Metamodel-based nested sampling for model selection in eddy-current testing

Caifang Cai, Sandor Bilicz, Thomas Rodet, Marc Lambert, Dominique Lesselier

Metamodel-based nested sampling for model selection in eddy-current testing

Caifang Cai*, Sándor Bilicz†, Thomas Rodet‡, Marc Lambert§, and Dominique Lesselier*
*Laboratoire des Signaux et Systèmes, CentraleSupélec–CNRS–Univ. Paris-Sud, Univ. Paris-Saclay,
3 rue Joliot-Curie, 91192 Gif-sur-Yvette CEDEX, France †Dept. of Broadband Infocommunications and
Electromagnetic Theory, Budapest Univ. of Technology and Economics,
Egry J. u. 18, 1111 Budapest, Hungary‡SATIE, ENS-Cachan, Univ. Paris-Saclay, 61 avenue du Président Wilson,
94230 Cachan CEDEX, France §GeePs | Group of electrical engineering - Paris, UMR CNRS 8507,
CentraleSupélec, Univ. Paris-Sud, Univ. Paris-Saclay,
Sorbonne Univ, UPMC Univ. Paris 06,
3 & 11 rue Joliot-Curie, Plateau de Moulon 91192 Gif-sur-Yvette CEDEX, France

Abstract

In Non-Destructive Testing, model selection is a common problem, e.g, to determine the number of defects present in the inspected workpiece. Statistical model selection requires to approximate the marginal likelihood also called model evidence. Its numerical approximation is usually computationally expensive. Nested Sampling (NS) offers a good compromise between estimation accuracy and computational cost. But it requires to evaluate the forward model many times. Here, we first propose a general framework where data-fitting surrogate models are used to accelerate the computation. Then, improvements benefiting from surrogate modeling are introduced into the traditional NS algorithm to further reduce the computational cost. These improvements include the use of a sparse grid surrogate model to deal with the "curse-of-dimensionality" in large dimensional problems and of the pre-estimated posterior space to save warming-up time. Based on eddy-current simulations, we show that this improved model selection approach has high model selection ability and can jointly perform model selection and parameter inversion.

Index Terms

Nested Sampling, model selection, marginal likelihood, non-destructive testing, eddy-current, surrogate model.

I. Introduction

In Non-Destructive Testing (NDT) including Eddy Current Testing (ECT), model selection and parameter inversion are common inverse problems in quantitative analysis. Parameter inversion is to estimate parameters of interest from measurements. These parameters can be characteristics of the flaw [1], [2], properties of the inspected material [3], [4] or configurations inspection set-ups [5], like lift-off, probe frequency, etc.

Yet, the unknown parameters are always subject to a given model. E.g., if we are interested in the size of the flaw, we often assume that we already know the properties of the inspected material, the shape of the flaw and the inspection set-ups.

In contrast, model selection is about taking a decision between two (or several) competing models. Each model has unknown parameters as well. Model selection is often used in predicting the source of measurements in NDT, e.g, in ECT flaw characterization, to choose between an air bubble and a surface crack within the inspected piece. In general, model selection is more difficult than parameter inversion.

To tackle the model selection, we need to choose a proper criterion from which decisions are made. Akaike Information Criterion (AIC) [6], Bayesian Information Criterion (BIC) [7] and Bayes factor [8] are three commonly used criteria. Bayes factor is defined by the ratio of posterior integral over all unknown parameters between two competing models. AIC and BIC are approximations of the Bayes factor when the likelihood has an exponential form and plays a more significant role than the prior in the posterior distribution, refer to [9] for more detail.

In other words, using AIC and BIC for model selection, decisions are based on the maximum likelihoods and the number of unknown parameters. Bayes factor uses the total information from the posterior distribution. Especially when we have important prior information, model selection based on Bayes factor is much more reliable, and we choose to use it as our criterion in this work.

The next step, which is also the most difficult part, is to approximate the corresponding evidences. There exist several types of algorithms to this effect: Reversible-jump Markov Chain Monte Carlo (RJ-MCMC) [10], Importance Sampling (IS) [11], path sampling [12] and Nested Sampling (NS) [13]–[15]. All follow the same procedure: drawing a series of samples following the posterior distribution for the considered model, making decision based on evidences approximated from these samples.

They differ from one another in how the samples are drawn. Depending on the complexity of posterior distribution and the number of unknown parameters, they can be chosen accordingly. RJ-MCMC is efficient for single-modal and isotropic

Corresponding author: Caifang Cai (email: caifang.cai@l2s.centralesupelec.fr).

distributions with small number of unknown parameters while IS can be very efficient for specific types of distributions [16]. In contrast, NS can manage complicated forward models and it can jointly solve the parameter inversion and the model selection problems. The only constraint is that there should be a limited number of unknowns in competing models. For this reason, we suggest to use NS algorithm for ECT applications, because, in ECT, the direct model is often non-linear, the number of unknown parameters is usually limited, parameter inversion and model selection are both within the scope of investigation.

NS has been proposed by Skilling in [13] and improved in [14], [15], [17], [18]. The main idea of NS is to generate samples with increasing likelihoods by exponentially narrowing down the sampling space. The core of a NS algorithm is to get a random sample in the remaining prior volume with constraint on likelihood. In the original work of Skilling [13], a Markov Chain Monte Carlo (MCMC) algorithm is advocated, yet it suffers from low acceptance rate. In [14], the authors proposed an ellipsoidal approximation method to locate the remaining sampling area, but the acceptance rate is only improved for single-modal distributions. In [15], the authors further introduced a multi-ellipsoidal approximation method, the acceptance rate of which is much better for multimodal distributions.

None of these improvements was made for narrow posterior distributions. In practice, the posterior distribution is concentrated only on a small part of the prior volume. Since the sampling space is narrowed exponentially from the prior space in a NS algorithm, many iterations are wasted on warming-up. Our first contribution in this work is to propose an efficient NS algorithm where the initial sampling space is narrowed via a pre-estimated posterior volume, the pre-estimation costs being negligible due to use of a global surrogate model. Secondly, we improve the *MultiNest* algorithm proposed in [15] by using a Minimum-Volume Enclosing Ellipsoid (MVEE) method to better approximate the remaining sampling volume.

To reduce the computational cost, surrogate forward models [19]–[24] have been used in Bayesian inversion approaches. A surrogate model is made of a pre-trained database and a fast interpolator. The database contains a large number of input-output pairs simulated by an accurate forward model. In the inversion, only the interpolator is called. Our last contribution in this work is to use a sparse grid surrogate model [25] where the input nodes in the database are on fixed sparse grids, which reduces the computational burden for large dimensional problems and allows, in natural manner, a parallel implementation.

The paper is organized as follows. In § II, we introduce the Bayesian framework for joint model selection and parameter inversion based on a sparse grid surrogate model. We discuss in detail the improved algorithm called *N-MultiNest* in § III to approximate the evidences as required in model selection. Simulations follow in § IV based on a ECT example.

II. Framework of Bayesian model selection

A. Bayesian model selection

For a model $\mathcal{M}$, the relation between observations $\boldsymbol{y}$ and unknown parameters $\boldsymbol{x}$ can read as

$$\boldsymbol{y} = f\left(\boldsymbol{x}, \mathcal{M}\right) + \boldsymbol{\varepsilon}, \tag{1}$$

where $\boldsymbol{\varepsilon}$ denotes the additive noise and $\boldsymbol{x}$ is the joint unknown parameters subject to model $\mathcal{M}$. The most important element in Eq. (1) is the forward model $f(\boldsymbol{x}, \mathcal{M})$ which describes the physical relation between unknown parameters $\boldsymbol{x}$ and observations $\boldsymbol{y}$. The complexity of parameter inversion or model selection depends on the complexity of this function, and this function describes the complexity of the problem.

In model selection, we usually have a single set of measurements $\boldsymbol{y}$ yet several potential models. Model selection is to determine from which model these measurements are derived. For simplicity, we focus on selection between two models, the hurdles being the same for multiple model selection, and strategies proposed in § III being directly applicable.

Let us assume that we have two competing models $\mathcal{M}_1$ and $\mathcal{M}_2$. The simplest way to make a model selection choice is to check the residual between simulated $f(\boldsymbol{x}, \mathcal{M})$ and measured $\boldsymbol{y}$ for each model. However, this will only work for models with the same number of unknowns. Otherwise, the model with a larger number of unknowns has more freedom in dimension. It is more probable to get a simulated $f(\boldsymbol{x}, \mathcal{M})$ closer to the measurement $\boldsymbol{y}$. This means that we will always be biased in favor of the model which has more unknown parameters. This is also the reason why Bayesian model selection often works on the Bayes factor defined by

$$r(\mathcal{M}_1, \mathcal{M}_2) = \frac{p(\mathcal{M}_1|\boldsymbol{y})}{p(\mathcal{M}_2|\boldsymbol{y})} = \frac{p(\boldsymbol{y}|\mathcal{M}_1)\, p(\mathcal{M}_1)}{p(\boldsymbol{y}|\mathcal{M}_2)\, p(\mathcal{M}_2)}. \tag{2}$$

If $r(\mathcal{M}_1, \mathcal{M}_2)$ is larger than 1, we can conclude that the observations $\boldsymbol{y}$ are from the $\mathcal{M}_1$, otherwise from $\mathcal{M}_2$. In Eq. (2), $\{p(\boldsymbol{y}|\mathcal{M}_i), i = 1, 2\}$ are the model evidences while $\{p(\mathcal{M}_i), i = 1, 2\}$ are the model prior probabilities. The latter are chosen between 0 and 1 based on the prior information and should verify $p(\mathcal{M}_2) + p(\mathcal{M}_1) = 1$. Whithout any prior information $p(\mathcal{M}_1) = p(\mathcal{M}_2) = 0.5$ is a common choice and will be ours in the following. According to Bayes law, the Bayes factor in Eq. (2) can be expressed as

$$\begin{aligned} r(\mathcal{M}_1, \mathcal{M}_2) &= \frac{p(\boldsymbol{y}|\mathcal{M}_1)}{p(\boldsymbol{y}|\mathcal{M}_2)} \\ &= \frac{\int p(\boldsymbol{y}|\boldsymbol{x}_1, \mathcal{M}_1)\, p(\boldsymbol{x}_1|\mathcal{M}_1)\, \mathrm{d}\boldsymbol{x}_1}{\int p(\boldsymbol{y}|\boldsymbol{x}_2, \mathcal{M}_2)\, p(\boldsymbol{x}_2|\mathcal{M}_2)\, \mathrm{d}\boldsymbol{x}_2}, \end{aligned} \tag{3}$$

where $\{p(\boldsymbol{y}|\boldsymbol{x}_i, \mathcal{M}_i), i = 1, 2\}$ are the likelihoods while $\{p(\boldsymbol{x}_i|\mathcal{M}_i), i = 1, 2\}$ are the parameter priors.

The importance of using the integral ratio in Eq. (3) to replace the evidence ratio is to render the calculation of this ratio practically possible. This is because the model evidence $p(\boldsymbol{y}|\mathcal{M})$ is not tractable, but likelihood $p(\boldsymbol{y}|\boldsymbol{x}, \mathcal{M})$ and parameter prior $p(\boldsymbol{x}|\mathcal{M})$ are. In most applications, the additive noise $\boldsymbol{\varepsilon}$ can be approximated by an independent identically distributed (i.i.d.) Gaussian model

$$\boldsymbol{\varepsilon} \sim \mathcal{N}(0, \sigma_y^2), \tag{4}$$

σ_y^2 being the noise variance, usually known or pre-estimated easily. Based on this noise model, recalling the forward model given in Eq. (1), it is not difficult to find out that the likelihood is fully computable given the model of concern $\mathcal{M}$ and the corresponding parameter $\boldsymbol{x}$. It has the following analytical expression for a model with M measurements:

$$p(\boldsymbol{y}|\boldsymbol{x}, \mathcal{M}) = (2\pi\sigma_y^2)^{-\frac{M}{2}} \exp\left\{-\frac{\|\boldsymbol{y} - f(\boldsymbol{x}, \mathcal{M})\|^2}{2\sigma_y^2}\right\}. \tag{5}$$

Regarding to the prior, the following uniform model can be used if no further information is available

$$p(\boldsymbol{x}|\mathcal{M}) = \frac{\chi_{\mathcal{U}}(\boldsymbol{x})}{\text{volume}(\mathcal{U})},\ \mathcal{U} = [a_1, b_1] \times \cdots \times [a_N, b_N] \tag{6}$$

where $\mathcal{U}$ is the uniform parameter space with N denoting the number of unknown parameters. $\chi_{\mathcal{U}}(\boldsymbol{x})$ is the indicator function on $\mathcal{U}$. All values of our interest for the i-th parameter x_i in the vector $\boldsymbol{x}$ should be included in the corresponding range between a_i and b_i.

B. Sparse-grid surrogate model

In the Bayesian framework, model selection decisions are made based on the Bayes ratio defined in Eq. (3). It can be calculated by integrating the product of two tractable elements: likelihood and prior. Yet, if we look at the Gaussian likelihood in Eq. (5), the forward model $f(\boldsymbol{x}, \mathcal{M})$ is present, so evaluation of the likelihood requires to tackle it. Considering that model evidences are usually approximated by Monte Carlo simulation methods, thousands of likelihood evaluations might be required. If the evaluation of $f(\boldsymbol{x}, \mathcal{M})$ is costly, it will be very hard to approximate the model evidence.

In ECT, there exist many modeling methods for $f(\boldsymbol{x}, \mathcal{M})$, such as the Method of Moments (MoM), the Finite Element Method, etc. In inversion problems, be they parameter inversion or model selection, proper compromise is often needed between modeling accuracy and computational cost. For a model selection, the computational cost is particularly crucial since it requires at least one order-of-magnitude more of likelihood evaluations compared to parameter inversion. Data-fitting surrogate models have been proposed [21], [23], [24] to reduce the computational cost in the inversion. Here, we use the same idea to overcome the computational cost problem.

A data-fitting surrogate model usually includes two essential parts: database and interpolator. The database is made of many pairs of $\{\boldsymbol{x}^j, f(\boldsymbol{x}^j, \mathcal{M})\}$ where $f(\boldsymbol{x}^j, \mathcal{M})$ is an accurate but relatively expensive model. In our case, a MoM is used [26]. The database is trained once at the beginning for a given model and re-used for all other inverse problems.

In the later model selection, only interpolations are performed. Since the goal is to reduce the computational cost without losing too much modeling accuracy, interpolation accuracy is considered in database training, i.e., it is based on the accuracy requirement of the global approximation of the interpolation.

Let us use $\mathcal{D}$ to denote the database, then

$$\mathcal{D} = \{\boldsymbol{x}^j, f(\boldsymbol{x}^j, \mathcal{M})\}, \quad j = 1, 2, \cdots, J. \tag{7}$$

J is the dimension of the total database, and the number of required forward evaluations as well.

In our situation, we use a surrogate model based on the sparse grid interpolation. This technique is presented in detail in [27], and it has recently been applied in the context of ECT [25]. In the present work, the sparse grid database is trained on the prior space $\mathcal{U}$.

C. Model selection framework with surrogate forward model

Being different from the forward model used in the database training, we denote the data-fitting based surrogate model as $\tilde{f}(\boldsymbol{x}, \mathcal{M}, \mathcal{D})$. In the following likelihood evaluations, it is used to replace $f(\boldsymbol{x}, \mathcal{M})$ in Eq. (5). Then the Bayesian model selection can be sketched as shown in Fig. 1 for a dual-model case.

In terms of computational cost, the two most expensive parts are database training and model evidence estimation. Compared to the on-line interpolation called in model evidence estimation, the database can be trained off-line. Interpolation costs much less time than accurate forward model evaluation. Consequently, on-line computational time can be reduced a lot.

The main interest of using surrogate model here is that we can treat the database training separately from the model evidence estimation. Acceleration strategies can then be employed independently for these two expensive operations.

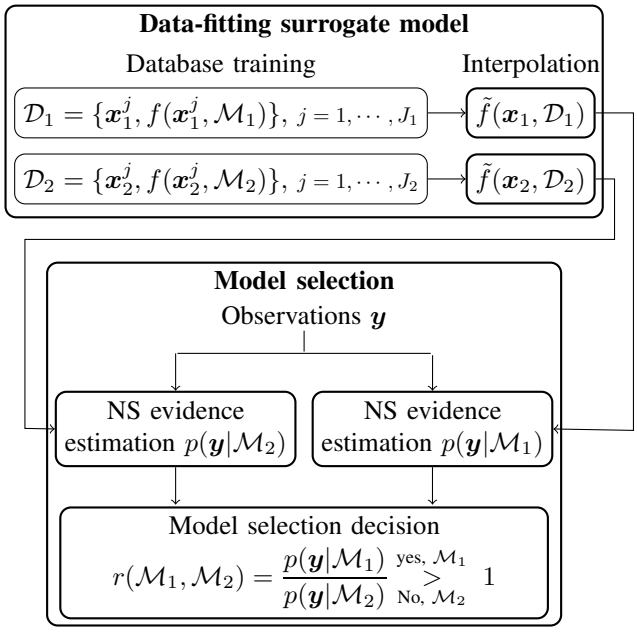


Fig. 1. Sketch of Bayesian model selection based on data-fitting surrogate model for dual-model selection.

Algorithm 1 A general algorithm of nested sampling.

procedure NESTED SAMPLING
  Set $k = 1$, $\hat{Z}^0 = 0$ and $u^0 = 1$
  Initialize $N_a{}^*$ active samples $\{\boldsymbol{x}^{m_1}, \boldsymbol{x}^{m_2}, \cdots, \boldsymbol{x}^{m_{N_a}}\}$ randomly from the prior $p(\boldsymbol{x}|\mathcal{M})$
  Sort $\{\boldsymbol{x}^{m_1}, \boldsymbol{x}^{m_2}, \cdots, \boldsymbol{x}^{m_{N_a}}\}$ in ascending order of likelihoods $p(\boldsymbol{y}|\boldsymbol{x}^{m_1}, \mathcal{M}) < p(\boldsymbol{y}|\boldsymbol{x}^{m_2}, \mathcal{M}) < \cdots < p(\boldsymbol{y}|\boldsymbol{x}^{m_{N_a}}, \mathcal{M})$
  while $\frac{p(\boldsymbol{y}|\boldsymbol{x}^{m_{N_a}}, \mathcal{M})\ u^k}{\hat{Z}^{k-1}} > 10^{-3}$ **do**
    Generate a random sample $\boldsymbol{x}^{(p)}$ from the remaining prior volume so that $p(\boldsymbol{y}|\boldsymbol{x}^{(p)}, \mathcal{M}) > p(\boldsymbol{y}|\boldsymbol{x}^{m_1}, \mathcal{M})$
    Update model evidence $\hat{Z}^k = \hat{Z}^{k-1} + \Delta u^k\, p(\boldsymbol{y}|\boldsymbol{x}^{m_1}, \mathcal{M})$
    Replace $\boldsymbol{x}^{m_1}$ by $\boldsymbol{x}^{(p)}$ in the active sample set and rearrange them in ascending order of likelihoods
    Increase iteration $k++$
  end while
  Make use of the active samples from the last iteration $\hat{Z} = \hat{Z}^k + \sum_{n=1}^{N_a} \Delta u^{k+n}\, p(\boldsymbol{y}|\boldsymbol{x}^{m_n}, \mathcal{M})$
  return
end procedure

$*$ N_a is the the number of active samples, fixed initially based on the number of unknown parameters and the accuracy demand on evidence approximation.

III. EVIDENCE APPROXIMATION FOR NARROW POSTERIOR DISTRIBUTIONS

A. Principle of nested sampling

Seen from the framework sketched in Fig. 1, one of the most important tasks is to get the model evidence $p(\boldsymbol{y}|\mathcal{M})$ for all models of concern. NS originally proposed in [13] is one of the efficient algorithms.

Let us recall the evidence estimation problem

$$Z = p(\boldsymbol{y}|\mathcal{M}) = \int p(\boldsymbol{y}|\boldsymbol{x}, \mathcal{M})\ p(\boldsymbol{x}|\mathcal{M})\ \mathrm{d}\boldsymbol{x}. \tag{8}$$

Since $\boldsymbol{x} \in \mathbb{R}^N$, it is a multi-dimensional integration problem.

The NS algorithm proposed by Skilling [13] first transforms this multi-dimensional integral into an one-dimensional integral, then approximate it numerically by using a series of samples specifically generated for this objective. We rewrite in Algorithm 1 the classical NS algorithm proposed by Skilling. The series of samples used for evidence approximation are the collection of samples $\boldsymbol{x}^1$ excluded from the active sample set at each iteration and all active samples at the last iteration.

For a better understanding of the NS algorithm, we summarize its principle in the following and more details being found in [13].

Let us define the prior volume by

$$u = \varphi(\lambda) = \int_{p(\boldsymbol{y}|\boldsymbol{x}',\mathcal{M})>\lambda} p(\boldsymbol{x}'|\mathcal{M})\ \mathrm{d}\boldsymbol{x}', \tag{9}$$

which is a scalar function on $[0, 1]$. Its relation with the likelihood is illustrated in Fig. 2. $\varphi(\lambda)$ represents the volume of parameter space by setting a threshold λ on the likelihood distribution. It equals to one when λ is set to zero and tends to zero when λ is increased up to the maximal likelihood value.

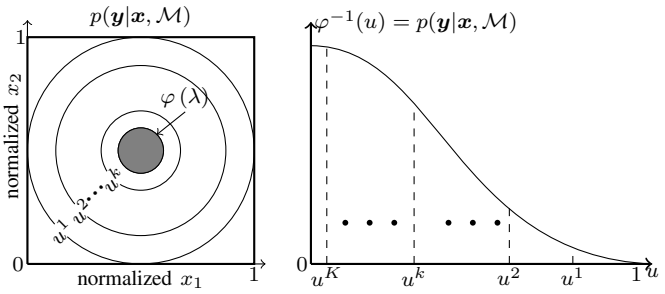


Fig. 2. 2D illustration of the relation between the prior volume and the likelihood function.

By taking Eq. (9) into Eq. (8), the multi-dimensional integral for evidence calculation can be transformed into an one-dimensional integral

$$Z = \int p(\boldsymbol{y}|\boldsymbol{x},\mathcal{M})p(\boldsymbol{x}|\mathcal{M})\ \mathrm{d}\boldsymbol{x} = \int_0^1 \varphi^{-1}(u)\ \mathrm{d}u, \tag{10}$$

where $\varphi^{-1}(u)$ denotes the inverse function of $\varphi(\lambda)$.

If we have a series of samples $\{\boldsymbol{x}^1, \boldsymbol{x}^2, \cdots, \boldsymbol{x}^K\}$ where the corresponding prior volumes are $\{u^1, u^2, \cdots, u^K\}$ with $u^k = \varphi\left(\lambda = p(\boldsymbol{y}|\boldsymbol{x}^k, \mathcal{M})\right)$, then the evidence in Eq. (10) can be approximated numerically by

$$\hat{Z} \approx \sum_{k=1}^{K} \Delta u^k \varphi^{-1}(u^k) = \sum_{k=1}^{K} \Delta u^k p(\boldsymbol{y}|\boldsymbol{x}^k, \mathcal{M}). \tag{11}$$

We see that the sample $\boldsymbol{x}^k$ appears within the evidence approximation.

The way how samples are generated in NS algorithm ensures that the corresponding prior volume is narrowed down exponentially, so we have

$$u^k = \exp\left\{-\frac{k}{N_a}\right\},\ k = 1, 2, \cdots, K. \tag{12}$$

Proofs can be found in [13]. For a trapezoidal approximation,

$$\Delta u^k = \begin{cases} \dfrac{u^{k-1} - u^{k+1}}{2}, & 1 \leq k < K \\ \dfrac{u^{k-1}}{2}, & k = K \end{cases} \tag{13}$$

The most difficult part in a NS algorithm is the step 5 in Algorithm 1, known as "sampling problem with constraint on likelihood". In [13], the use of an independant MCMC algorithm is proposed but it suffers from low acceptation rate and the resulting samples are not totally independent. Many contributions [14], [17] have been made to improve this step, but none of them is dedicated to narrow likelihood distributions. In the following sub-sections, we first introduce the sampling difficulties for narrow likelihood distributions. Then, we address a practical strategy benefiting from the use of the data-fitting surrogate model to overcome them.

B. Difficulties for narrow posterior distributions

The prior space is usually much larger than the posterior space, especially when using a data-fitting surrogate model. The database must be trained on a large space including all possible values for the unknown parameters. This means that the likelihood value is only important on a tiny part of the entire prior space. Unfortunately, a NS algorithm always starts from the prior unit and narrows down exponentially. So it might take many iterations for a NS algorithm to reach down to the posterior space. These iterations, called warm-ups, do not contribute to the evidence approximation but bring extra computational cost.

In order to quantify the narrowness of the posterior distribution, we introduce u_p as the normalized posterior volume

$$u_p = \frac{\text{volume}\big(p(\boldsymbol{y}|\boldsymbol{x},\mathcal{M}) > \epsilon_L\big)}{\text{volume}(\mathcal{U})}. \tag{14}$$

The normalization is with respect to the prior volume. Here, ϵ_L is a small positive value which thresholds "non-zero" likelihoods. $\mathcal{U}$ is the same prior volume as used in Eq. (6). In a classical NS algorithm, a new sample is always drawn in the remaining prior volume. For an exponentially decreasing speed given in Eq. (12), the equivalent number of warm-ups required to get a sample within the posterior volume can be given approximately as follows

$$N_{\text{warm-up}} = \lfloor -N_a \ \ln(u_p) \rfloor , \tag{15}$$

N_a being the number of active samples.

Tab. I shows several examples for $N_{\text{warm-up}}$ at different u_p and N_a. The number of warm-up iterations increases rapidly with the number of active samples N_a and the decrease of u_p.

TABLE I
EXAMPLES OF WARM-UP ITERATIONS IN A NESTED SAMPLING ALGORITHM WITH N_a ACTIVE SAMPLES FOR A POSTERIOR DISTRIBUTION WITH NARROWNESS u_p.

	$N_a = 100$	$N_a = 200$	$N_a = 500$	$N_a = 1000$
$u_p = 0.1$	230	460	1 151	2 302
$u_p = 0.01$	460	921	2 302	4 605
$u_p = 0.001$	690	1 381	3 453	6 907
$u_p = 0.0001$	920	1 842	4 605	9 210

For large dimensional problems, most of the volume will be concentrated near the outer surface of the prior space, and u_p will be very small. Meanwhile, to guarantee approximation accuracy for the model evidence, a large number of active samples must be used. So, we need to deal with small u_p, large N_a. This means that a classical NS algorithm, single-nest or multi-nest, wastes too much on warming-up.

C. N-MultiNest: Multi-nest sampling with narrowed searching space

In [17], the authors proposed an efficient multi-nest sampling algorithm for multimodal posterior distributions. Our work is based on the same sampling algorithm but with narrowed searching space at the initialization. In order to distinguish from the single-nest algorithm in [13], [14], we call the algorithm in [17] *MultiNest*, and ours with narrowed searching space *N-MultiNest*. The narrowing is performed by approximating the posterior space using nodes existing in the metamodel database.

In our case, a data-fitting surrogate model is used instead of an expensive MoM forward model. In this surrogate model, we have already previously trained a database $\mathcal{D}$ for which the corresponding likelihoods can be easily computed according to Eq. (5).

Then, a thresholding method can be applied on these likelihood values to locate all important samples, where "important" means that the corresponding likelihood is sufficienty large to make a significant contribution to the model evidence approximation in Eq. (11). From the important samples, we can approximate the posterior space u_p by the total intersection volume of the enclosing multi-ellipsoids with the prior space. In the 2D case, this procedure of posterior space approximation can be summarized by the scheme in Fig. 3.

The four essential steps marked by 1) - 4) in Fig. 3 are as follows:

1) Likelihood matching based on Eq. (5)

$$\begin{aligned} \{\boldsymbol{x}^j, f(\boldsymbol{x}^j, \mathcal{M})\} \rightarrow & \{\boldsymbol{x}^j, p(\boldsymbol{y}|\boldsymbol{x}^j, \mathcal{M})\}, \\ & \boldsymbol{x}^j \in \mathcal{D}, j = 1, 2, \cdots, J, \end{aligned} \tag{16}$$

where J is the total number of nodes in metamodel database.

2) Thresholding of important samples

$$\begin{aligned} & \mathcal{I} = \{\boldsymbol{x}^{i_1}, \boldsymbol{x}^{i_2}, \cdots, \boldsymbol{x}^{i_{N_s}}\}, \\ & \frac{p(\boldsymbol{y}|\boldsymbol{x}^{i_n}, \mathcal{M})}{\max\{p(\boldsymbol{y}|\boldsymbol{x}^j, \mathcal{M}), j = 1, 2, \cdots, J\}} < \epsilon_L, \\ & i_n = i_1, i_2, \cdots, i_{N_s}. \end{aligned} \tag{17}$$

$\mathcal{I}$ denotes the set of important samples and N_s is the number of important samples.

3) Locating the enclosing ellipsoid(s)

$$\begin{aligned} \mathcal{E} = & \{\mathcal{E}_1, \mathcal{E}_2, \cdots, \mathcal{E}_{N_e}\} \\ & = \arg \min_{(\mathcal{E}_{m_1}, \cdots, \mathcal{E}_{m_{N_e}})} \left\{ \text{volume}(\mathcal{E}_{m_1} \cap \cdots \cap \mathcal{E}_{m_{N_e}}) \right\}, \\ & \text{s.t. any } \boldsymbol{x}^i \in \mathcal{I}, \text{ there exists } \mathcal{E}_m = (\boldsymbol{\mu}_m, \boldsymbol{C}_m) \in \mathcal{E} \\ & \text{that } (\boldsymbol{x}^i - \boldsymbol{\mu}_m) \boldsymbol{C}_m^{-1} (\boldsymbol{x}^i - \boldsymbol{\mu}_m)^T < 1, \end{aligned} \tag{18}$$

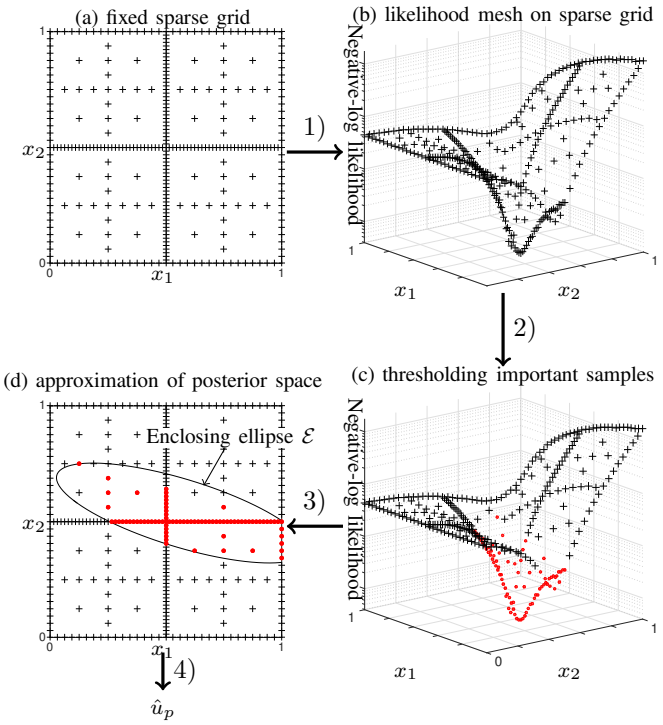


Fig. 3. Approximation of the posterior space by using a multi-ellipsoidal enclosing method on important samples. Where + denotes the sparse grid nodes and • the important samples.

where N_e is the number of ellipsoids, $\boldsymbol{\mu}_m$ and $\boldsymbol{C}_m$ are ellipsoid center and bounding matrix of $\mathcal{E}_m$.

4) Approximation of the posterior volume

$$\hat{u}_p \approx \text{volume}(\mathcal{E} \cap \mathcal{U}), \tag{19}$$

where $\mathcal{U}$ is the uniform prior space, the same as used in Eq. (6).

Applying *MultiNest* on the estimated posterior space results in *N-MultiNest* as described in algorithm 2, where u^k is given

$$u^k = \hat{u}_p \exp\left\{-\frac{k}{N_a}\right\}, \tag{20}$$

l_m is the average likelihood for all non-important samples in the metamodel database. With reference to the scheme shown in Fig. 3, it is the average likelihood value for all black crosses in Fig. 3 (d).

Algorithm 2 Multi-nested sampling algorithm with narrowed searching space at the initialization.

1: **procedure** *N-MultiNest*
2: Initialize with N_a random samples $\{\boldsymbol{x}^{m_1}, \boldsymbol{x}^{m_2}, \cdots, \boldsymbol{x}^{m_{N_a}}\}$ from estimated posterior space $\mathcal{E} \cap \mathcal{U}$ and sort them in ascending order of likelihoods.
3: Set $\hat{Z}^0 = l_m(1 - \hat{u}_p)$ and $u^0 = \hat{u}_p$ with $\hat{u_p}$ given in Eq. (19).
4: Same steps 4-10 as in Algorithm 1 but with u^k given in Eq. (20) instead of Eq. (12).
5: **end procedure**

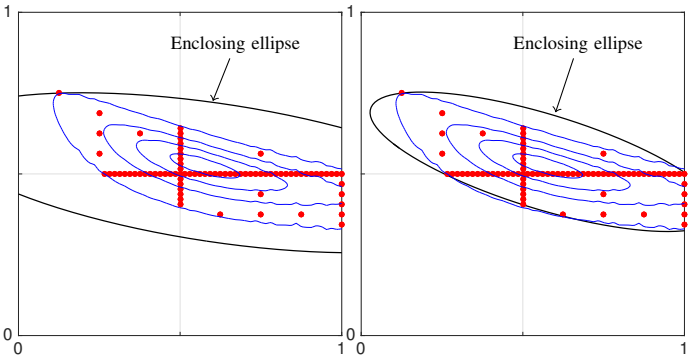
Compared with Algorithm 1, Algorithm 2 runs only on the estimated posterior space instead of the entire prior space. For narrow likelihood distributions, it can save on warming-up iterations. $\hat{Z}^0$ is now the marginal likelihood on the complement of the posterior space, called initial evidence as well. By comparing it with the estimated evidence $\hat{Z}$, it can confirm whether the chosen value for ϵ_L is appropriate or not. A detailed discussion is given in § IV-D.

D. Ellipsoidal enclosing of important samples

In the example shown in Fig. 3 (d), there is only one ellipse enclosing all important samples. This is usually the case for unimodal, isotropic likelihoods. For multimodal or anisotropic likelihoods, multiple ellipses are used. In this case, we need to determine the number of ellipses N_e, their centers and bounding matrices. This is done by clustering the important samples first and then drawing an enclosing ellipse for each cluster.

In [15], [17], a brief review of clustering algorithms is given. A recursive partitioning and bounding algorithm has also been proposed in [17] based on a hierarchic K-means algorithm, wherein the partition method works jointly with the bounding method to minimize the total volume. In our implementation, a similar algorithm is used, with improvements made on how to get the ellipse bounding matrix after recursive K-means clustering.

Let us assume that we have a sub-set of important samples $\mathcal{I}_m \subset \mathcal{I}$, $\mathcal{I}_m = \{\boldsymbol{x}^{n_1}, \boldsymbol{x}^{n_2}, \cdots, \boldsymbol{x}^{n_{N_m}}\}$. All its elements are clustered in the same group and enclosed by an ellipsoid $\mathcal{E}_m = (\boldsymbol{\mu}_m, \boldsymbol{C}_m)$.



(a) Feroz method [17] (b) Khachiyan method [28], [29]

Fig. 4. 2D example of enclosing ellipse for sparse grid important samples for "banana" shaped likelihood distribution where • represents the important samples and – the likelihood density lines.

In [17], the authors proposed to use the mass center for $\boldsymbol{\mu}_m$ and the covariance matrix for $\boldsymbol{C}_m$. After applying this ellipsoidal enclosing method to our problem, we observed that the result is relatively poor. We give in Fig. 4a the result for a "banana" shaped likelihood distribution. We see that the area enclosed by the ellipse is far away from the true distributed area of the important samples. Two reasons can explain why this so-called Feroz method does not work for our problem. First, the likelihood that we are dealing with is a very narrow distribution, and there is a limited number of important samples. Second, due to the use of a sparse grid surrogate model, the important samples are not uniformly distributed according to the true likelihood. But the Feroz method therein is efficient only when sufficient samples are available and the samples should be quasi-uniformly distributed in the remaining posterior space. For our problem here, these two conditions are clearly not satisfied.

So, we propose to use the following way to find the MVEE:

$$\begin{aligned} (\boldsymbol{\mu}_m, \boldsymbol{C}_m) &= \arg\min \{\det(\boldsymbol{C}_m)\} \\ s.t. \quad & (\boldsymbol{x}^i - \boldsymbol{\mu}_m)\boldsymbol{C}_m^{-1}(\boldsymbol{x}^i - \boldsymbol{\mu}_m)^T \leq 1, i = n_1, n_2, \cdots, n_{N_m} \\ & \boldsymbol{C}_m > 0, \end{aligned} \tag{21}$$

where $\boldsymbol{C}_m > 0$ means $\boldsymbol{C}_m$ is positive definite.

Since the ellipsoidal volume is proportional to $\det(\boldsymbol{C}_m)$, the optimization given in Eq. (21) is a joint optimization problem minimizing the volume of the enclosing ellipsoid. It can be solved by the Khachiyan method [28]. More efficient implementations are discussed in [29]–[31]. In our implementation, the dual-optimization algorithm described in [29] is used. Fig. 4b shows the result for the same ellipsoidal enclosing problem as in Fig. 4a. The ellipse is closer to the true posterior volume because the Khachiyan method does not require a lot of samples and the samples do not need to be uniformly distributed within the posterior volume.

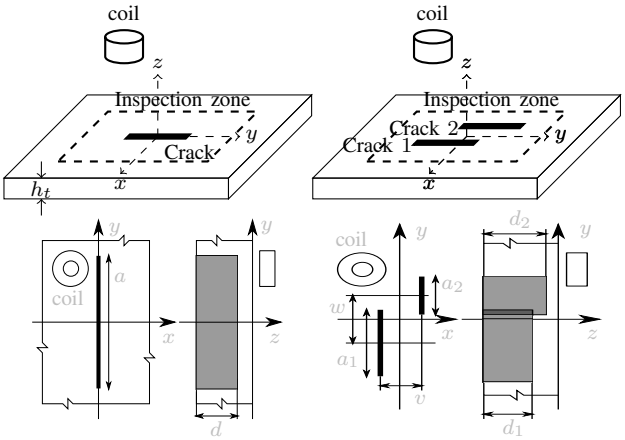
Considering the improved efficiency of this MVEE method compared to the Feroz method [17], we apply it not only here for posterior volume approximation but also in optimal ellipsoidal decomposition in evidence estimation in Algorithm 2.

IV. Simulation Test in Eddy-Current Testing

A. Model selection between single crack and double cracks

To analyze the performance of our approach, simulations are performed based on a ECT example sketched in Fig. 5. A planar infinite non-ferromagnetic plate is affected by surface cracks. The number of cracks is unknown. To simplify the discussion, we only consider two possible situations: one with a single crack, the other with two cracks. In terms of observations, a surface scan of impedance variations is performed via an air-cored probe coil driven by a time-harmonic current signal. More information about the configurations can be found in Tab. II.

Let us name the single-crack model as $\mathcal{M}_1$ and the double-crack model as $\mathcal{M}_2$. In $\mathcal{M}_1$, only two parameters are used to characterize the flaw while six parameters are required for $\mathcal{M}_2$. We can imagine that the measurements from $\mathcal{M}_1$ and $\mathcal{M}_2$ will be very close to each other when v is very small. To test the model selection ability, we apply our approach sketched in Fig. 1 on simulated data at a typical SNR $= 20\,\mathrm{dB}$, the interest here being to find the smallest value v for reaching a correct decision in the model selection.



(a) single crack (b) double cracks

Fig. 5. Eddy-current test examples: a non-ferromagnetic plate effected by a single crack (a) or by two cracks (b).

TABLE II
PLATE, COIL AND INSPECTION CONFIGURATIONS

Plate			
thickness (h_t)	1.25 mm	conductivity	1 MS/m
Coil			
turns	140	outer radius	1.6 mm
inner radius	0.6 mm	height	0.8 mm
frequency	150 kHz		
Inspection			
lift-off	0.5 mm	spacing step (x)	0.5 mm
number of measurements	9×33	spacing step (y)	0.5 mm

B. Accuracy and computational demand of sparse grid surrogate model

Before conducting Bayesian model selection, we fisrt need to train databases for both models of concern. In this work, databases are trained by using the adaptive method described in [25] where outputs are simulated by a MoM [26], input parameters are located on sparse grids on the uniform prior space $\mathcal{U}$ the bounds of which are given in Tab. III.

TABLE III
PARAMETER RANGES USED IN THE DATABASE TRAINING.

Model	**parameters ranges**					
	a			d		
$\mathcal{M}_1$ inferior bound (mm)	2			0.125		
superior bound (mm)	10			1.125		
	a_1	d_1	a_2	d_2	w	v
$\mathcal{M}_2$ inferior bound (mm)	2	0.125	2	0.125	0	0.01
superior bound (mm)	10	1.125	10	1.125	2	1

For a sparse grid database, once we know the number of unknown parameters N, the number of nodes in the database J depends only on the depth of the sparse grid d_s, definition seen [27]. The relation among J, N and d_s can be approximately described by an asymptotic expression

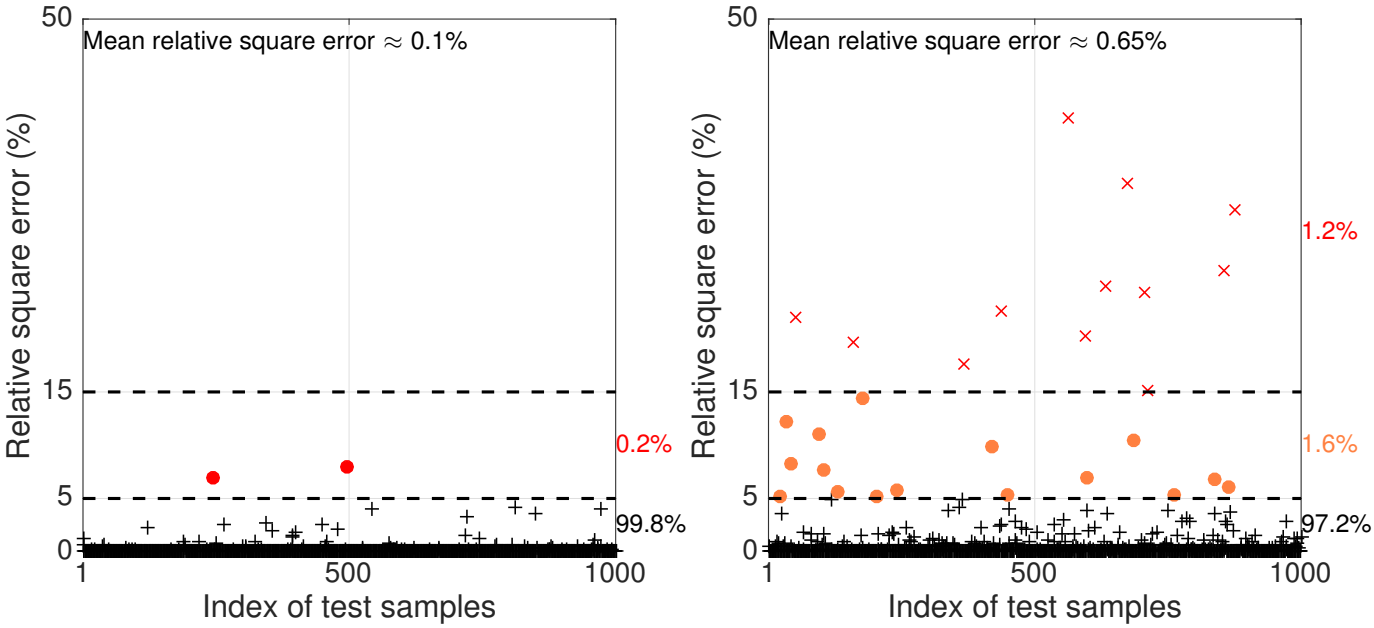
$$J = \mathcal{O}\Big(2^{d_s}\big(d_s \log 2\big)^{N-1}\Big). \tag{22}$$

The choice for d_s usually compromises between the accuracy of surrogate model and the computational cost of the database training [25], [27].

In our implementation, d_s is chosen adaptively according to the smoothness of the observations of impedance variations. We start our database training from $d_s = 0$ and increase it until the requirement on interpolation accuracy of the model is satisfied. We found that $d_s = 6$ is a good compromise between the interpolation accuracy and the computational cost for both models $\mathcal{M}_1$ and $\mathcal{M}_2$. The corresponding number of nodes in sparse grid databases $J = 176$ for $\mathcal{M}_1$ and $J = 10256$ for $\mathcal{M}_2$, respectively.

To verify the interpolation accuracy, we conducted an additional analysis where the interpolated data from the surrogate model are compared to the simulated data from MoM [26] at 1000 random test samples. Fig. 6 shows the results of the relative square errors. We see that the average relative error is less than 1% for both models. However, there are still a few of them

with relative errors larger than 15%. By checking on the impedance variation signal, we found that those correspond to cases where both cracks are of small sizes. Since the impedance variation is small, the relative error becomes large whereas the absolute error remains small. So globally, we consider that the surrogate model with sparse grid depth $d_s = 6$ is accurate enough for our following analysis.



(a) single-crack model (b) double-crack model

Fig. 6. Relative square errors between simulated data and interpolated data with sparse grid at depth 6 for 1000 random test samples.

In terms of the computational cost, on a standard 3.4 GHz PC, one simulation by the MoM [26] takes seconds while one sparse grid interpolation costs milliseconds even for the 6-parameter case. This provides a considerable speed-up when many repeated runs are required in the inversion. Noticing that MoM simulations in database training is totally independent among nodes, they can be run in parallel on multi-threads and multi-processors to reduce the time consumption in database training. In practical ECT, more complicated electromagnetic simulations can be dealt with; therein the gain provided by applying a sparse grid surrogate model would be higher [25].

C. Data simulation

As already indicated, measurements from models $\mathcal{M}_1$ and $\mathcal{M}_2$ can be very close to each other if v is small. Then, to distinguish one from the other is challenging. So, the smallest v for which the proposed approach is able to make the correct selection should be a good indicator for its model selection ability.

In our test, eight simulations of synthetic noise-free data are performed, one using the single-crack model, the others using the double-crack model with varying distances v from $0.04\,\mathrm{mm}$ to $0.3\,\mathrm{mm}$. Five random draws of Gaussian noise of same variance are then added to noise-free data leading to a set of forty noisy data simulations. The noise variance σ_y^2 is obtained using $\mathrm{SNR} = 20\,\mathrm{dB}$ in the following equation

$$\sigma_y^2 = \|f(\boldsymbol{x}, \mathcal{M})\|^2_{\mathrm{mean}} 10^{-\frac{\mathrm{SNR}}{10}}, \tag{23}$$

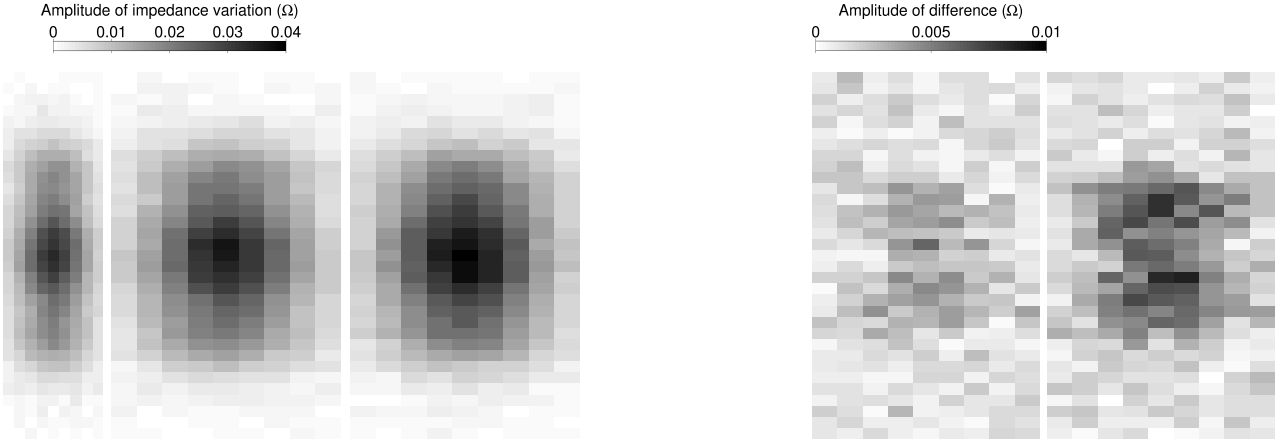
$\|f(\boldsymbol{x}, \mathcal{M})\|^2_{\mathrm{mean}}$ being the average l_2 norm of the noise-free data on the entire scanning map. The *N-MultiNest* is then applied on this set of simulated data. Using five simulations for each given v enables us to analyze the variance evolution of evidence approximation versus noise draws for the same SNR, all the other parameters remaining the same as the ones given in Tab. IV. In such a configuration, the single-crack case is equivalent to the double-crack case when $v = 0\,\mathrm{mm}$. In order to show the

TABLE IV
FIXED FLAW PARAMETERS USED IN DATA SIMULATION.

Model	fixed parameters (mm)
$\mathcal{M}_1$	$a = 6.6$, $d = 0.65$
$\mathcal{M}_2$	$a_1 = 5$, $d_1 = 0.65$, $a_2 = 5$, $d_2 = 0.65$, $w = 1.6$

difficulty of the model selection problem, Fig. 7 shows the map of the amplitude of the variation of impedance of the simulated

noisy data for $v = 0\,\mathrm{mm}$ (Fig. 7a, single-crack model), $v = 0.12\,\mathrm{mm}$ (Fig. 7b) and $v = 0.3\,\mathrm{mm}$ (Fig. 7c). As exemplified in Fig. 7d which gives the map of the absolute values of the difference between the variation of impedance for $v = 0\,\mathrm{mm}$ (Fig. 7a) and for $v = 0.12\,\mathrm{mm}$ (Fig. 7b), it is difficult to distinguish between $\mathcal{M}_1$ and $\mathcal{M}_2$ for $v = 0.12\,\mathrm{mm}$ whereas the difference increases when v increases as expected and as shown in Fig. 7e for $v = 0.3\,\mathrm{mm}$. Hence, it is interesting to see whether the Bayesian model selection approach is able to distinguish between the two models despite these slight differences.



(a) single crack (b) double cracks $v = 0.12\,\mathrm{mm}$ (c) double cracks $v = 0.3\,\mathrm{mm}$ (d) difference (b)-(a) (e) difference (c)-(a)

Fig. 7. Amplitude images of simulated impedance variations at SNR $= 20\,\mathrm{dB}$ for $v = 0.12\,\mathrm{mm}, 0.3\,\mathrm{mm}$ and their differences compared with the single-crack case.

D. Choice of threshold ϵ_L

As we mentioned in § III-B, ϵ_L is a small positive value thresholding the "non-zero" likelihood values. Here, "non-zero" means that its contribution to the evidence estimation is not ignorable. Therefore, the choice of ϵ_L depends on the likelihood distribution $p(\boldsymbol{y}|\boldsymbol{x}, \mathcal{M})$. It compromises between the saved warm-ups and the approximation error introduced by pre-estimation of posterior volume in evidence. In our algorithm 2, ϵ_L plays a role only in the estimation of the initial evidence $\hat{Z}^0$ which corresponds to the integral of the likelihood on $\mathcal{D} \cap \mathcal{I}'$. Even though the likelihoods are small on $\mathcal{D} \cap \mathcal{I}'$, $\hat{Z}^0$ can still be significant if the volume of $\mathcal{I}'$ is large enough. Thus, the choice of ϵ_L should be small enough so that the initial evidence $\hat{Z}^0$ remains relatively small compared to the total evidence $\hat{Z}$.

In our examples here, $p(\boldsymbol{y}|\boldsymbol{x}, \mathcal{M})$ is in the order between 10^{-100} and 10^{-2000}. We chose to use a relative small $\epsilon_L = 10^{-100}$ in order to guarantee that the additional error introduced on the evidence is ignorable. To justify this choice, we can compare the initial evidence $\hat{Z}^0$ with the final estimated evidence $\hat{Z}$, as done in § IV-E. In the implementation, all calculations involving likelihoods are performed after applying the logarithm in order to avoid exceeding the data precision limit.

E. N-MultiNest evidence estimation

We apply *N-MultiNest* described in § III on the forty sets of simulated noisy data respectively for $\mathcal{M}_1$ and $\mathcal{M}_2$. Fig. 8 shows the estimated model evidences and their averages at the eight v values. We observe that there is no overlap between stars and circles when $v > 0.12\,\mathrm{mm}$, which indicates that the evidence of $\mathcal{M}_2$ is always larger than that of $\mathcal{M}_1$. Even if, as shown in Fig. 7d and said earlier, no difference can be observed on the measurements for $\mathcal{M}_1$ and $\mathcal{M}_2$, the proposed approach can perform correct model selection with a probability close to 100%.

Tab. V shows the average results over the five independent simulations. Due to the estimation uncertainty for model evidence, decisions on model selection should be made according to

$$\begin{cases} \log\left[r(\mathcal{M}_1, \mathcal{M}_2)\right] > 10 & \mathcal{M}_1 \\ \left|\log\left[r\left(\mathcal{M}_1, \mathcal{M}_2\right)\right]\right| \leq 10 & \text{difficult to tell} \\ \log\left[r(\mathcal{M}_1, \mathcal{M}_2)\right] < -10 & \mathcal{M}_2 \end{cases} \tag{24}$$

instead of comparing $r(\mathcal{M}_1, \mathcal{M}_2)$ directly with 1.

Let us remind that one of our main contributions in this work is the use of an improved NS algorithm to approximate the model evidence. The proposed *N-MultiNest* saves warming-up time due to the use of pre-estimated posterior volume. In order to show the interest of our approach, we provide in Tab. V the pre-estimated posterior volumes $\hat{u}_p$, the equivalent warm-up iterations $N_{\text{warm-up}}$ and the total number of *N-MultiNest* samples K, $N_{\text{warm-up}}$ being calculated by taking $\hat{u}_p$ into Eq. (15).

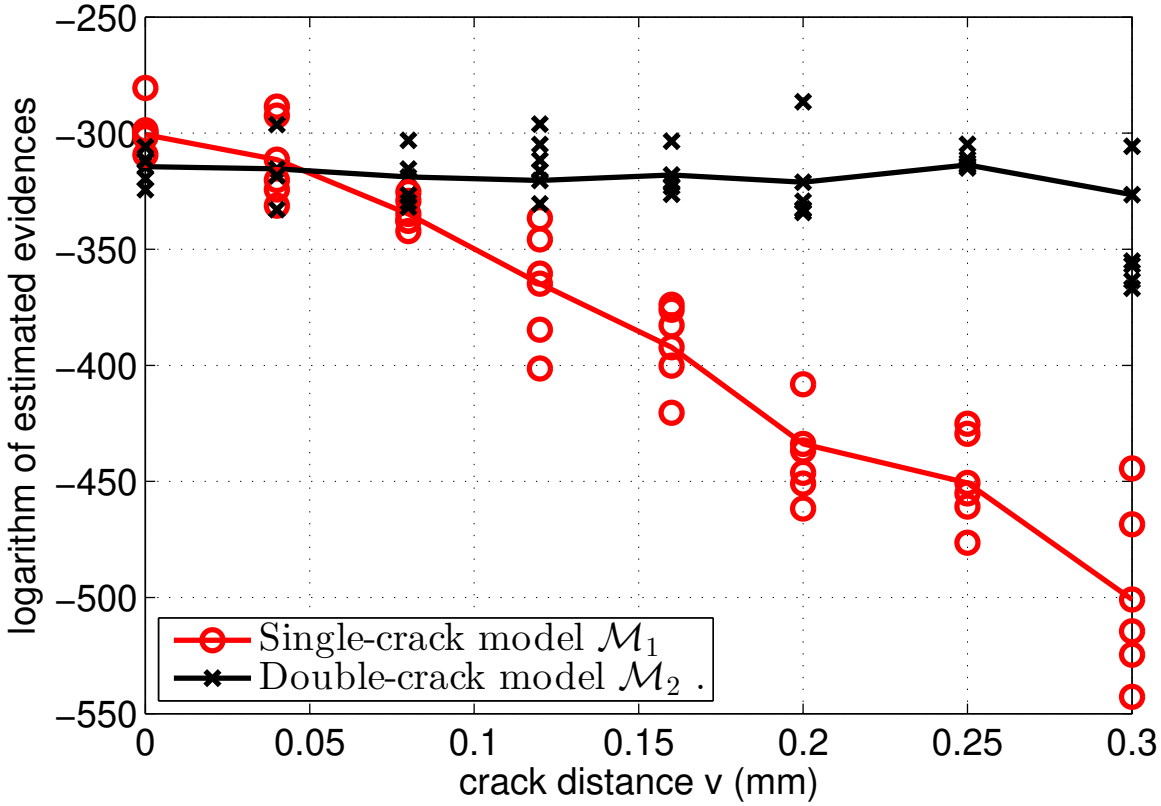


Fig. 8. Estimated model evidences by using *N-MultiNest*, solid lines being averages of five independent simulations.

TABLE V
AVERAGE PRE-ESTIMATED POSTERIOR VOLUMES, SAVED WARM-UP ITERATIONS, TOTAL *N-MultiNest* SAMPLES AND BAYES FACTORS VARYING v FROM 0 mm TO 0.3 mm. NUMBER OF ACTIVE SAMPLES $N_a = 200$, SNR $= 20$ dB.

Crack distance v (mm)	**single-crack model** $\mathcal{M}_1$			**double-crack model** $\mathcal{M}_2$			**Bayes factor**
	$\frac{\hat{u}_p}{(10^{-2})}$	$N_{\text{warm-up}}$	K	$\frac{\hat{u}_p}{(10^{-2})}$	$N_{\text{warm-up}}$	K	See Eq. (24)
0.00	0.27	1180	1412	6.79	553	5410	6
0.04	0.41	1072	1640	6.13	558	5339	1
0.08	0.47	1097	1667	5.89	566	4363	−6
0.12	0.41	1076	1663	7.03	531	4242	−22
0.20	0.48	1001	1837	0.70	992	3579	−48
0.30	0.53	1072	1844	0.06	1469	3399	−72

From $\hat{u}_p$ shown in Tab. V, we observe that the posterior volumes are all less than 10% of the prior volume. This indicates that we are dealing with narrow posterior distributions. For the double-crack model, when $v = 0.3\,\text{mm}$, the likelihood concentrates only on less than 0.1% of the total prior volume. Compared to *MultiNest*, the proposed *N-MultiNest* saved up to 45% of iterations (case $v = 0$ mm for $\mathcal{M}_1$).

We also compared the initial evidence $\hat{Z}^0$ with the final estimated evidence $\hat{Z}^0$ to see whether the use of pre-estimated posterior volume will introduce additional error on evidence estimation, and we observed that $\frac{\hat{Z}^0}{\hat{Z}}$ are all less than 10^{-100} for all the tested cases. This indicates that $\epsilon_L = 10^{-100}$ is small enough to ensure that the additional error introduced on evidence estimation by the use of pre-estimated posterior volume is totally ignorable. Of course, a larger ϵ_L can be used to save more warm-ups but consequently it will increase the estimation uncertainty for evidence and enlarge the range of v for which the proposed model selection approach has difficulties to make a decision.

For practical applications, the metamodel database is often built on very large parameter ranges to include all possibilities. Therefore, the ratio between the posterior volume and the prior volume can be very small. For these cases, it is more interesting to use the pre-estimated posterior volume as the initial sampling space.

TABLE VI
ESTIMATED PARAMETERS AND STANDARD DEVIATIONS FOR ONE SIMULATION AT $v = 0\,\mathrm{mm}$ AND $v = 0.3\,\mathrm{mm}$, SNR $= 20\,\mathrm{dB}$.

True model and parameters (mm)	**estimated parameters $\hat{\boldsymbol{x}} \pm \boldsymbol{\sigma}$ (mm)**							
	single-crack model $\mathcal{M}_1$		double-crack model $\mathcal{M}_2$					
	a	d	a_1	d_1	a_2	d_2	w	v
$\mathcal{M}_1$ $a = 6.6$, $d = 0.65$	**6.603±0.168**	**0.649±0.007**	6.265±1.067	0.440±0.202	6.277±1.113	0.483±0.196	0.162±0.457	0.571±0.494
$\mathcal{M}_2$ $a_1 = 5$, $d_1 = 0.65$ $a_2 = 5$, $d_2 = 0.65$ $w = 1.6$, $v = 0.3$	6.180±0.128	0.716±0.006	**5.196±0.373**	**0.644±0.014**	**5.074±0.436**	**0.636±0.018**	**1.478±0.356**	**0.366±0.088**

F. Estimated parameters

As discussed in § III, in *N-MultiNest*, the model evidence is approximated by a series of samples with increasing likelihoods. These samples, also called *N-MultiNest* samples, follow the posterior distribution for the model of concern. So the unknown parameters can be estimated from these samples without extra computation as well as the variances:

$$\begin{cases} \hat{x}_i = \dfrac{1}{K}\displaystyle\sum_{k=1}^{K} x_i^k \\ \sigma_i^2 = \dfrac{1}{K(K-1)}\displaystyle\sum_{k=1}^{K}(x_i^k - \hat{x}_i)^2 \end{cases}, \quad i = 1, 2, \cdots, N \tag{25}$$

where the subindex i is the parameter index while the upper-index k is the *N-MultiNest* sample index. So the proposed approach can jointly solve the parameter inversion and the model selection.

As examples, we give in Tab. VI the parameter estimation results for one simulation of two situations $v = 0\,\mathrm{mm}$ and $v = 0.3\,\mathrm{mm}$ respectively, one with the single-crack model as the true model used in the simulation, the other with the double-crack model as the true model. On the diagonal axis, it shows the estimation results for the correct model. They are all close to the true parameter values. In terms of estimation uncertainty, w and v are difficult to estimate accurately compared with lengths and depths.

In order to illustrate how *N-MultiNest* behaves during the iterations, we show in Fig. 9 the distribution of *N-MultiNest* samples and their evolutions versus iterations for the situation $v = 0\,\mathrm{mm}$. The data are simulated from $\mathcal{M}_1$ in this case, but we know that both models can be valid. This has been confirmed by the estimated evidences displayed in Fig. 8.

From Fig. 9a, we see that the posterior distribution for $\mathcal{M}_1$ is relatively simple, i.e., it is similar with a Gaussian distribution and has one global maximum only. The nested sample obtained at each iteration yields decreasing likelihoods. If we draw the active samples at each iteration, their distribution narrows from pre-estimated posterior region down to a point.

For the double-crack model, it is difficult to show their distribution in 6D. We display in Figs. 10a, 10b their evolutions vs. iterations for $v = 0\,\mathrm{mm}$ and in Figs. 11a, 11b for $v = 0.3\,\mathrm{mm}$. For display simplicity, all parameter values are normalized with respect to the bounds given in Tab. III. We observe that the distribution of the *N-MultiNest* samples for the case $v = 0\,\mathrm{mm}$ is more complicated than for the case $v = 0.3\,\mathrm{mm}$. For the case $v = 0\,\mathrm{mm}$, we can imagine that there exist infinite combinations for the six unknown parameters. One obvious possibility is to make one of the cracks disappear. This means a_1 and d_1 (or a_2 and d_2) tending to zero. This is exactly the situation shown in Figs. 10a, 10b. Limited by the metamodel bounds, a_2 and d_2 cannot be zero, but are approaching their inferior bounds. For the case with a large gap in-between $v = 0.3\,\mathrm{mm}$, the evolution of *N-MultiNest* samples is simpler. They all converge to their true values.

Since the series of samples $\{\boldsymbol{x}^1, \boldsymbol{x}^2, \cdots, \boldsymbol{x}^K\}$ follow the posterior distribution $p(\boldsymbol{x}|\boldsymbol{y}, \mathcal{M})$, $1D$ marginal distributions for each parameter can also be estimated by the histograms of these samples, as displayed in Fig. 10d for the case $v = 0\,\mathrm{mm}$ and in Fig. 11d for the case $v = 0.3\,\mathrm{mm}$. The peaks of the $1D$ marginal distributions represent local or global maxima of the marginal posterior distributions. We can see that the case $v = 0\,\mathrm{mm}$ is much complicated than the case $v = 0.3\,\mathrm{mm}$.

G. Computational cost

The computational cost depends on the number of unknown parameters and the complexity of the posterior distributions. For the single-crack model, we see from Fig. 9a that the posterior distribution is similar with a Gaussian distribution and there are only two unknown parameters. Tab. VII gives the total computational times and the sampling acceptance rates for different cases. The computational time is obtained on a standard PC with 3.4 GHz CPU. For the double-crack model, the total computational time is inversely proportional with the acceptance rate which is only 0.3% at $v = 0\,\mathrm{mm}$ and reaches 18% at $v = 0.3\,\mathrm{mm}$. This is because the posterior distribution is much more complicated when $v = 0\,\mathrm{mm}$. As illustrated in Tab. V and Fig. 10d, the posterior volume is much larger; there are multiple local maxima and global maxima.

So far, we only discuss dual-model selection, yet the approach can be applied directly for multiple models. It only needs to run the *N-MultiNest* algorithm on each competing model. Since the computation is independent for each model, it can be performed in parallel, and the total computational time remains the same.

TABLE VII
NUMBER OF ACCEPTED SAMPLES K, ACCEPTANCE RATE r_{ACC} AND COMPUTATIONAL TIME t OF *N-MultiNest* ON A 3.4 GHZ PC FOR CASES $v = 0\,\text{mm}$, $0.12\,\text{mm}$ AND $0.3\,\text{mm}$.

Case	$v = 0\,\text{mm}$			$v = 0.12\,\text{mm}$			$v = 0.3\,\text{mm}$		
	K	r_{acc} (%)	t (min.)	K	r_{acc} (%)	t (min.)	K	r_{acc} (%)	t (min.)
$\mathcal{M}_1$	1412	45	3	1663	42	2	1844	20	4
$\mathcal{M}_2$	5410	**0.3**	**679**	4242	**3**	**67**	3399	**18**	**11**

V. CONCLUSION

In this work, we first introduce a general framework of Bayesian model selection based on a data-fitting surrogate model. The use of a surrogate model enables us to train the database off-line. It can help us to reduce significant computational time on-line in the inversion. Second, we propose an efficient sampling algorithm *N-MultiNest* for approximating the model evidence. It makes use of the fact that posterior distribution is often concentrated on a very small part in the prior space. Instead of sampling on the entire prior space as done by other nested sampling algorithms, we sample on the pre-estimated posterior space only. The pre-estimation of the posterior volume is carried out by using the nodes existing in the database trained off-line.

Based on an ECT example, numerical simulations show that the proposed approach has high model selection ability. To distinguish between a single-crack and a double-crack model, it is able to make a correct decision even when the gap between the two cracks is $0.12\,\text{mm}$ only. Considering that the measurements are performed with a spatial resolution of $0.5\,\text{mm}$, we believe that the proposed model selection has very high model selection ability in general.

Furthermore, the proposed approach is directly applicable to multiple model selection and can be used in all kinds of non-destructive applications. The only constraint is that the number of unknown parameters should not be too large. If not the algorithm will suffer both from the curse-of-dimensionality in the database training and the heavy computational cost in the model evidence approximation.

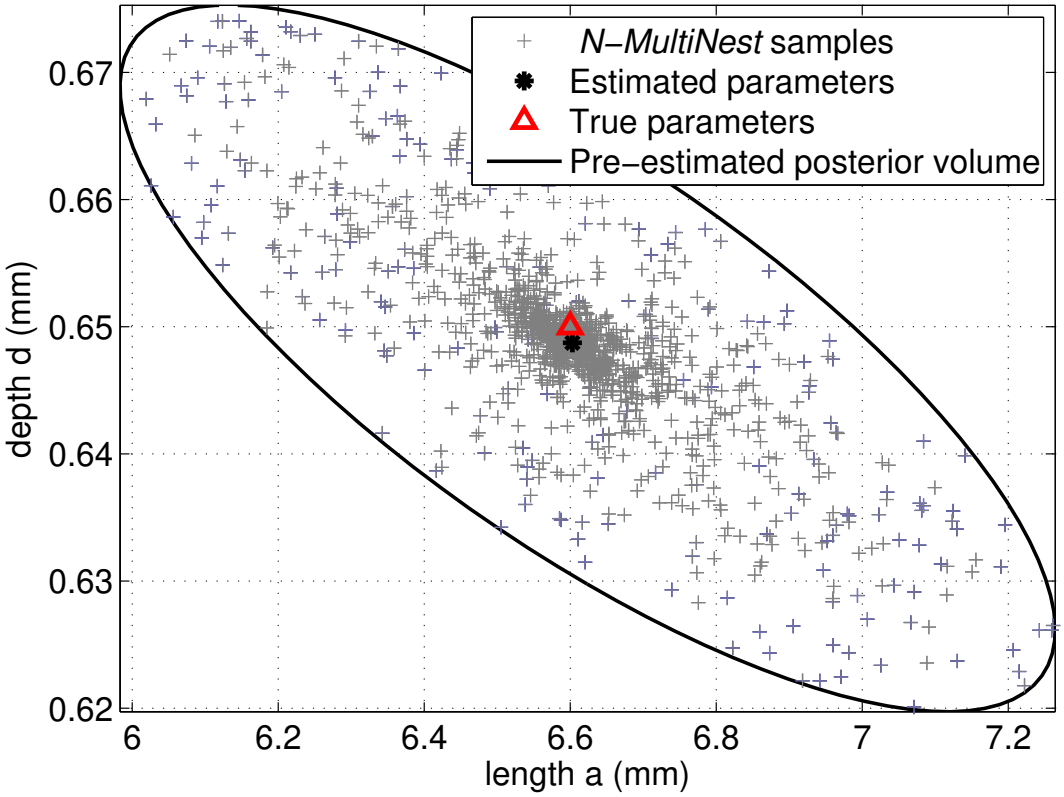
ACKNOWLEDGMENT

This work has been supported by the French National Research Agency in the framework of ByPASS project and the Hungarian Scientific Research Fund under grant no. K-111987.

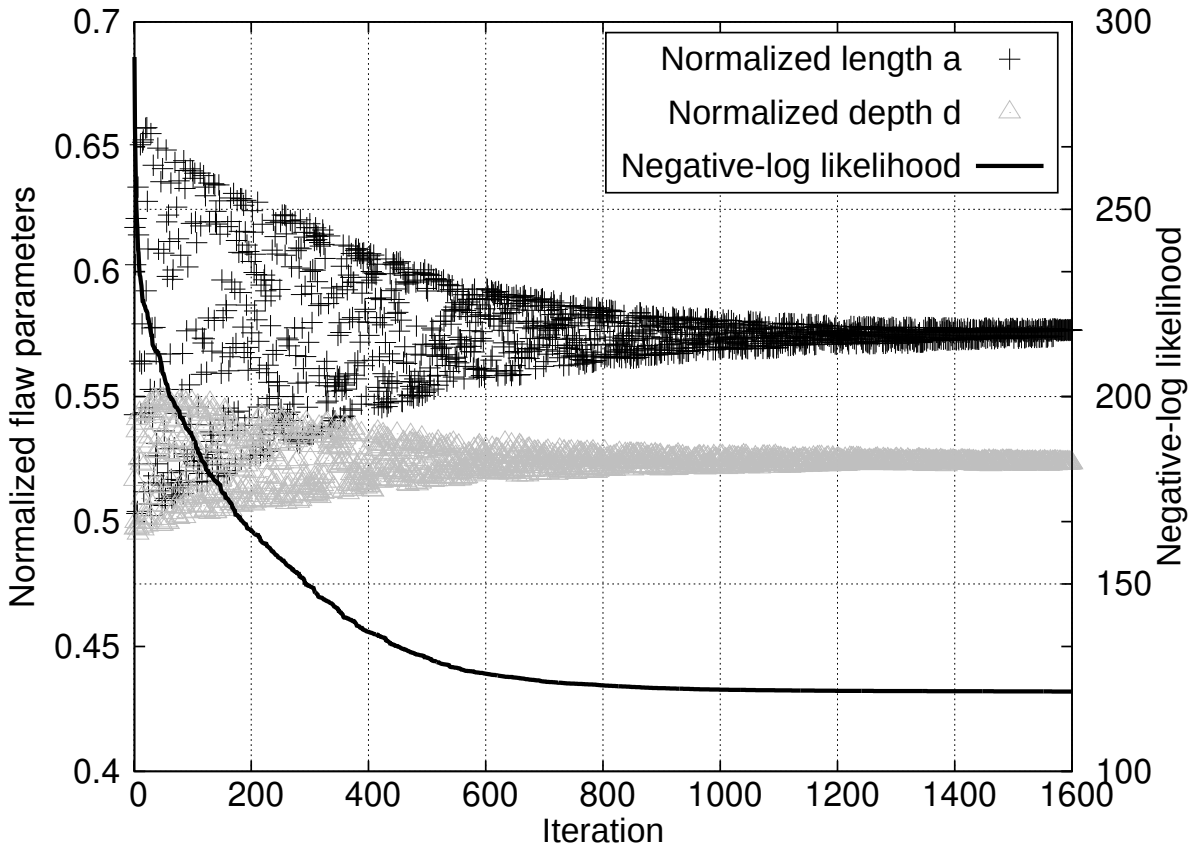
REFERENCES

[1] C. Cai, T. Rodet, and M. Lambert, "Influence of partially known parameter on flaw characterization in Eddy Current Testing by using a random walk MCMC method based on metamodeling," *Journal of Physics: Conference Series*, vol. 542, no. 1, pp. 012 009 1–6, 2014.
[2] S. Bilicz, S. M. Lambert, and J. Pávó, "Solution of inverse problems in nondestructive testing by a kriging-based surrogate model," *IEEE Transactions on Magnetics*, vol. 84, no. 2, pp. 495–498, 2012.
[3] H. A. Sabbagh, D. J. Radecki, S. Barkeshli, B. Shamee, J. Treece, and S. A. Jenkins, "Inversion of eddy-current data and the reconstruction of three-dimensional flaws," *IEEE Transactions on Magnetics*, vol. 26, no. 2, pp. 626–629, 1990.
[4] Y. Li, Z. Chen, and Y. Qi, "Generalized analytical expressions of liftoff intersection in PEC and a liftoff-intersection-based fast inverse model," *IEEE Transactions on Magnetics*, vol. 47, no. 10, pp. 2931–2934, 2011.
[5] A. Tamburrino, R. Fresa, S. S. Udpa, and Y. Tian, "Three-dimensional defect localization from time-of-flight/eddy current testing data," *IEEE Transactions on Magnetics*, vol. 40, no. 2, pp. 1148–1151, 2004.
[6] H. Akaike, "Prediction and entropy," in *A Celebration of Statistics*, A. C. Atkinson and S. E. Fienberg, Eds. Springer New York, 1985, pp. 1–24.
[7] R. E. Kass and L. Wasserman, "A reference Bayesian test for nested hypotheses and its relationship to the Schwarz criterion," *Journal of the American Statistical Association*, vol. 90, no. 431, pp. 928–934, 1995.
[8] R. E. Kass and A. E. Raftery, "Bayes factors," *Journal of the American Statistical Association*, vol. 90, no. 430, pp. 773–795, 1995.
[9] J. B. Johnson and K. S. Omland, "Model selection in ecology and evolution," *Trends in Ecology & Evolution*, vol. 19, no. 2, pp. 101 – 108, 2004.
[10] P. Green, "Reversible jump MCMC computation and Bayesian model determination," *Biometrika*, vol. 82, pp. 711–732, 1995.
[11] R. M. Neal, "Annealed importance sampling," *Statistics and Computing*, vol. 11, no. 2, pp. 125–139.
[12] N. Friel and A. N. Pettitt, "Marginal likelihood estimation via power posteriors," *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, vol. 70, no. 3, pp. 589–607, 2008.
[13] J. Skilling, "Nested sampling for general Bayesian computation," *Bayesian Analysis*, vol. 1, no. 4, pp. 833–680, 2006.
[14] P. Mukherjee, D. Parkinson, and A. R. Liddle, "A nested sampling algorithm for cosmological model selection," *The Astrophysical Journal*, vol. 638, pp. L51–L54, 2006.
[15] F. Feroz and M. P. Hobson, "Multimodal nested sampling: an efficient and robust alternative to mcmc methods for astronomical data analysis," *Monthly Notices of the Royal Astronomical Society*, vol. 384, no. 2, pp. 449–463, 2008.
[16] J. E. Lee and C. P. Robert, "Importance sampling schemes for evidence approximation in mixture models," *Bayesian Analysis*, vol. TBA, pp. 1–25, 2015.
[17] F. Feroz, M. P. Hobson, and M. Bridges, "MultiNest: an efficient and robust Bayesian inference tool for cosmology and particle physics," *Monthly Notices of the Royal Astronomical Society*, vol. 398, no. 4, pp. 1601–1614, 2009.
[18] F. Feroz, M. Hobson, E. Cameron, and A. N. Pettitt, "Importance nested sampling and the multinest algorithm," *Instrumentation and Methods for Astrophysics*, vol. preprinted on arXiv, pp. 1–28, 2014.
[19] J. P. C. Kleijnen and W. C. M. Van Beers, "Application-driven sequential designs for simulation experiments: kriging metamodeling," *The Journal of the Operational Research Society*, vol. 15, no. 8, pp. 303–325, 2004.
[20] N. V. Queipo, R. T. Haftka, W. Shyy, T. Goel, R. Vaidyanathan, and P. K. Tucker, "Surrogate-based analysis and optimization," *Progress in Aerospace Sciences*, vol. 41, no. 1, pp. 1–28, 2005.

[21] M. Bensetti, Y. Choua, L. Santandrea, Y. L. Bihan, and C. Marchand, "Adaptive mesh refinement and probe signal calculation in eddy current NDT by complementary formulations," *IEEE Transactions on Magnetics*, vol. 44, no. 6, pp. 1646–1649, 2008.
[22] A. I. J. Forrester and A. J. Keane, "Recent advances in surrogate-based optimization," *Progress in Aerospace Sciences*, vol. 45, no. 1-3, pp. 50–79, 2009.
[23] S. Bilicz, M. Lambert, and Sz. Gyimóthy, "Kriging-based generation of optimal databases as forward and inverse surrogate models," *Inverse Problems*, vol. 26, no. 7, p. 074012, 2010.
[24] R. Douvenot, M. Lambert, and D. Lesselier, "Adaptive metamodels for crack characterization in eddy-current testing," *IEEE Transactions on Magnetics*, vol. 47, no. 4, pp. 746–755, 2011.
[25] S. Bilicz, "Sparse grid surrogate models for electromagnetic problems with many parameters," *IEEE Transactions on Magnetics*, vol. 52, no. 3, pp. 1–4, 2016.
[26] J. Pávó and D. Lesselier, "Calculation of eddy current testing probe signal with global approximation," *IEEE Transactions on Magnetics*, vol. 42, no. 4, pp. 1419–1422, 2006.
[27] H.-J. Bungartz and M. Griebel, "Sparse grids," *Acta Numerica*, vol. 13, no. 5, pp. 147–269, 2004.
[28] L. G. Khachiyan and M. J. Todd, "On the complexity of approximating the maximal inscribed ellipsoid for a polytope," *Mathematical Programming*, vol. 61, no. 1, pp. 137–159, 1993.
[29] N. Moshtagh, "Minimum volume enclosing ellipsoid," *Convex Optimization*, vol. 111, pp. 111–112, 2005.
[30] P. Kumar and E. A. Yıldırım, "Minimum-volume enclosing ellipsoids and core sets," *Journal of Optimization Theory and Applications*, vol. 126, no. 1, pp. 1–21, 2005.
[31] M. J. Todd and E. A. Yıldırım, "On Khachiyan's algorithm for the computation of minimum-volume enclosing ellipsoids," *Discrete Applied Mathematics*, vol. 155, no. 13, pp. 1731–1744, 2007.

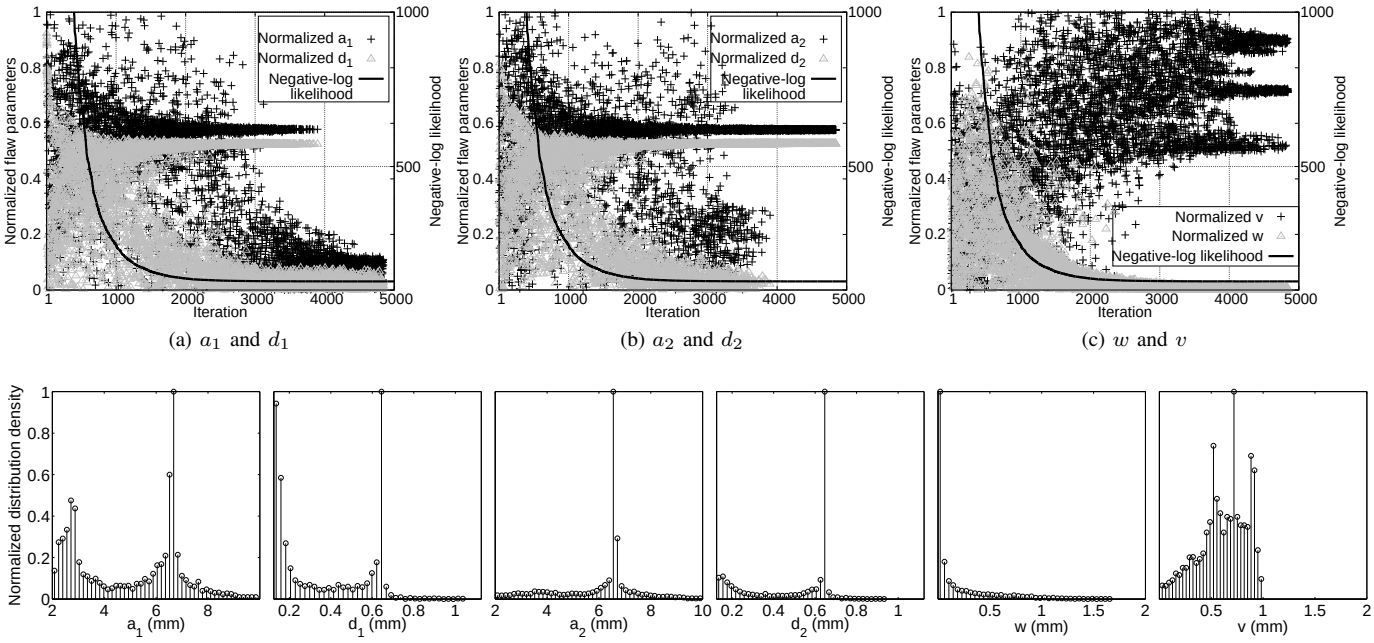


(a) $2D$ distribution of *N-MultiNest* samples



(b) normalized *N-MultiNest* samples and negative-log likelihoods vs. iterations.

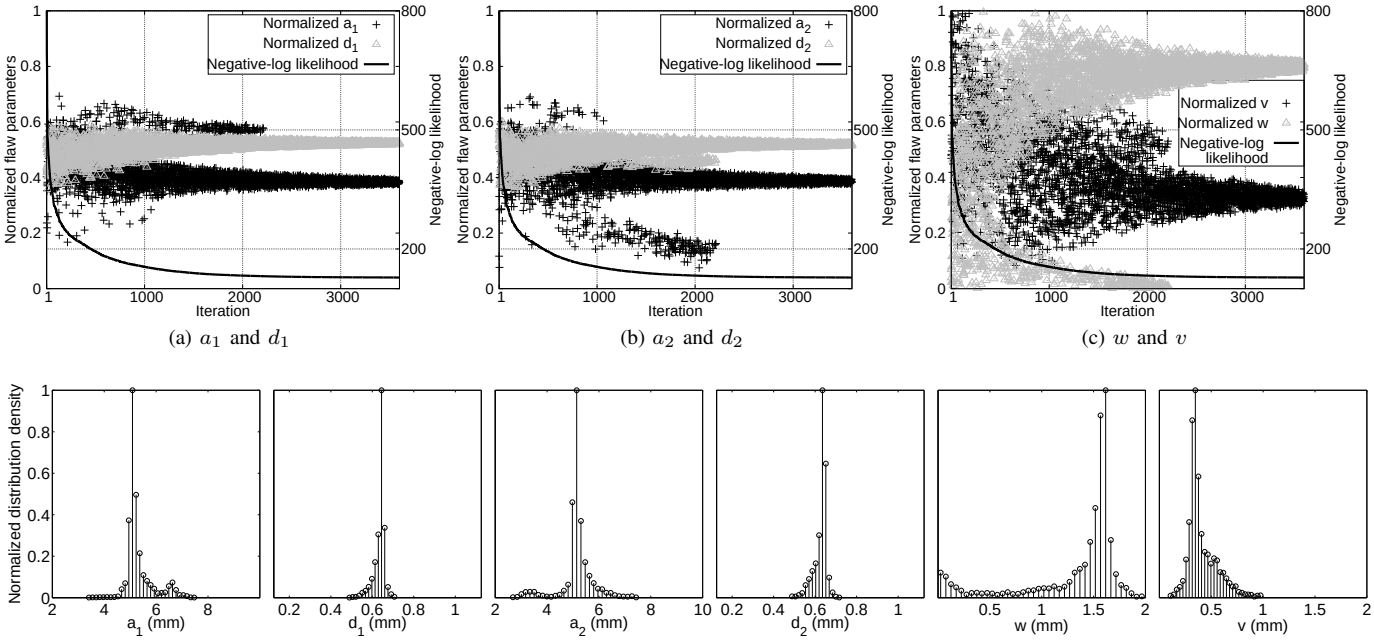
Fig. 9. *N-MultiNest* samples for the single-crack model at $v = 0\,\mathrm{mm}$, SNR$= 20\,\mathrm{dB}$. Normalizations are performed with respect to the parameter ranges given in Tab. III.



(a) a_1 and d_1 (b) a_2 and d_2 (c) w and v

(d) 1D marginal distributions

Fig. 10. Normalized *N-MultiNest* samples, negative-log likelihoods vs. iterations and 1D marginal distribution densities for the double-crack model at $v = 0\,\mathrm{mm}$, SNR $= 20\,\mathrm{dB}$. Normalizations are performed with respect to the parameter ranges given in Tab. III.



(a) a_1 and d_1 (b) a_2 and d_2 (c) w and v

(d) 1D marginal distributions

Fig. 11. Normalized *N-MultiNest* samples, negative-log likelihoods vs. iterations and 1D marginal distribution densities for the double-crack model at $v = 0.3\,\mathrm{mm}$, SNR $= 20\,\mathrm{dB}$. Normalizations are performed with respect to the parameter ranges given in Tab. III.