Technical article

Use of Adaptive Databases for Surrogate Modeling of Electromagnetic Problems, Illustration with NDE Examples

Abstract — **Numerical methods are frequently used for the analysis and design of electromagnetic devices or for the solution of inverse problems of computational electromagnetics. The precision provided by such simulators is usually fine but at the price of computational cost. In some applications this cost might be crucial or in some other applications users might require an easy-to-use environment dedicated to their particular problem. These lead us to consider cheap *surrogate models* in order to reduce the computation time still meeting the precision requirements. Among all available surrogate models, we deal with the generation and application of *adaptive database* of pre-calculated results combined with certain interpolators. In this paper, three adaptive sampling algorithms are investigated. Two of them –one being based on *simplex-mesh* refinement, the other is supported by the *kriging* interpolation– aim at generating a so-called *output-equidistant* database, i.e., the stored results are required to be equally repartitioned in the space of the measurable output data. The third sampling strategy is designed to improve directly the precision of the kriging interpolator (subsequently fitted to the samples stored in the database). The yielded databases can be used for the approximate solution of both forward and inverse problems, they can also be applied to build a problem-oriented easy-to-use simulator that can be operated without any preliminary knowledge of computational electromagnetics. Beyond the mere approximations, the structure of the output-equidistant databases carries considerable *meta-information* on the modeled forward problem. This meta-information can be exploited for the quantitative characterization of the related inverse problem. All approaches are illustrated by examples drawn from eddy-current nondestructive evaluation.**

I Introduction

Computational electromagnetics is widely used for the solution of various engineering problems. In a number of applications (e.g., certain nondestructive tests), computational load and time consumption are crucial, whereas the precision requirements are still demanding. However, an end-user can neither be expected to know the mathematical model in details nor have time to run overlong simulations. So, nowadays more and more emphasis is being put on the *emulation*, or *surrogate modeling*, [1]. Surrogate models intend to imitate the behavior of the true model as far as possible, but at a much less computational cost, however.

Many surrogate modeling techniques are sample-based, i.e., the true model is evaluated for certain feasible configurations (e.g., defect prototypes in nondestructive evaluation), and an interpolator is fitted to the observed data. The configuration is specified by the *input parameters* (spanning the input space), the measurable quantities are called the *output data*, and the set of corresponding input-output pairs is referred to as a *database* in the following. Once the database is generated, the "end-user" does not need the simulator any more, but only the database and the appropriate interpolator. Most of the computational load is then concentrated into the database generation stage, performed only once and off-line using the simulator operated by an "expert" who is not necessarily the end-user of the computational results. Consequently, the time consumption of the database generation is not of main interest. On the contrary, at the end-user's stage, the originally time consuming issues (e.g., the solution of an inverse problem or design optimization) can be addressed within a reduced computational budget.

The key points of such interpolation-like surrogate modeling are the choice of the samples and the applied interpolator. The sampling strategy can be chosen form a wide range provided in the framework of *Design-Of-Experiments* (DOE, or Experimental Design). DOE is the collective name of information-gathering techniques related to observations of a given phenomenon. At the beginning of DoE's long history, the word "experiment" referred to classical (e.g., physical, chemical) experiments, whereas nowadays running a computer code of a simulation is also considered as a "computer experiment".

As DOE plays a central role in our approaches, Subsection A is devoted to some traditional sampling strategies in order to provide a historical basis. Then, we give the biographic background of the domains in which our research efforts are concentrated. The simplex-mesh and the kriging-based meshless adaptive sampling strategies are concerned in Subsections B and C, respectively. Both method aims at generating a database consisting of samples being "equally spaced" in the space of the output data. Whereas such equidistant sampling can easily be achieved in the input space, the output-equidistant sampling is much more challenging. In Subsection D, a specific interpolation tool –the functional kriging– is briefly introduced.

A Traditional Design-Of-Experiments

Classical approaches (see, e.g., [2], [3], [4] as recent textbooks) are based on only geometrical considerations related to the repartition of input samples. These methods do not take the interpolator into account, i.e., the sampling is independent from the subsequent use of the samples. Well-known classical designs are the full-factorial design (spacing the input samples at the nodes of a regular grid in the input space) and the Latin hypercube sampling. The latter was introduced to ensure the uniform representation of the whole range of all input parameters in the sample-set, but at a much smaller sample number than at a full-factorial design ([5]). The essential work [6] gives a thorough review of the first attempts about computer experiments, moreover, proposes a method for the interpolation based on the samples. More recently, classical DOE approaches are discussed by [7] and [8] (with applications in control theory, citing almost 200 references).

A step towards adaptive DOE is to take into account the interpolator subsequently fitted to the samples (but still not accounting for the actually observed output values).

An example of such model-oriented DOE is the approach of sample-pattern optimization presented in [9], whose goal is to reduce the uncertainty of the kriging prediction in a geostatistical application. Kriging –a stochastic tool for function approximation– is dealt with in more details in Subsection C and in Sections V and VI, respectively.

Fully adaptive DOE methods account not only for the model but for the actually observed output samples as well. Thus, most of the adaptive methods are sequential sampling techniques, and the choice of the subsequent sample depends on all the previous observations. An up-to-date general overview is, e.g., [10], with more than 100 references. In the following three subsections, our specific domains of adaptive DOE are introduced.

B Mesh-based sampling

A remarkable manner of adaptive sampling is based on meshing methods, inspired by meshes of Finite Element Methods. Several papers deal with the optimal spatial sampling of multivariate scalar functions in order to interpolate them by means of simplex meshes [11, 12]. The key of optimal sampling is the equidistribution of the interpolation error among the mesh elements. In particular, it is proved to be efficient to use an *anisotropic* mesh whose edge lengths are adjusted according to the directions of the principal curvatures of the function. This is because in order to equidistribute the error of linear interpolation, the edge length should be longer in a low curvature direction, and shorter in a high curvature direction. These criteria are usually tackled by introducing a specific local metric based on the Hessian of the function, so that the triangular elements of the optimal sampling are locally isotropic and their volumes are globally equidistributed under this metric [13, 14].

Since our objective is the interpolation of multivariate multicomponent functions of arbitrary dimensions (or even functional data), we cannot use the above results – derived for scalar functions – directly. An approach worth mentioning here is the so-called co-triangulation [15], which was successfully applied to color image data ($m = 3$). However, the method seems to be computationally too expensive for higher number of data components (which can count some tens or even hundreds in e.g., a typical nondestructive testing application).

In order to extend the above concept of optimal sampling for multicomponent functions, we developed a specific sampling method that relies on the adaptive anisotropic triangulation of the input space as well [16]. In this method, the goal of adaptive mesh refinement is to balance the *distance* between pairs of points connected by an edge, all over the mesh. This "distance" however, is associated with the maximum estimated interpolation error along the given edge, instead of being measured in a specific local metric based on the Hessian, as mentioned above. It can be seen that specifically for nearest neighbour (NN) interpolation this edge length relates to the Euclidean distance of the data stored in the connected mesh nodes, which allows us to call it an "output-equidistant" sampling.

Among others, this paper addresses (in Section IV) the technical issue of generating an optimal sampling of a multicomponent function, with respect to its NN interpolation over simplicial mesh, by using existing and available meshing tools. To overcome the problem of anisotropy, in our improved method the mesh-based sampling is carried out in a hypothetical "control-space" in which the optimal mesh is seen isotropic and quasi-uniform (see Section IV). Note that although the mapping between the input space and some control space (computational space) is a concept already introduced in several papers (see e.g. [12, 13]), up to now the control space has not been used directly as the "site" of mesh generation, rather just as a useful abstraction to make the idea of anisotropy comprehensible.

Furthermore, in the proposed method, we apply mesh deformation based on spring analogy to let the data points find their position in the control space (see Section IV for more details). Very similar relaxation technique based on spring equilibrium is described in [17, 18]. Further related techniques based on particle system model are reported in [19, 20]. In this latter approach, proximity-based interacting forces (both attractive and repulsive) are defined among the data points, and a force-balancing configuration is to be found using dynamic solution, similarly to a spring system.

Although this work was solely inspired by the authors' experience in mesh generation and NDT, after extending the scope of looking for related works, it was soon realized that very similar techniques (mesh mapping, spring analogy, etc.) appear in various engineering applications. As a matter of fact, the common idea behind most of these works (including this one) can be designated as "generalized surface fitting" or "distance mapping". Accordingly, the applied methods are intended to capture low-dimensional structure of data living in high-dimensional spaces or at least to provide a low-dimensional representation of them. Popular application fields include geophysical modeling, computer graphics, data visualization, data mining, clustering and signal processing.

Two prevalent methods have to be highlighted. One of them is the self-organizing map (SOM), a type of artificial neural network invented by Kohonen. It converts complex, nonlinear statistical relationships between high-dimensional data items into simple geometric relationships on a low-dimensional (typically two-dimensional) display. This is carried out through a competitive self-learning algorithm that develops a topological ordering of the data. Its application for scattered data interpolation and surface fitting is presented in [21]. The other method is called generative topographic mapping (GTM) [22]. GTM is a statistically based approach intended for modeling continuous, intrinsically low-dimensional probability distributions embedded in high-dimensional spaces. It can be seen as a non-linear form of principal component analysis (PCA), the result of which is called the "principal manifold". One realization of GTM using elastic grids – based on a physical analogy with elastic membranes – together with its application to data visualization are presented in [23].

C Kriging-based meshless sampling

As an other approach, a meshless adaptive sampling strategy for the generation of output-equidistant databases is also discussed in this paper (Section V), [24, 25, 26]. The approach consists in an iterative optimization procedure, adding samples one-by-one, according to a given criterion of optimality. The sample-insertion rule aims at spreading out the output samples as much as possible in the output space. Since the optimization problem –to which the

sample-insertion rule boils down– appears to be computationally demanding, the realization of the method involves the *kriging* interpolation of the auxiliary objective functions.

By now, kriging –invented in the domain of geostatistics in the 1960s– has become a widely used tool of function approximation. Thorough discussions can be found in the textbooks, e.g., [27], [28] (from the original domain, geostatistics), [29] (a rigorous mathematical approach), [3] and [2] (containg a short but purposive summary from an engineering viewpoint). Among the vast amount of journal papers, let us mention [6] (first introducing kriging to model computer experiments), [30] as a quite recent overview and [31] discussing an application for the solution of eddy-current testing inverse problems.

D A novel manner of interpolation: functional kriging

A recent development of mathematics is the so-called functional kriging, being an extension of the original theory (of scalar function interpolation) to the interpolation of functional output data [32, 33]. Functional kriging has successfully been applied for surrogate modeling in eddy-current nondestructive evaluation, and has been compared to traditional (nearest neighbour, piecewise linear and radial basis functions (RBF)) techniques [34]. However, in [34], the samples were chosen in a naive way: a rectangular grid was defined in the input space, and samples were placed to the grid-nodes.

In Section VI, a functional kriging surrogate model is dealt with, but the sampling strategy is adapted both to the modeled forward problem and to the fitted kriging interpolator [35]. The method is a sequential procedure, solving an optimization task in each cycle in order to reduce the uncertainty of the resulted interpolator as much as possible. Our method is a combination of the one discussed in [36] (based on the so-called jackknife variance estimation) and a slight improvement inspired by an idea occurring in [37].

* * *

The paper is organized as follows. In Section II, the basic idea of surrogate modeling by means of a database (along with the formal definition of the output-equidistant database) is discussed. Following this, the illustrative NDT test example –being used throughout the paper– is outlined. Two methods for the generation of special *output-equidistant* databases are presented in Sections IV (the mesh-based approach) and V (the meshless method, using the kriging prediction). The use of functional kriging interpolation for surrogate modeling –along with an adaptive sampling strategy– is dealt with in Section VI. In Section VII, the possible practical applications of the methodologies are highlighted. Finally, some numerical results are given in Section VIII. Beyond the comparison of the interpolation capabilities of the yielded surrogate models, a manner of visualization of the sample distributions is presented. A simple tool is also given to draw quantitative conclusions on the related inverse problem.

Finally, let us suggest for further reading the papers coauthored by the authors of the present publication where most of the topics touched here are discussed in more details [16, 31, 38, 39, 40, 34, 41, 24, 25, 26].

II Surrogate modeling by means of a database

A Terms and notations

For the sake of generality, the approach is formalized in a general manner. Let us define a well-behaved forward operator $\mathcal{F}$ representing a forward problem ($\mathcal{F}$ is usually realized via a numerical simulation of the mathematical model of the underlying physical phenomena), an input vector $\mathbf{x} = [x_1, x_2, \ldots, x_p]^{\mathsf{T}}$ of p real parameters, and the corresponding output scalar function[1] $y(t)$. They are related as

$$\begin{aligned} \mathcal{F}: \quad \mathbb{X} &\mapsto \mathbb{Y} \\ \mathbf{x} &\mapsto y(t) = \mathcal{F}\{\mathbf{x}\}, \quad t \in T, \end{aligned} \tag{1}$$

$y(t)$ is square-integrable over the domain T, $\mathbb{X} \subset \mathbb{R}^p$ being the input space and the output space $\mathbb{Y}$ is defined as the codomain of $\mathcal{F}\{\mathbf{x}\}$:

$$\mathbb{Y} = \{y(t) : y(t) = \mathcal{F}\{\mathbf{x}\}, \forall \mathbf{x} \in \mathbb{X}\}. \tag{2}$$

$\mathbb{Y}$ is then a subset of the $L^2(T)$ space consisting of all square-integrable functions over T.

Let us assume that each x_k ($k = 1, 2, \ldots, p$) has a lower and an upper bound, $x_k^{\min}$ and $x_k^{\max}$, respectively. $\mathbb{X}$ can then be written as

$$\mathbb{X} = \left[x_1^{\min}, x_1^{\max}\right] \times \left[x_2^{\min}, x_2^{\max}\right] \times \cdots \times \left[x_p^{\min}, x_p^{\max}\right]. \tag{3}$$

An appropriate norm is defined on both $\mathbb{X}$ and $\mathbb{Y}$. In the input space, we have

$$||\mathbf{x}||_x = \sqrt{\sum_{k=1}^{p} \eta_k^2 \left(x_k - x_k^{\min}\right)^2}, \tag{4}$$

with the scaling factors η_k introduced to put more or less emphasis on a priori chosen parameters. Equal emphasis on all parameters is obtained if $\eta_k = \left(x_k^{\max} - x_k^{\min}\right)^{-1}$. In the $L^2(T)$ space (and so, in $\mathbb{Y}$) the following norm is used:

$$||y(t)||_y = \sqrt{\frac{1}{|T|} \int_T |y(t)|^2 \mathrm{d}t}, \tag{5}$$

where $|T|$ is the measure of the domain T in a certain appropriate sense, i.e., the norm of an output function is equal to its root mean square.

Let us define a database $\mathbb{D}_n$ as a set of n samples, each of them being a pair of input sample ($\mathbf{x}_i$) and output sample ($y_i(t) = \mathcal{F}\{\mathbf{x}_i\}$), $\forall i = 1, 2, \ldots, n$. Formally:

$$\mathbb{D}_n = \{(\mathbf{x}_1, y_1(t)), (\mathbf{x}_2, y_2(t)), \ldots, (\mathbf{x}_n, y_n(t))\}. \tag{6}$$

The surrogate model $\mathcal{G}_n$ is based on $\mathbb{D}_n$ and it approximates $\mathcal{F}$ everywhere in $\mathbb{X}$ as

$$\begin{aligned} \mathcal{G}_n: \quad \mathbb{X} &\mapsto \mathbb{Y} \\ \mathbf{x} &\mapsto \mathcal{G}_n\{\mathbf{x}\} = \mathcal{F}\{\mathbf{x}\}, \quad \forall \mathbf{x} \in \{\mathbf{x}_1, \ldots, \mathbf{x}_n\} \\ &\quad\ \mathcal{G}_n\{\mathbf{x}\} \approx \mathcal{F}\{\mathbf{x}\}, \quad \forall \mathbf{x} \in \mathbb{X} / \{\mathbf{x}_1, \ldots, \mathbf{x}_n\}. \end{aligned} \tag{7}$$

The main advantage of such sample-based surrogate modeling consists in concentrating the computational burden

[1] The output is not necessarily a function, it can be a (usually high dimensional) vector as well. Moreover, even if the output is theoretically considered as a function, in practice, only its discrete representation can be measured.

(the forward simulations) to the first stage. The generation of the database is needed to be performed only once, practically by "experts", equipped with the numerical simulator (which might be an expensive commercial software and/or more complex than an "end-user" can use it). In the second stage, the "end-user" must have only the tools for the interpolation and the database, respectively. In other words, the $\mathcal{G}_n$ surrogate model represents an "off-line" simulator of the forward problem.

In this paper, two interpolators are dealt with. In the case of the so-called output-equidistant databases (defined in the next subsection, generated by one of the two methods discussed in the sections IV and V), the Nearest Neighbour (NN) interpolation method is used, formalized as

$$\mathcal{G}_n\{\mathbf{x}\} = y_j(t), \text{ where } j = \arg\min_{i=1,2,\ldots,n} ||\mathbf{x} - \mathbf{x}_i||_x \qquad (8)$$

The main advantages of NN are easy interpretability (close inputs – supposedly similar outputs) and low computational cost. Moreover, the error of the NN interpolation is believed to be reduced by the output-equidistant sampling.

The second interpolation method used in this paper is the *functional kriging*. This tool –providing smooth and usually more precise interpolations than NN– will briefly be discussed in Section VI.

B Output-equidistant database

An output-equidistant database is characterized by the following two properties:

- For all outputs $y(t) \in \mathbb{Y}$, an output sample $y_i(t)$ exists which is "not too far" from $y(t)$. Let us define $\gamma(\mathbf{x})$ as the distance to the nearest sample:

$$\gamma(\mathbf{x}) = \min_{i=1,2,\ldots,n} ||\mathcal{F}\{\mathbf{x}\} - y_i(t)||_y \,, \qquad (9)$$

  then the first criterion is formalized as

$$\max_{\mathbf{x}\in\mathbb{X}} \gamma(\mathbf{x}) < \Delta, \qquad (10)$$

  with the positive constant Δ.

- The output samples are evenly distributed in the sense that none of the output samples is "too close" to an other one, and every output sample has its nearest neighbour "not too far". For a formal definition, let us introduce the minimal and maximal distances between the nearest neighbours:

$$d_{\min} = \min_{i=1,2,\ldots,n} \left(\min_{j=1,2,\ldots,n,\ j\neq i} ||y_i(t) - y_j(t)||_y \right),$$
$$d_{\max} = \max_{i=1,2,\ldots,n} \left(\min_{j=1,2,\ldots,n,\ j\neq i} ||y_i(t) - y_j(t)||_y \right). \qquad (11)$$

  Then we require –as second criterion– the equivalence of $d_{\min}$ and $d_{\max}$:

$$\frac{d_{\max}}{d_{\min}} = 1. \qquad (12)$$

If $\mathbb{D}_n$ simultaneously fulfils the criteria (10) and (12), we call it an output-equidistant database. The algorithms presented in the sections IV and V are designed to generate such output-equidistant databases. Beyond being adequate for the NN interpolation, the output-equidistant databases provide certain *meta-information* on the studied forward problem, which might be used to characterize the related inverse problem as well (Section VIII).

III Test example

The methods and considerations are illustrated with a simple example taken from the field of eddy current testing (ECT). Here we define the investigated configuration (Fig. 1). An infinitesimally thin crack affects a non-magnetic, homogeneous infinite metal plate with a thickness of 1.25 mm and an electrical conductivity of $\sigma = 10^6$ S/m. An air-cored probe coil (driven by a time-harmonic current of frequency $f = 150$ kHz) is scanning a centered rectangular surface of 5 mm × 20 mm above the damaged zone. The impedance change of the coil ΔZ (influenced by the crack) is measured at 11×41 regularly spaced positions along α and β. The output signal $y(t)$ is then the impedance of the coil, t being related to the position of the coil (denoted by crosses in Fig. 1).

The EM phenomenon is modeled using a surface integral approach. The numerical simulation –representing $\mathcal{F}$– is based on the Method-of-Moments (for details, see [42]).

Table 1: Bounds of the input parameters.

Ex.	A (mm)		B (mm)		L (mm)		D (%)	
	min	max	min	max	min	max	min	max
2-par.	0		0		1	10	5	90
3-par.	−1.5	1.5	0		1	10	5	90
4-par.	−1.5	1.5	−1.5	1.5	1	10	5	90

The crack is rectangular-shaped, perpendicular to the surface of the plate and it opens to the side opposite with the coil. The configuration can be described using four parameters: L and D are the length and the depth (in % of the thickness of the plate) of the crack, respectively; A and B are the co-ordinates of the center of the crack along α and β. Three configurations are then used to illustrate our approaches:

- 2-parameter example, where only two parameters (L and D) are varying, A and B being both 0.
- 3-parameter example, the x co-ordinate of the center of the crack (A) is added as input parameter.
- 4-parameter example, all four parameters are now considered as input parameters.

The minimum and the maximum values of each parameter are defined in Table 1.

IV Generation of output-equidistant databases: the meshing approach

In this section a procedure for obtaining output-equidistant sampling based on meshing methods inspired by meshes of finite element method is discussed. The obtained database is known to be optimal for NN interpolation, however, the described procedure can be adapted to other interpolation methods through the appropriate definition of the distance. The basic idea behind the database generation method can be explained as follows.

We introduce an auxiliary space of points $\boldsymbol{\xi} \in \Xi \subset \mathbb{R}^p$. Ξ is called the *control space*. We attempt to establish a local mapping $\ell : \Xi \to \mathbb{X}$ in such a way that the projection of the manifold $\mathcal{F}(\mathbb{X})$ onto Ξ approximately preserves the distances of points, at least locally within small neighbourhoods. This mapping is realized by the local barycentric mapping between two corresponding meshes living in $\mathbb{X}$ and

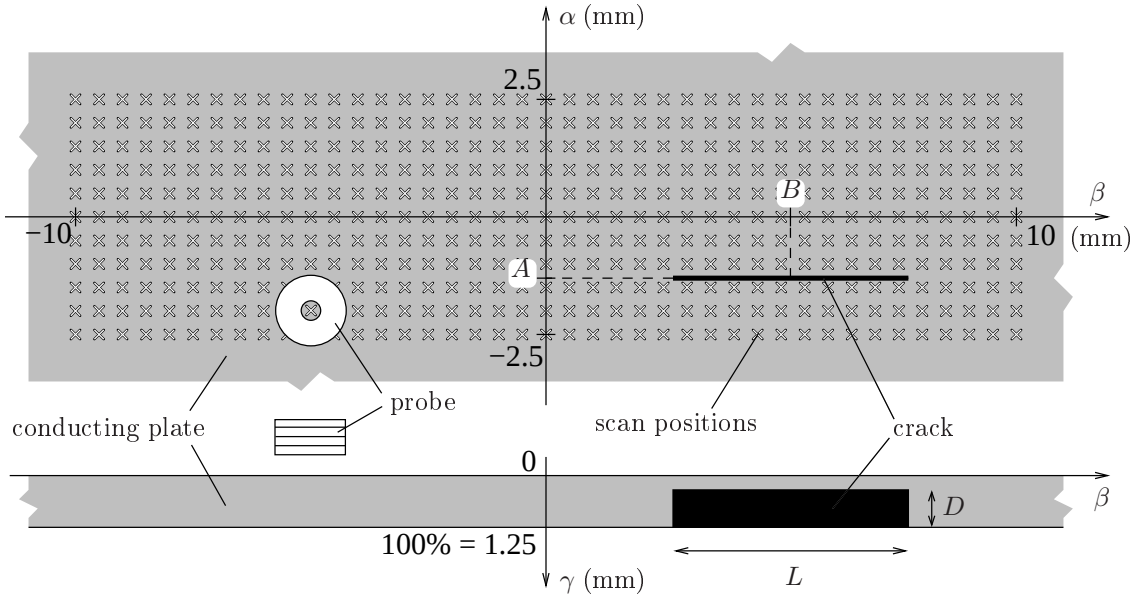


Figure 1: Illustration of the studied benchmark problem. Probe positions are marked by crosses (X).

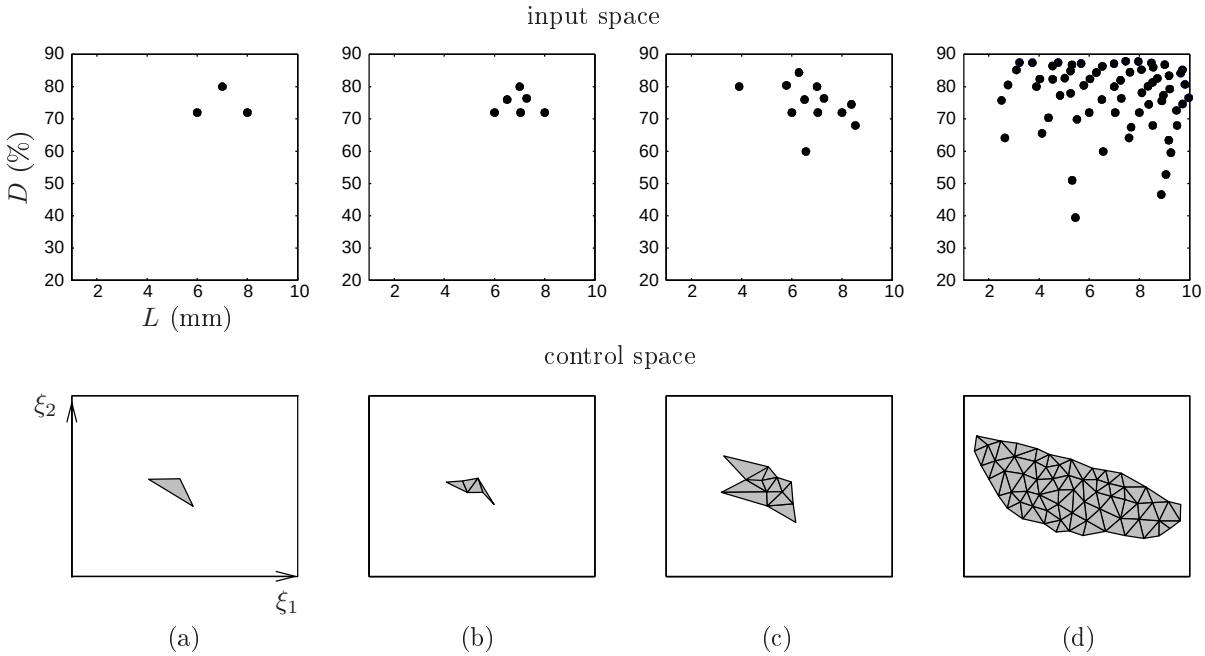


Figure 2: Steps of database generation in the 2-parameter example: the seed element (a), refinement (b), growth (c) and the final database after subsequent steps of refinement and growth (d).

in Ξ, respectively. The nodes of the mesh in $\mathbb{X}$ represent the $\mathbf{x}_i$ input-samples. Therefore the corresponding mesh in Ξ has to be generated in a way that its edge lengths reflect the output-distances between the pairs of nodes:

$$||\boldsymbol{\xi}_i - \boldsymbol{\xi}_j||_\xi \approx ||\mathcal{F}(\ell(\boldsymbol{\xi}_i)) - \mathcal{F}(\ell(\boldsymbol{\xi}_j))||_y \qquad (13)$$

for all nodes numbered i and j that are connected by an edge. The norm $||\cdot||_\xi$ in Ξ is defined similarly to the one in $\mathbb{X}$ defined by (4).

After all, the sampling of the forward operator can be considered optimal in the above terms if the corresponding mesh in Ξ looks isotropic and uniform. However, in the practice we use this relationship in the opposed direction. We generate a "quasi-uniform" mesh in the control space using adaptive refinement and smoothing. This procedure controls the appropriate sampling of the forward operator, which normally yields a "quasi-optimal" sampling. The method is described in the followings. For an illustration, see Fig. 2.

We start with $p+1$ predefined or random samples, $(\mathbf{x}_i, y_i(t))$, $i = 1, \ldots, p+1$, which form a simplex called the *seed element* (Fig. 2a). The first task is to find the picture of the seed element in Ξ. Having no better idea we let all $\boldsymbol{\xi}_i = \mathbf{x}_i$. From this point the following iterative procedure is carried out.

Step 1: relaxation. To let the points find their "right places" in Ξ, we utilize a mesh relaxation technique based upon spring analogy, which is often applied in the finite element solution of moving boundary problems [43]. In our implementation, the edges of the mesh are considered as spring segments, having an equilibrium length equal to the output-distance between the connected points. Point positions are updated one-by-

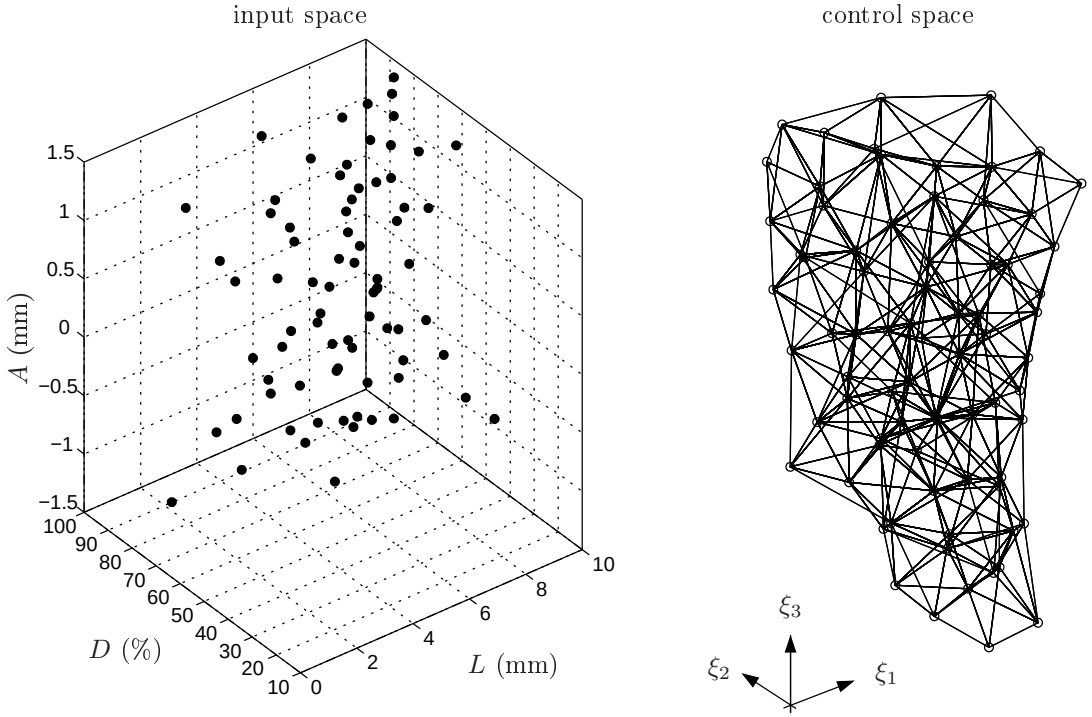


Figure 3: Optimized database in the 3-parameter example: distribution of the input samples (left) and the corresponding 3-dimensional mesh in the control space (right).

one to satisfy the local equilibrium condition,

$$\mathbf{0} = \sum_{j=1}^{\nu} \vartheta_{ij} \left(1 - \frac{||y_i(t) - y_j(t)||_y}{||\boldsymbol{\xi}_i - \boldsymbol{\xi}_j||_\xi} \right) (\boldsymbol{\xi}_i - \boldsymbol{\xi}_j) \quad (14)$$

where ϑ_{ij} is the spring stiffness, ν is the number of points connected to the ith point in the mesh. This is repeated until the migration of the points has an end. Note that no constraints are placed on the boundary nodes of the mesh.

Step 2: refinement. Each simplex having at least one edge longer than the prescribed limit Δ in (10) is refined by inserting a new point, say $\boldsymbol{\xi}^*$, into its circumcenter (Fig. 2b). Each point insertion has to be followed by the evaluation of the forward problem, i.e. the computation of $\mathcal{F}(\ell(\boldsymbol{\xi}^*))$, usually by calling an external program.

Step 3: growth. In order to explore the manifold $\mathcal{F}(\mathbb{X})$, some new points outside the existing mesh are inserted (Fig. 2c). Let $\boldsymbol{\xi}^*$ be a tried point. If the output-distance of $\boldsymbol{\xi}^*$ measured from any existing point is smaller than Δ, or $\mathbf{x}^* = \ell(\boldsymbol{\xi}^*) \notin \mathbb{X}$, the point is ignored. If there was no new point inserted in Step 2 and Step 3, the procedure terminates, otherwise it returns to Step 1.

Remarks:

- The iteration procedure involving local equilibrium in Step 1, which is similar to the Gauss-Seidel iteration for solving linear systems, may exhibit poor convergence properties especially in higher dimensions. More sophisticated solutions on spring system equilibrium can be found in [18].
- Node collision and the flip of simplices during relaxation can be more or less prevented by choosing the spring constant equal the inverse of the data-distance (rest length): $\vartheta_{ij} = 1/\,||y_i(t) - y_j(t)||_y$ [43].
- For the point insertion in Step 2 and Step 3 we use the free software *Qhull* [44], which is a powerful algorithm for creating Delaunay-meshes. However, its limitation in dimensions to $p < 9$ imposes the same practical limit for the database generation method.
- The mapping ℓ, which is used in Steps 2–3, is subject to slight changes during Step 1 because of node relocations, but no critical behaviour has been experienced in connection with this effect yet.
- The method assumes that the manifold $\mathcal{F}(\mathbb{X})$ (i.e., the output space) is continuous and relatively smooth. Even if this is the case, the resulting mesh in $\mathbb{X}$ is not necessarily valid, meaning that it contains tangles or holes. This is not a fundamental problem if the connectivity information is not used by the interpolator (e.g. the samples of $\mathcal{F}$ are used for training neural networks).
- The method can be adapted to other *local* interpolation methods (in which the refinement of the sampling improves the quality of interpolation on a local basis) through the condition (13) by changing its left hand side to the maximum interpolation error estimated along the given edge. We have successfully completed and tested this extension for linear interpolation using the original database generation method [16]. Note that the mesh database optimized for linear interpolation can be much more economic than the database optimized for NN interpolation.

In Fig. 2, an output-equidistant database generated for the 2-parameter example is shown, whereas in Fig. 3, the final database in the case of the 3-parameter example is visualized. Both pictures show the equally distributed mesh

in the control space and the irregular sampling of the input space.

V Generation of output-equidistant databases: the kriging approach

A Incremental distance-based sampling procedure

An other possible way of generating an output-equidistant database is explained in this section: the sought equidistant output sample distribution is attempted to achieve by using the *maximin method*. In our context, *maximin* means maximizing the minimal distance between any pair of output samples. In so doing, one can heuristically expect for databases being output-equidistant in the sense discussed in Section II. An incremental algorithm for the database generation is proposed, consisting in two main stages:

1. at the beginning of the procedure, some initial samples are chosen by hand (e.g., by using one of the classical DOE tools, such as Latin hypercube sampling);
2. then, samples are inserted one-by-one, according to a maximin criterion.

For a formal description, let us introduce the so-called *distance functions*:

$$Q_k(\mathbf{x}) = ||\mathcal{F}(\mathbf{x}) - y_k(t)||_y, \text{ where } k = 1, 2, \ldots, n. \quad (15)$$

The database consists of n samples, and $Q_k(\mathbf{x})$ is related to the kth sample. If one intends to add the next, $(n+1)$th sample to the database (in the incremental loop), the evaluation of the following expression is needed:

$$\mathbf{x}_{n+1} = \arg\max_{\mathbf{x}\in\mathbb{X}} \left[\min_{k=1,2,\ldots,n} Q_k(\mathbf{x}) \right]. \quad (16)$$

Since the algorithm looks only one-step ahead and no back step is allowed, the sample distribution is not an exact solution of the maximin problem. Moreover, the initial samples are not chosen according to the criterion (16). To overcome these pitfalls, a sample-removing strategy must also be included in the sampling strategy. After, say, every two sample insertions, one sample is removed from the database as follows. Let $d_{\min}^{(-r)}$ the distance between the pair of the nearest output samples if the rth sample was removed:

$$d_{\min}^{(-r)} = \min_{i=1,2,\ldots,n,\ i\neq r} \left(\min_{j=1,2,\ldots,n,\ j\neq r,i} ||y_i(t) - y_j(t)||_y \right). \quad (17)$$

In other words, the sample which would increase the most the $d_{\min}$ minimal distance (12) if it was not included in the database, it is indeed removed.

B Realization by kriging interpolation

The optimization problem (16) is computationally expensive since the forward operator $\mathcal{F}$ is involved in the distance functions $Q_k(\mathbf{x})$. An approximate solution is proposed by using the *kriging* interpolation. Some references for kriging have already been given in the Introduction.

Kriging models the unknown $Q_k(\mathbf{x})$ function by a $\xi(\mathbf{x})$ Gaussian random process. The process is assumed to be stationary, with a known covariance function $k(\cdot)$. The observations of $Q_k(\mathbf{x})$ at the locations $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n$ are considered as realizations of the Gaussian random variables $\xi(\mathbf{x}_1), \xi(\mathbf{x}_2), \ldots, \xi(\mathbf{x}_n)$. The method computes the *best linear unbiased prediction* (BLUP) of $\xi(\mathbf{x})$ in the space spanned by these n Gaussian random variables. The prediction is denoted by $\widehat{\xi}(\mathbf{x})$ and expressed as

$$\widehat{\xi}(\mathbf{x}) = \sum_{i=1}^{n} \lambda_i(\mathbf{x}) \xi(\mathbf{x}_i). \quad (18)$$

The coefficients $\lambda_i(\mathbf{x})$ are computed by using the covariance function $k(\cdot)$, via the solution of a linear system of n equations. The *unbiasedness* of the prediction means that the mean of $\xi(\mathbf{x})$ and $\widehat{\xi}(\mathbf{x})$ are equal. The term *best* refers to the fact that the prediction (18) has the smallest error (in the sense of error variance) among all unbiased linear predictors.

Having evaluated $Q_k(\mathbf{x})$ at the locations $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n$ (this is straightforward since $Q_k(\mathbf{x}_l) = ||y_k(t) - y_l(t)||_y$, $k, l = 1, 2, \ldots, n$), the mean of the predictor process (considered as the prediction for $Q_k(\mathbf{x})$) is

$$\widehat{Q}_k(\mathbf{x}) = \sum_{i=1}^{n} \lambda_i(\mathbf{x}) Q_k(\mathbf{x}_i). \quad (19)$$

Let us note that the coefficients $\lambda_i(\mathbf{x})$ do *not* depend directly on the observed values of $Q(\mathbf{x})$ but only on the location of the $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n$ input samples. This fact makes possible to use the *same* $\xi(\mathbf{x})$ process for modeling all $Q_k(\mathbf{x})$ distance functions.

The original optimization problem (16) can then be solved approximately at a much less computational cost by replacing $Q_k(\mathbf{x})$ by $\hat{Q}_k(\mathbf{x})$. The sampling algorithm then becomes tractable even when carrying out an exhaustive search in $\mathbb{X}$ to solve (16).

Further details on this kriging-based sampling algorithm can be found in our recent works [25, 24, 26].

VI Functional kriging interpolation

The output-equidistant sampling methods (Section IV and V) are basically designed for the use of nearest neighbour interpolation based on the final dataset. As it will be presented in Section VIII, such output-equidistant databases might provide additional meta-information on the studied forward problem. However, if the main goal is the interpolation itself, one might consider more sophisticated interpolators than the NN interpolator.

We have successfully applied a certain version of kriging interpolation to construct sample-based surrogate models. The original kriging framework (slightly touched in Section V) has recently been extended to the prediction of functional data [32, 33]. Similarly to the scalar form (19), the functional output $y(t)$, can directly be predicted as

$$\mathcal{F}\{\mathbf{x}\} \approx \mathcal{G}_n\{\mathbf{x}\} = \sum_{i=1}^{n} \lambda_i(\mathbf{x}) y_i(t), \quad (20)$$

where $\mathbf{x}_i$ and $y_i(t)$ are the samples stored in the database. The coefficients $\lambda_i(\mathbf{x})$ are determined in a stochastic framework; a Gaussian random process model (similar to (18)) is applied. However, since the modeled data is functional, the modeling random processes are also functional processes. The so-called trace-covariance has to be introduced and estimated as well. However, these details are out of our scope herein, see [34] for a detailed discussion.

Though the above kriging interpolator usually performs well even when the samples are generated in a naive way (e.g., by sampling regularly on a grid in the input space), the precision can further be improved by the adaptive generation of the database. To this end, we have developed a sequential sampling algorithm aiming at the generation of databases adapted to both the modeled problem and the subsequently used functional kriging interpolator [35]. The procedure starts with a small n number of initial input samples $(\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n)$, chosen according to a naive design (e.g., a regular grid in the input space). Observations of $\mathcal{F}$ are performed at these locations, and a temporary kriging interpolator (20) is fitted. Then, an iterative loop starts, inserting new samples one-by-one according to a certain criterion being related to the estimated prediction error. The latter is obtained by using the so-called jackknifing technique (a well-known tool –being similar to cross-validation– in statistics to estimate the uncertainty of predictors). Jackknifing provides the $\sigma^2_{\mathrm{Jack}}(\mathbf{x})$ estimated variance of the kriging predictor, which can be considered as a measure of the expected interpolation error to a certain extent. Based on this variance, the following heuristic sample-insertion rule can be formalized:

$$\mathbf{x}_{n+1} = \arg\max_{\mathbf{x}\in\mathbb{X}} \left[\min_{i=1,2,\ldots,n} ||\mathbf{x} - \mathbf{x}_i||_x \, \sigma^2_{\mathrm{Jack}}(\mathbf{x}) \right], \qquad (21)$$

where the first factor (the minimal distance in the input space) is involved in order to avoid clustering samples in the input space. By sequentially repeating the above sample-insertion, the yielded database is expected to be a balance of input-space filling and adaptive sampling.

VII Application of adaptive databases in computational electromagnetics

The generated database together with a properly selected interpolator can be used in various applications related to computational electromagnetics. In this section, we would like to highlight some of such possibilities.

Lightweight simulator. In many cases, the end user of a particular design problem is not an expert in using general purpose EM field calculation softwares. At the same time, such design problems usually have only a few number of degree of freedoms. As an example, in the case of the design of a new type of eddy current testing (ECT) probe, the basic geometry of the probe is usually fixed, only certain geometric parameters are enabled to vary between pre-defined bounds. To qualify the performance of such probes, one needs to know the detected response of the probe for only one or two defect prototypes. All the information needed for the design considering the given constrains can be provided to the end-user in the form of a database and an interpolator, embedded in a very simple graphical interface. This design tool is called as "lightweight simulator".

The generation of the database can be done independently by an expert in EM field simulation who can apply the most effective software and model for the calculations. Also, during the database generation one can consider practical constrains of the design, like the noise of the measurements or the manufacturing tolerances of the design parameters of the probe.

Teaching neural networks. An optimal database can be used to teach neural networks. We compared the performance of neural networks taught with databases being equidistant in the input space (this is the database type usually used for teaching in the literature) and in the output space (the one we call output-equidistant database). It was found that the neural networks perform similarly if the input data are free from noise, however, the neural network taught with output-equidistant database outperforms the other one if the observations are subjected to measurement noise [40].

Solution of inverse problems. The most important application of the adaptive databases is, however, in the field of the solution of inverse problems using optimization methods, direct search or interpolation in the database. Such problems related to various different applications have recently been solved. Examples can be read in [16, 39, 24, 45].

Qualification of inverse problems. In addition to the successful solution of an inverse problem, the output-equidistant database can also be used to gather additional information (meta-information) on the inverse problem itself. We believe that this is a very important application, having theoretical importance as well. In the following Section we discuss this point in more details.

VIII Evaluation of the databases and means of inverse problem characterization

In this section, the use of the generated databases is in the focus. The precision provided by the surrogate models is assessed in terms of the interpolation error (Subsection A) and a method for the qualitative comparison of output sample distributions is given (Subsection B). Finally, a simple way of inverse problem characterization (enabling us to draw quantitative conclusions on the possible ill-posedness of the problem) is presented in Subsection C.

The two approaches of output-equidistant sampling (Sections IV and V) yield rather similar databases. A common limitation of both methods is the number of input parameters: the meshing is limited by $p < 9$ (due to the applied *Qhull* mesh generator), whereas in the kriging-based approach, our numerical studies showed that the evaluation of the sample insertion criterion (16) becomes numerically demanding with increasing number of input parameters. One can say that both approaches have the same practical upper limit for p, however, there is no theoretical limitation in any of the two methods. An advantage of the mesh-based methodology is the connectivity information represented by the mesh, enabling higher order interpolation (e.g., linear) over the simplices. On the other hand, the distance functions of the kriging-based strategy make possible the quantitative characterization of the inverse problem in an easy way.

The numerical studies presented herein are all based on the test example defined in Section III. The mesh-supported sampling has already been illustrated in Section IV, hereafter the output-equidistant sampling is performed by the kriging-based approach.

A Precision of the forward interpolation

The most plausible way to compare the different surrogate models is to compute the discrepancy between the true outputs (by evaluating $\mathcal{F}$) and the output of the surrogate model, i.e., to consider the interpolation error

$$\varepsilon(\mathbf{x}) = ||\mathcal{G}_n\{\mathbf{x}\} - \mathcal{F}\{\mathbf{x}\}||_y \,. \tag{22}$$

In Fig. 4, four surrogate models are compared in the sense of $\varepsilon(\mathbf{x})$. The true output –also $\varepsilon(\mathbf{x})$– is computed over a 37×18 grid in the input space. The worst performance is shown by the combination of the naive grid sampling and the NN interpolation (Fig. 4a). NN interpolation yields smaller error when using an output-equidistant database (Fig. 4b). The smooth functional kriging interpolators provide much finer precision even by grid-sampling (Fig. 4c), whereas the adaptive sampling strategy of functional kriging results in outstanding precision (Fig. 4d).

The distribution of the input samples in the output-equidistant database (Fig. 4b) has an important message concerning the modeled forward problem. In the large-crack region, much more samples have been inserted than in the region of small (mainly in the sense of the depth D) cracks. Since the output sampling is output-equidistant, one can conclude that $\mathcal{F}$ varies rapidly in the large-crack region and has almost no variation in the small-crack region. This *meta-information* carried by the structure of the database will be further discussed and exploited in Subsection C.

B Visualization of the output sample distribution

Since the output data lie in a function space, direct visualization is not possible, but the *multidimensional scaling* (MDS) method [46] can be applied instead. The method gives a low-dimensional representation of the output samples whereas the distance relations remain more-or-less unchanged. Let us consider a database including n samples. The main idea of MDS is to space n points $(\boldsymbol{\pi}_1, \boldsymbol{\pi}_n, \ldots, \boldsymbol{\pi}_n)$ in a k-dimensional Euclidean space so that the pairwise distances between the new points are as similar as possible to the distances between the original output samples. Formally, the so-called stress

$$S = \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} \Big(||y_i(t) - y_j(t)||_y - ||\boldsymbol{\pi}_i - \boldsymbol{\pi}_j|| \Big)^2 \tag{23}$$

is to be minimized by the appropriate spacing of the points $\boldsymbol{\pi}_i$ (herein $||\cdot||$ stands for the Euclidean norm). Specially, if $k = 2$, the resulted repartition can easily be plotted. This plot can then be considered as the low-dimensional representation of the original dataset.

In Fig. 5, the MDS plots of the databases in the 2 and 4-parameter cases are presented. The output-equidistant sampling yields balanced output sample distribution whereas the regular input sampling causes distorted output sampling. In the latter case, some samples appears to cluster in certain regions of the output space; other regions are unexplored at the same time. According to the visualized repartition of the output samples, the proposed sampling algorithm indeed outperforms the naive grid sampling approach.

C Ill-posedness and inverse-mapped noise levels

As it is well known, inverse problems (determining the configuration knowing the measured EM field) might be ill-posed: the solution might not exist, not be unique and not be stable. The latter refers to the fact that small variations in the measured data can cause high variations in the reconstructed configuration. This sensitivity (being a basic property of $\mathcal{F}$) can be assessed to a certain extent by using the output-equidistant databases. In Subsection A, we have already pointed out the relation between the input sample density and the assumed behaviour of the forward operator.

Based on the distance functions (15), the ill-posedness of the inverse problem can quantitatively be addressed. Let us assume that a uniform noise level δ –being an aggregated representation of all sources of uncertainty– affects the measurement at hand. For a given measured output $\tilde{y}(t)$, one then has to consider all functions $y(t)$, being closer to $\tilde{y}(t)$ than a distance δ, as possible real (noise-free) data. Expressively, one has to imagine a sphere with radius δ and center $\tilde{y}(t)$ in the $L^2(T)$ function space. The uncertainty of the solution of the inverse problem is in fact related to the *image* of this sphere in the input space. Via the distance functions, such inverse imaging boils down to a simple inequality:

$$\Phi_i^\delta = \Big\{ \mathbf{x} \in \mathbb{X} : ||\mathcal{F}\{\mathbf{x}\} - y_i(t)||_y \leq \delta \Big\} = \\ = \{ \mathbf{x} \in \mathbb{X} : Q_i(\mathbf{x}) \leq \delta \} \,, \tag{24}$$

where Φ_i^δ is the "inverse-mapped" noise region, if the measured data $\tilde{y}(t)$ is assumed to be equal to the ith sample in the database.

The shape and the dimensions of the different Φ_i^δ cells provide valuable information about the underlying problem, since they show how the noise influences the uncertainty of the solution of the inverse problem. Where $\mathcal{F}$ varies rapidly, a smaller Φ_i^δ is expected than in the regions where $\mathcal{F}$ is flat, i.e., when the problem tends to be ill-posed. This formalization gives an explicit expression of the within reach precision at a given noise level.

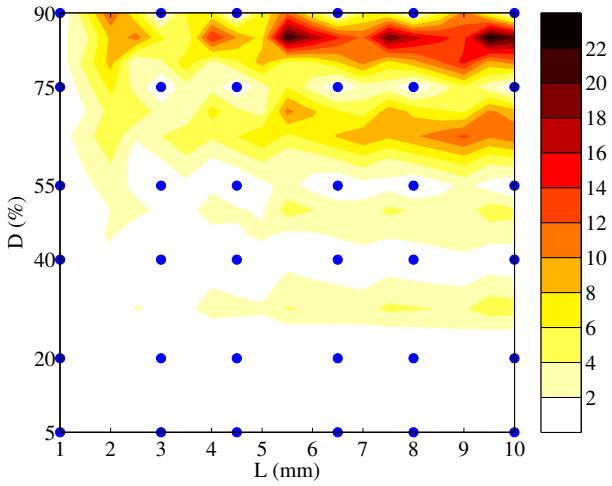
Let us also mention that if the noise level δ is known when generating the database, the latter can be optimized by using $\Delta = 2\delta$ in (10).

In the realization, due to the high computational cost, not the exact distance functions but their kriging predictions are used. Thus, the approximate noise-cells $\hat{\Phi}_i^\delta$ can be determined as

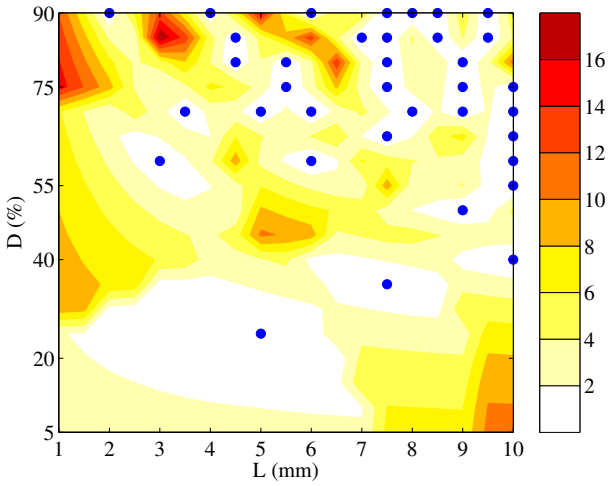
$$\hat{\Phi}_i^\delta = \Big\{ \mathbf{x} \in \mathbb{X} : \hat{Q}_i(\mathbf{x}) \leq \delta \Big\} \,. \tag{25}$$

In Fig. 6, the approximate noise-cells in the 2-parameter example for three different δ levels are shown. Let us notice that the database is approximately output-equidistant, even so the shape and the dimensions of the different cells $\hat{\Psi}_i$ in the input space are very diverse. The large cells in the region of small cracks highlight that (i) a finer sampling is not needed, and (ii) the difficulty of retrieving such small defects. On the contrary, the small cells in the region of large cracks justify the dense sampling. These results are in good agreement with the experience and the explanation of the evaluation of the forward interpolation error (Subsection A).

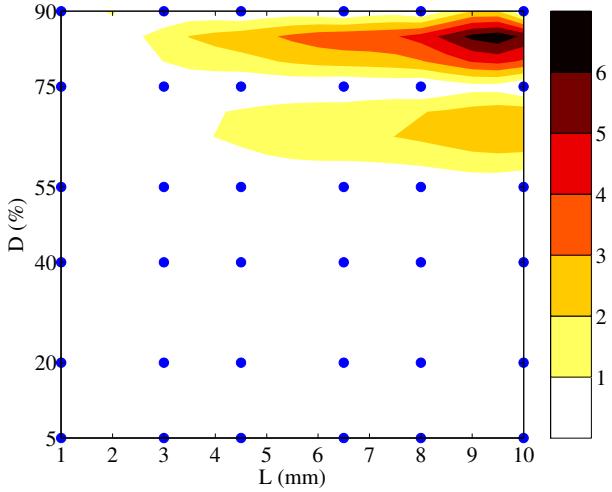
Further discussion on the inverse mapping and more numerical examples can be found in [24].



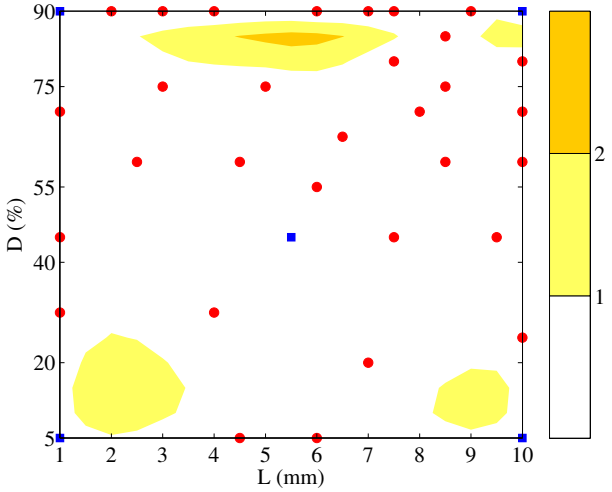
(a) Regular sampling + NN interpolation.



(b) Output equidistant sampling + NN interpolation.



(c) Regular sampling + functional kriging interpolation.



(d) Adaptive sampling + functional kriging interpolation. (Blue squares: initial samples; red dots: sequentially added samples.)

Figure 4: Interpolation error ($\varepsilon(\mathbf{x})$ in mΩ) using different sampling and interpolation methods. All databases include 36 samples.

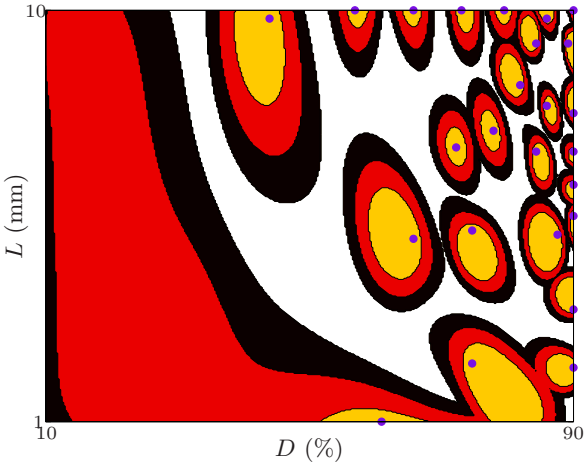


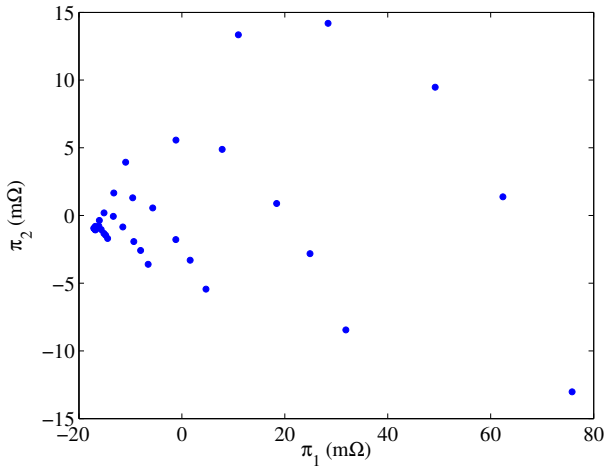
Figure 6: Inverse-mapped noise levels and input samples for three different noise levels in the 2-parameter example, using $n = 25$ samples. (Taken form [24].)

IX Conclusion

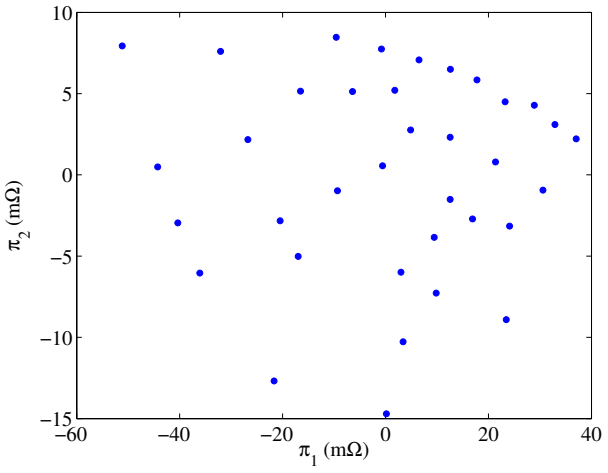
Different sample-based adaptive surrogate modeling approaches have been presented in the paper. The output-equidistant databases are combined with the nearest neighbour interpolation. The use of functional kriging as interpolator has also been presented, along with a sampling algorithm which improves the precision of the surrogate model being built.

For the generation of output-equidistant databases, two methods are proposed. First a p-dimensional simplex mesh has been used as a generic interpolator of the forward problem. Adaptive mesh refinement and smoothing techniques are applied for exploring the forward operator and for learning the topological and metric relationships of its data samples. By introducing the control space, where meshing takes place, one can avoid the usage of special mesh generators. In the second approach, the generation of the database is recast as an optimization problem, which is solved iteratively, by inserting the next input sample with respect to the distance between points in the output space. This is an expensive-to-compute optimization problem, thus, the use of a kriging interpolator is proposed. In spite of the essential difference of the working principles, the two methodologies yield similar final databases.

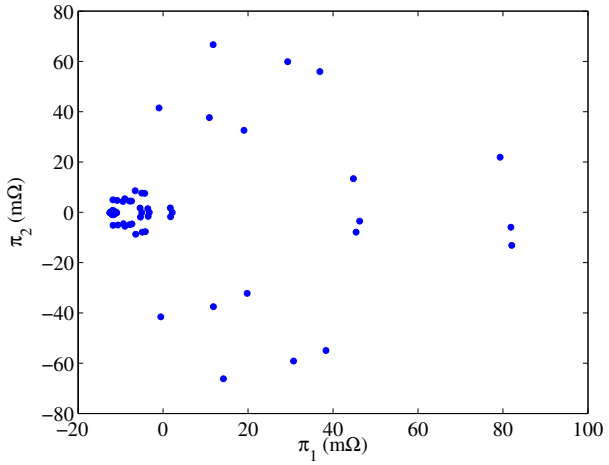
In the case of functional kriging interpolation (instead of nearest neighbour), the database generation involves the



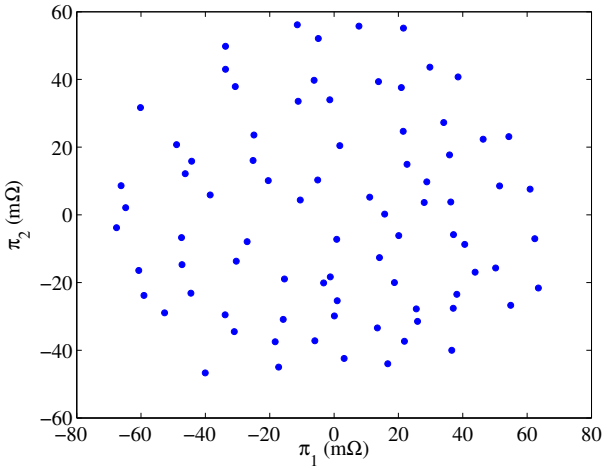
(a) 2-parameter example, regular sampling with 36 samples (see Fig. 4a for the corresponding input sampling).



(b) 2-parameter example, output-equidistant sampling with 36 samples (see Fig. 4b for the corresponding input sampling).



(c) 4-parameter example, regular sampling with 81 samples (input samples are spaced on a regular grid of $3 \times 3 \times 3 \times 3$ nodes).



(d) 4-parameter example, output-equidistant sampling with 81 samples.

Figure 5: *Multidimensional scaling* representation of output samples in the 2 and 4-parameter cases, with 36 samples. The output-equidistant databases have been generated by the kriging method discussed in Section V.

estimated interpolation error via the jackknife variance estimation. The variance of the prediction error is a natural measure of the quality of the interpolation in the stochastic framework of kriging. The resulted sampling strategy is thus fully adapted to both the modeled forward operator and the applied interpolator.

The efficiency of all approaches is illustrated by an eddy-current nondestructive application using two, three and four defect parameters.

The surrogate models involving functional kriging interpolation appear to outperform the strategies based on output-equidistant databases in the sense of interpolation precision. However, the special structure of the output-equidistant databases carries some meta-information on the modeled forward problem, making possible the characterization of the related inverse problem to a certain extent. An inverse mapping of a noise level is discussed, which results in a quantitative assesment of the uncertainty of the reconstructed input parameters in the presence of measurement noise.

As a common limitation of the presented approaches, we must note the relatively small p number of considered input parameters. Though there is no theoretical limitation, the increasing computational load arises a limit of, say, $p < 9$. However, we believe that further development can raise this upper limit.

The future research will be focused on (i) the generation of databases optimized for inverse interpolation, and (ii) on the analysis of uncertainty propagation (from different sources, e.g., imprecise knowledge on the material parameters, simulation errors, measurement noise) involving the kriging-based tools of database generation.

Acknowledgement

Authors would like to acknowledge the significant contribution of Marc Lambert, Dominique Lesselier (Laboratoire des Signaux et Systèmes, CNRS-SUPELEC-Univ Paris-Sud, France), Yann Le Bihan (Laboratoire de Génie Electrique de Paris, France) and Emmanuel Vazquez (SUPELEC, France) to the results of the research made in cooperation in the field covered by this paper.

X References

[1] J. K. Sykulski, "New trends in optimization in electromagnetics," *Przeglad Elektrotechniczny*, vol. 83, no. 6, pp. 13–18, 2007.

[2] T. J. Santner, B. J. Williams, and W. I. Notz, *The Design and Analysis of Computer Experiments*. Springer, 2003.

[3] K.-T. Fang, R. Li, and A. Sudjianto, *Design and Modeling for Computer Experiments*. London: Chapman & Hall/CRC, 2006.

[4] T. P. Ryan, *Modern Experimental Design*. Wiley, 2007.

[5] M. D. McKay, R. J. Beckman, and W. J. Conover, "A comparison of three methods for selecting values of input variables in the analysis of output from a computer code," *Technometrics*, vol. 42, pp. 55–61, 1979.

[6] J. Sacks, W. Welch, T. Mitchell, and H. Wynn, "Design and analysis of computer experiments," *Statist. Science*, vol. 4, pp. 409–435, 1989.

[7] A. A. Giunta, S. F. Wojtkiewicz Jr., and M. S. Eldred, "Overview of modern design of experiments methods for computational simulations," in *Proc. of the 41th AIAA Aerospace Sciences Meeting And Exhibit*, Reno, Jan 6-9 2003.

[8] L. Pronzato, "Optimal experimental design and some related control problems," *Automatica*, vol. 44, no. 2, pp. 303 – 325, 2008.

[9] D. J. Brus and G. B. Heuvelink, "Optimization of sample patterns for universal kriging of environmental variables," *Geoderma*, vol. 138, no. 1-2, pp. 86 –95, 2007.

[10] F. A. C. Viana, C. Gogu, and R. T. Haftka, "Making the most out of surrogate models: tricks of the trade," in *Proceedings of the ASME 2010 International Design Engineering Technical Conferences & Computers and Information in Engineering Conference IDETC/CIE 2010*, Montreal, Canada, 8 2010, p. 28813.

[11] J. R. Shewchuk, "What is a good linear element? Interpolation, conditioning, and quality measures," in *11th International Meshing Roundtable*, 2002, pp. 115–126.

[12] L. Chen, P. Sun, and J. Xu, "Optimal anisotropic meshes for minimizing interpolation errors in L^p-norm," *Mathematics of Computation*, vol. 76, pp. 179–204, 2007.

[13] H. Borouchaki, P. J. Frey, and P. L. George, "Adaptive triangular-quadrilateral mesh generation," *International Journal of Numerical Methods in Engineering*, vol. 41, pp. 915–934, 1998.

[14] F. Alauzet, A. Loseille, A. Dervieux, and P. Frey, "Multi-dimensional continuous metric for mesh adaptation," in *15th International Meshing Roundtable*, Birmingham, Alabama, September 2006, pp. 191–214.

[15] H. Weimer, J. D. Warren, J. Troutner, W. Wiggins, and J. Shrout, "Efficient co-triangulation of large data sets," in *IEEE Visualization*, 1998, pp. 119–126.

[16] J. Pávó and S. Gyimóthy, "Adaptive inversion database for electromagnetic nondestructive evaluation," *NDT&E International*, vol. 40, no. 3, pp. 192–202, 2007.

[17] G. Greiner and K. Hormann, "Interpolating and approximating scattered 3D data with hierarchical tensor product B-splines," in *Surface Fitting and Multiresolution Methods*. Vanderbilt University Press, 1997, pp. 163–172.

[18] K. Hormann, B. Lévy, and A. Sheffer, "Mesh parameterization: Theory and practice," in *SIGGRAPH 2007 Course Notes*, vol. 2. ACM Press, Aug. 2007, pp. 1–122.

[19] F. Bossen and P. Heckbert, "A pliant method for anisotropic mesh generation," in *5th International Meshing Roundtable*, October 1996.

[20] K. Shimada, A. Yamada, and T. Itoh, "Anisotropic triangular meshing of parametric surfaces via close packing of ellipsoidal bubbles," in *6th International Meshing Roundtable*, 1997, pp. 375–390.

[21] G. K. Knopf and A. Sangole, "Interpolating scattered data using 2D self-organizing feature maps," *Graphical Models*, vol. 66, no. 1, pp. 50–69, 2004.

[22] M. Svensén, "Gtm: The generative topographic mapping," Ph.D. dissertation, Aston University, Birmingham, 1998.

[23] A. Gorban and A. Zinovyev, "Elastic maps and nets for approximating principal manifolds and their application to microarray data visualization," in *Principal Manifolds for Data Visualisation and Dimension Reduction*, A. Gorban, B. Kégl, D. Wunsch, and A. Zinovyev, Eds. Springer, 2007, pp. 97–131.

[24] S. Bilicz, M. Lambert, and S. Gyimóthy, "Kriging-based generation of optimal databases as forward and inverse surrogate models," *Inverse Problems*, vol. 26, no. 7, p. 074012, 2010.

[25] S. Bilicz, M. Lambert, E. Vazquez, and S. Gyimóthy, "Combination of maximin and kriging prediction methods for eddy-current testing database generation," *Journal of Physics: Conference Series*, vol. 255, no. 1, p. 012003, 2010.

[26] ——, "A new database generation method combining maximin method and kriging prediction for eddy-current testing," in *Electromagnetic Nondestructive Evaluation (XIII)*, ser. Studies in Applied Electromagnetics and Mechanics, J. Knopp, M. Blodgett, B. Wincheski, and N. Bowler, Eds. Amsterdam: IOS Press, 2010, vol. 33, pp. 199–206.

[27] N. A. C. Cressie, *Statistics for Spatial Data*. Wiley, 1993.

[28] J. Chilès and P. Delfiner, *Geostatistics, Modeling Spatial Uncertainty*. Wiley, 1999.

[29] M. Stein, *Interpolation of Spatial Data: Some Theory for Kriging*. New York: Springer, 1999.

[30] J. P. Kleijnen, "Kriging metamodeling in simulation: A review," *European Journal of Operational Research*, vol. 192, no. 3, pp. 707–716, February 2009.

[31] S. Bilicz, E. Vazquez, M. Lambert, S. Gyimóthy, and J. Pávó, "Characterization of a 3D defect using the expected improvement algorithm," *COMPEL*, vol. 28, no. 4, pp. 851–864, 2009.

[32] R. Giraldo, P. Delicado, and J. Mateu, "Geostatistics for functional data: An ordinary kriging approach," Universitat Politècnica de Catalunya, Tech. Rep., 2007.

[33] P. Delicado, R. Giraldo, C. Comas, and J. Mateu, "Statistics for spatial functional data: some recent contributions," *Environmetrics*, 2009.

[34] S. Bilicz, E. Vazquez, S. Gyimóthy, J. Pávó, and M. Lambert, "Kriging for eddy-current testing problems," *IEEE Transactions on Magnetics*, vol. 46, no. 8, pp. 3165–3168, 2010.

[35] S. Bilicz, S. Gyimóthy, M. Lambert, and J. Pávó, "Adaptive kriging metamodels for expensive-to-run electromagnetic simulations," in *Proceedings of the 14th International IGTE Symposium on Numerical Field Calculation in Electrical Engineering*, Graz, Austria, 9 2010, p. 48.

[36] J. Kleijnen and W. van Beers, "Application-driven sequential designs for simulation experiments: Kriging metamodelling," *J. Oper. Res. Soc.*, vol. 55, pp. 876–883, 2004.

[37] R. Jin, W. Chen, and A. Sudjianto, "On sequential sampling for global metamodeling in engineering design," in *Proceedings of DETC'02 ASME 2002 Design Engineering Technical Conferences and Computers and Information in Engineering Conference*, Montreal, Canada, 2002.

[38] S. Gyimóthy and J. Pávó, "Qualification of the inverse problem of defect reconstruction using optimized mesh database," *COMPEL*, vol. 24, no. 2, pp. 436–445, 2005.

[39] S. Gyimóthy, J. Pávó, and H. Tsuboi, "Conceptual evaluation of inversion models used for layered structures," *IEEE Transactions on Magnetics*, vol. 42, no. 4, pp. 1091–1094, 2006.

[40] S. Gyimóthy, Y. L. Bihan, and J. Pávó, "Optimized database for training neural networks used in nondestructive testing," *International Journal of Applied Electromagnetics and Mechanics*, vol. 25, no. 1-4, pp. 717–721, 2007.

[41] S. Gyimóthy, I. Kiss, and J. Pávó, "Adaptive sampling technique based on moving meshes for building data-equidistant inversion databases for NDT," *International Journal of Applied Electromagnetics and Mechanics*, vol. 30, no. 3-4, pp. 309–319, 2009.

[42] J. Pávó and D. Lesselier, "Calculation of eddy current testing probe signal with global approximation," *IEEE Transactions on Magnetics*, vol. 42, pp. 1419–1422, April 2006.

[43] F. J. Blom, "Considerations on the spring analogy," *International Journal for Numerical Methods in Fluids*, vol. 32, pp. 647–668, 2000.

[44] Qhull, a public domain generic n-dimensional Delaunay-mesh generation code. Available at http://www.qhull.org.

[45] H. Acikgoz, L. Santandrea, Y. Le Bihan, S. Gyimóthy, J. Pávó, O. Meyer, and L. Pichon, "Generation and use of optimized databases in microwave characterization," *IEE Proceedings-Science Measurement and Technology*, vol. 2, no. 6, pp. 457–473, 2008.

[46] J. B. Kruskal and M. Wish, *Multidimensional Scaling*. SAGE, 1986.

Authors name and affiliation

József Pávó, Szabolcs Gyimóthy, Sándor Bilicz, Imre Kiss: Electromagnetic Simulation and Design Laboratory, Electromagnetic Theory Group, Department of Broadband Infocommunications and Electromagnetic Theory, Faculty of Electrical Engineering and Informatics, Budapest University of Technology and Economics. H-1521 Budapest, e-mail: pavo@evt.bme.hu