M.: Review of literature for Model Assisted Probability of Detection. Technical Report PNNL-23714, Pacific Northwest National Laboratory (2014)
23. Oberkampf, W., Roy, C.: *Verification and Validation in Scientific Computing*. Cambridge University Press, Cambridge (2010)
24. Rosell, A., Persson, G.: Model based capability assessment of an automated eddy current inspection procedure on flat surfaces. *Res. Nondestruct. Eval.* **24**, 154–176 (2013)
25. Rupin, F., Blatman, G., Lacaze, S., Fouquet, T., Chassignole, B.: Probabilistic approaches to compute uncertainty intervals and sensitivity factors of ultrasonic simulations of a weld inspection. *Ultrasonics* **54**, 1037–1046 (2014)
26. Sacks, J., Welch, W., Mitchell, T., Wynn, H.: Design and analysis of computer experiments. *Stat. Sci.* **4**, 409–435 (1989)
27. Sobol, I.: Sensitivity estimates for non linear mathematical models. *Math. Model. Comput. Exp.* **1**, 407–414 (1993)
28. Soize, C., Ghanem, R.: Physical systems with random uncertainties: chaos representations with arbitrary probability measure. *SIAM J. Sci. Comput.* **26**(2), 395–410 (2004)
29. Thomas, P., Goursaud, B., Maurice, L., Cordeiro, S.: Eddy-current non destructive testing with the finite element tool Code_Carmel3D. In: *11th International Conference on Non destructive Evaluation, Jeju* (2015)
30. Thompson, R.: A unified approach to the Model-Assisted determination of Probability of Detection. In: *34th Annual Review of Progress in Quantitative Nondestructive evaluation, July 2007*, pp. 1685–1692. Golden, American Institute of Physics, Melville, New-York (2008)
31. Walter, E., Pronzato, L.: *Identification of Parametric Models from Experimental Data*. Springer, New York (1997)