Postprint of: Kozieł S., Pietrenko-Dąbrowska A., Knowledge-based performance-driven modeling of antenna structures, KNOWLEDGE-BASED SYSTEMS (2021), 107698, DOI: 10.1016/j.knosys.2021.107698

Knowledge-Based Performance-Driven Modeling of Antenna Structures

Slawomir Koziel[1,2] and Anna Pietrenko-Dabrowska[2]

[1] Engineering Optimization & Modeling Center, Reykjavik University, 101 Reykjavik, Iceland, koziel@ru.is,

[2] Faculty of Electronics, Telecommunications and Informatics, Gdansk University of Technology, 80-233 Gdansk, Poland, anna.dabrowska@pg.edu.pl

Keywords: Antenna modeling; surrogate modeling; domain confinement; performance-driven modeling; EM-based design; inverse surrogates.

Abstract

The importance of surrogate modeling techniques in the design of modern antenna systems has been continuously growing over the recent years. This phenomenon is a matter of practical necessity rather than simply a fashion. On the one hand, antenna design procedures rely on full-wave electromagnetic (EM) simulation tools. On the other hand, the computational costs incurred by repetitive EM analyses involved in solving common tasks (parameter tuning, uncertainty quantification, multi-criterial design, etc.), are often prohibitive. The replacement of full-wave simulations by fast surrogates may mitigate these issues; as a matter of fact, it is the only viable option for carrying out EM-driven design in many cases. Among available modeling approaches, data-driven surrogates are by far the most popular due to their accessibility and versatility. At the same time, a construction of reliable models is hindered by the curse of dimensionality, high nonlinearity of antenna characteristics, as well as broad ranges of parameters and operating conditions that the model has to cover to ensure its design utility. Recently proposed performance-driven modeling frameworks offer a workaround these issues by restricting the model domain to the parameter space regions that contain high-quality designs (w.r.t. the assumed performance metrics). However, the domain determination requires acquisition of a set of pre-optimized reference designs, which adds to the overall computational cost of the surrogate model setup in a significant manner. This work proposes a novel two-stage knowledge-based approach, where the confined domain is defined without using any reference designs. Instead, a preselected set of random observables is employed to establish an inverse regression model being a basis for domain determination of the final surrogate. Comprehensive numerical validation involving three antenna structures indicates that our methodology offers the computational benefits similar to those of the previous performance-driven methods while considerably reducing the initial setup cost, by a factor of sixty percent on the average, which has been achieved by exploiting the problem-specific knowledge.

1. Introduction

Although theoretical models and methods are still applicable in the design of antenna structures, a continuous shift towards simulation-based techniques has been observed over the recent years [1]-[4]. In particular, full-wave electromagnetic (EM) analysis, once mainly employed to design verification, is nowadays used at all design stages, including topology development [5], parametric studies [6], and, first and foremost, final tuning of geometry parameters [7], [8]. The primary reasons for the aforementioned shift include inadequacy of analytical (e.g., circuit-theoretical) antenna representations in accounting for, e.g., the mutual coupling, the effects of connectors, housing, installation fixtures, or substrate anisotropy. At the same time, antenna geometries have become increasingly involved to meet the demands of contemporary applications, and enable multi-band [9], broadband [10], or MIMO operation [11], circular polarization [12], or polarization/pattern diversity [13]. Realization of these capabilities is often achieved using dedicated components (stubs [14], slots [15], shorting pins [16], defected ground structures [17]). Evaluating the effects thereof can only be achieved through full-wave EM simulation. While ensuring accuracy, EM analysis tends to be computationally expensive, to the extent of being prohibitive when multiple simulations are required, e.g., when solving tasks such as parametric optimization [18] or uncertainty quantification (statistical analysis [19], tolerance-aware design [20]). This might be a problem even for local tuning of relatively simple components, let alone global optimization [21] or multi-criterial design [22] of complex devices (frequency selective surfaces [23], substrate integrated waveguide (SIW)-based components [24], multi-layer antennas [25], etc.).

Considerable research efforts have been directed towards expediting EM-driven design procedures, especially in the context of parametric optimization (both single- and multi-objective) but also uncertainty quantification [26], [27]. Some of available options include acceleration of gradient-based procedures using adjoint sensitivities [28], mesh deformation

methods [29], or sparse Jacobian updates [30], [31], as well as utilization of dedicated solvers [32]. Surrogate-assisted frameworks constitute a quickly growing class of techniques developed as computationally efficient alternatives to conventional methods. Physics-based metamodels are particularly popular for local optimization. Herein, the surrogate is constructed through a suitable correction of the underlying low-fidelity model. Exemplary approaches include space mapping [33], response correction (manifold mapping [34], adaptive response scaling [35], shape-preserving response prediction [36]). Data-driven (or approximation) models (kriging [37], radial-basis functions [38], neural networks [39], Gaussian process regression [40], support vector machines [41]) find applications in accelerating global search (including multi-criterial design) [42], [43], often in combination with sequential sampling procedures [44]. Polynomial chaos expansion (PCE) surrogates are well suited for handling uncertainty quantification tasks because statistical moments of the system outputs (given probability distributions of the inputs) can be determined directly from the model coefficients, without defaulting to Monte Carlo simulation [45]. Yet another approach is exploitation of the specific structure of the system response, e.g., allocation of multi-band antenna resonances [46] (feature-based optimization (FBO) [47], [48] or cognition-driven design [49]). The former allows for reducing the cost of the optimization process by reformulating the design task in terms of appropriately defined characteristic points of the responses, the coordinates of which are in slightly nonlinear dependence on the geometry and/or material parameters of the device at hand.

From the design perspective, a replacement of expensive full-wave simulations by a stand-alone surrogate would be the most economical option as it enables conducting all kinds of simulation-driven procedures, primarily optimization, at a low computational cost. In this context, the most suitable choices are approximation techniques. Numerous methods, such as kriging [50], neural networks (in many variations, e.g., deep neural networks [51],

convolutional neural networks [52]), radial basis functions [53], PCE [54], support vector regression (SVR) [55] are easily accessible using various toolboxes (e.g., [56], [57]). Approximation models are versatile, transferrable between the problem domains, and straightforward to handle. Notwithstanding, data-driven modeling of antenna structures is hindered by both the curse of dimensionality, and a typically high nonlinearity of antenna characteristics. Furthermore, in order for the surrogate to be useful for design purposes, it needs to cover broad ranges of geometry parameters, which makes the modeling task even more challenging. Many of the mitigation methods, such as high-dimensional model representation (HDMR) [58], or least-angle regression [59], are of limited use in the case of antenna structures. On the other hand, variable-fidelity techniques have been shown to work well for antenna modeling (co-kriging [60], two-stage GPR [61]).

A conceptually distinct, alternative approach to alleviating dimensionality and parameter range problems has been proposed in the form of performance-driven modeling methods [62]. The principal idea thereof is to restrict the domain of the surrogate to the vicinity of the region containing design that are optimum (or nearly-optimum) with respect to the assumed figures of interest or operating/implementation condition (allocation of operating frequencies, substrate permittivity, power split ratio in the case of coupling structures, etc.). Setting up the model within a confined domain requires a significantly smaller number of training samples and allows us to maintain good predictive power over broad ranges of the parameters [63]. Several variations of performance-driven frameworks have been proposed [64], [65], with a notable example of nested kriging [64], further generalized to variable-fidelity setup [66], as well as combined with dimensionality reduction methods [67]. The principal issue of the aforementioned techniques is that identification of the domain involves a set of pre-optimized reference designs, acquisition of which adds to the overall cost of the surrogate setup. With a typical number of reference designs ranging from ten to twenty (for two- or three-

dimensional objective space [64]), this cost can be as high as over a thousand EM evaluations of the antenna structure of interest. Utilization of advanced techniques such as gradient-enhanced kriging (GEK) may reduce these expenses somehow [68] but the underlying problem remains.

This paper proposes a novel knowledge-based approach to surrogate modeling of antenna structures. Our methodology capitalizes on the performance-driven modeling concept; however, it does not require any pre-optimized reference designs. The domain of the final surrogate is established using an inverse regression model constructed from a preselected set of random observables. The inverse model maps the objective space of the antenna at hand into its parameter space, and it is identified based on the knowledge about the system at hand in the form of performance figure values extracted from the EM simulated antenna responses across the observable set. To form the domain, the image of the inverse model is then orthogonally extended towards its normal vectors, with the amount of extension deduced from the allocation of the observables in the parameter space, which is yet another way of exploiting the system-specific knowledge by the proposed modeling technique to make it more reliable and efficient. Using three antenna examples it is shown that the presented procedure offers the computational benefits comparable or even exceeding those pertinent to the prior performance-driven frameworks (including nested kriging). At the same time, the initial setup cost is considerably reduced, by over sixty percent on the average. Design utility of the framework is demonstrated through applications to parametric optimization of the considered antenna structures.

2. Two-Stage Knowledge-Based Modeling with Inverse Regression

This section introduces the proposed two-stage modeling procedure. We start by recalling the concept of performance-driven modeling (Section 2.1). Section 2.2 describes the inverse regression surrogate (stage one) constructed from the set of pre-selected random

observables, and used to define the domain of the final surrogate (stage two). Definition of the surrogate model domain, design of experiments scheme, as well as final model construction are covered in Section 2.3. Section 2.4 summarizes the operating flow of the entire modeling framework. Numerical verification of the procedure and benchmarking will be discussed in Section 3.

2.1. Performance-Driven Modeling Concept

The modeling procedure proposed in this paper capitalizes on the performance-driven concept (also referred to as constrained modeling) [65], which is briefly recalled below. The principal idea is to restrict the surrogate domain to those regions of the parameter space that contain high-quality designs (with respect to the assumed performance figures). The volume of such a region is normally small because the antenna geometry parameter vectors corresponding to optimum designs are well correlated (e.g., re-designing the antenna for different operating frequencies requires synchronized adjustment of at least some of its parameters [69]). Reducing the domain size as compared to the conventional approach, normally being an interval determined by the lower/upper bounds of antenna parameters, results in a considerable reduction of the number of necessary training points and ensuring good predictive power of the surrogate even in higher-dimensional cases and for broad ranges of the parameters [63], [64].

Table 1. Performance-driven modeling: Notation

Description	Notation
Antenna parameter vector	$\boldsymbol{x} = [x_1 \ldots x_n]^T$
Conventional parameter space	$X = [\boldsymbol{l}, \boldsymbol{u}]$
Lower bounds on the parameters	$\boldsymbol{l} = [l_1 \ldots, l_n]^T$
Upper bounds on the parameters	$\boldsymbol{u} = [u_1 \ldots, u_n]^T$
Figures of interest	f_k, $k = 1, \ldots, N$
Objective space	F: $f_{k.\min} \le f_k^{(j)} \le f_{k.\max}$, $k = 1, \ldots, N$
Objective vector	$\boldsymbol{f} = [f_1 \ldots f_N]^T$

The standard notation used in the context of performance-driven modeling follows [64] and was highlighted in Table 1. The figures of interest include, e.g., operating frequency, bandwidth, substrate permittivity. The ranges of the figures of interest delimit the objective space F, which constitutes the intended region of validity of the surrogate model.

The critical component of performance-driven modeling is a definition of the surrogate model domain. This requires the notion of design optimality, which is understood in terms of minimizing the scalar merit function $U(\boldsymbol{x},\boldsymbol{f})$ that quantifies the design quality [65]. More specifically, for a given $\boldsymbol{f} \in F$, the optimum design $\boldsymbol{x}^*$ is obtained by solving

$$\boldsymbol{x}^* = U_F(\boldsymbol{f}) = \arg\min_{\boldsymbol{x}} U(\boldsymbol{x}, \boldsymbol{f}) \tag{1}$$

The set of designs that are optimum for all $\boldsymbol{f} \in F$ is denoted by

$$U_F(F) = \{U_F(\boldsymbol{f}) : \boldsymbol{f} \in F\} \tag{2}$$

Note that—assuming the uniqueness of solving (1)—$U_F(F)$ is an N-dimensional manifold embedded in the parameter space X.

The surrogate model domain is constructed to contain a possibly small vicinity of $U_F(F)$ [65]. This requires a relatively precise identification of the optimum designs, which is normally achieved using a set of reference designs $\boldsymbol{x}^{(j)} = [x_1^{(j)} \ldots x_n^{(j)}]^T$ $j = 1, \ldots, p$, pre-optimized with respect to the objective vectors $\boldsymbol{f}^{(j)} = [f_1^{(j)} \ldots f_N^{(j)}]$, allocated within the objective space. Having the dataset $\{\boldsymbol{f}^{(j)},\boldsymbol{x}^{(j)}\}$, $j = 1, \ldots, p$, a first-level surrogate $\boldsymbol{s}_I(\boldsymbol{f}) : F \rightarrow X$ is identified [64] to serve as an initial approximation of $U_F(F)$. A graphical illustration of the aforementioned concepts can be found in Fig. 1.

The necessity of acquiring the reference designs is the most serious practical issue related to performance-driven modeling methods. This is for two primary reasons:

- Depending on the objective space dimensionality, the number of reference designs varies between ten and twenty [63], [64]. Assuming that each design is obtained through

local (e.g., gradient-based) search, and the average parameter space dimensionality is ten, the cost of reference point acquisition may be as high as a thousand EM analyses of the antenna at hand. This cost adds to the surrogate model setup cost. Needless to say, it is detrimental to the computational efficiency of the modeling process.

- Acquisition of the reference vectors is a numerically challenging task because the antenna structure of interest has to be re-designed over broad ranges of operating conditions. In practice, it may entail several optimization attempts to obtain a single design, including manual adjustment of the starting point. Automation of this process is difficult [70].

As mentioned in the introduction, it is possible to reduce the number of reference points using advanced techniques such as the establishment of the first-level surrogate using gradient-enhanced kriging [68], or attempt to automate the acquisition process [70]. Nevertheless, the main problem, i.e., the necessity of gathering at least several designs allocated within the optimum design manifold, remains.

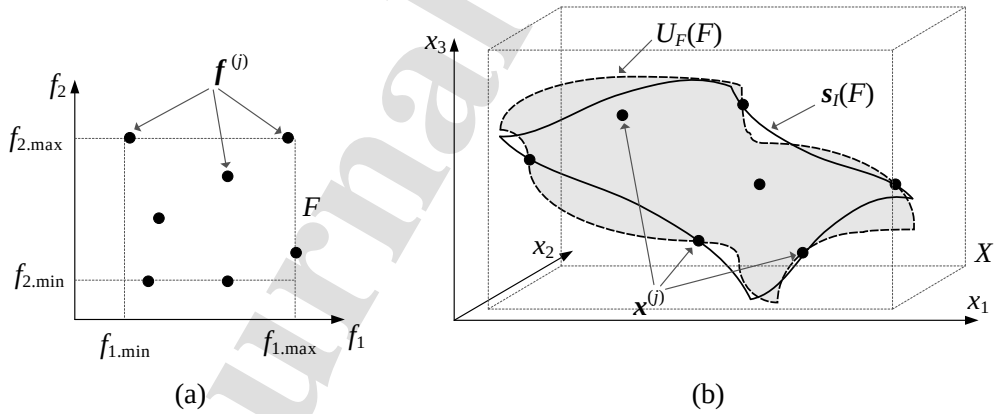


Fig. 1. Defining concepts of performance-driven modeling [64]: (a) the objective space F, (b) the parameter space X. The reference designs are shown as black circles, whereas the optimum design manifold $U_F(F)$ is indicated as a grey surface. The first-level model image $\boldsymbol{s}_I(F)$ provides a first approximation of the manifold, and has to be further extended to encapsulate the entire $U_F(F)$ (or, at least, the majority of it).

One of the main purposes of this work is to develop a modeling approach that benefits from the overall concept of domain confinement; however, it does not rely on reference designs produced through separate optimization procedures.

2.2. Two-Stage Modeling: Stage One – Knowledge-Based Inverse Regression Surrogate

The reference designs, as elaborated on in Section 2.1, provide a great deal of information about the spatial allocation of the optimum design manifold, which subsequently becomes the core of the surrogate model definition [65]. In this work, the reference designs are replaced by a pre-selected set of random observables generated in the parameter space X. Using the knowledge extracted from the observables, an inverse regression model is constructed that replaces the first-level surrogate discussed before. The inverse model will be then used to define the domain of the final surrogate, cf. Section 2.3.

Let $\boldsymbol{x}_r^{(j)}$, $j = 1, 2, \ldots$, be a sequence of random parameter vectors sequentially generated in X using a uniform probability distribution. Let $\boldsymbol{f}_r^{(j)}$ be the performance figure vector extracted from the EM-simulated antenna response at $\boldsymbol{x}_r^{(j)}$. For example, in the case of a dual-band antenna with the figures of interest being the target operating frequencies, the vector $\boldsymbol{f}_r^{(j)}$ contains the actual operating frequencies of the structure (regardless of the resonance depths). If $\boldsymbol{f}_r^{(j)} \in F$, the observable is accepted; otherwise (also, if the components of $\boldsymbol{f}_r^{(j)}$ fall beyond the assumed lower or upper limits for the figures of interest, or cannot be identified at all), it is rejected. Figure 2 shows a graphical illustration for the discussed case of a dual-band antenna. The sampling process is concluded when the prescribed number of observables has been found, here denoted as N_r (typically 50 to 100).

The dataset $\{\boldsymbol{x}_r^{(j)}, \boldsymbol{f}_r^{(j)}\}_{j=1,\ldots,Nr}$, is used to establish the inverse regression model s_r, which provides a rough allocation of the optimum design manifold. The inverse model is defined over the objective space F with the values in the parameter space X, or $s_r : F \rightarrow X$. In this work, the following analytical form of s_r is assumed:

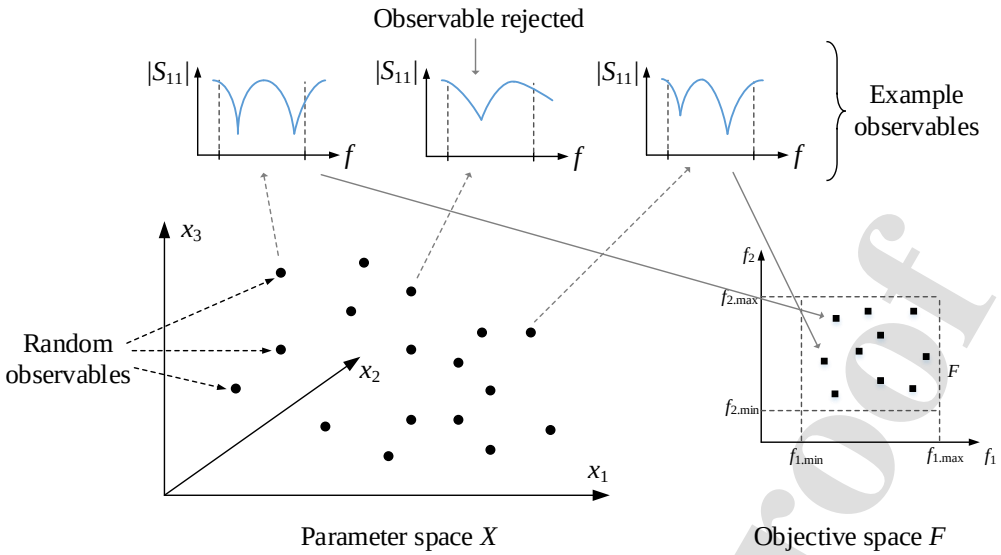


Fig. 2. Graphical illustration of the random observable generation procedure using an example of a dual-band antenna. The antenna is to be modelled over a two-dimensional objective space and within a three-dimensional parameter space. The random vectors that correspond to designs featuring operating frequencies within the assumed objective space are accepted, others are rejected. The final observable set $\{\boldsymbol{x}_r^{(j)}\}_{j=1,\ldots,Nr}$, will be used to generate the inverse regression model $s_r(\cdot)$.

$$s_r(\boldsymbol{f}) = \boldsymbol{s}_r\left([f_1 \ldots f_N]^{\mathrm{T}}\right) = \begin{bmatrix} s_{r.1}(\boldsymbol{f}) \\ \cdots \\ s_{r.n}(\boldsymbol{f}) \end{bmatrix} = \begin{bmatrix} a_{1.0} + a_{1.1}\exp\left(\sum_{k=1}^{N} a_{1.k+1} f_k\right) \\ \cdots \\ a_{n.0} + a_{n.1}\exp\left(\sum_{k=1}^{N} a_{n.k+1} f_k\right) \end{bmatrix} \tag{3}$$

which is motivated by a normally weakly-nonlinear dependence between the geometry parameters and operating figures such as resonant frequencies, bandwidths, etc. [64].

The model is identified by solving the nonlinear regression problems of the form

$$\left[a_{j.0}\ a_{j.1} \ldots a_{j.K+1}\right] = \arg\min_{[b_0\ b_1 \ldots b_{K+1}]} \sum_{k=1}^{N_r} w_k \left[s_{r.j}\left(\boldsymbol{f}_r^{(k)}\right) - x_{r.j}^{(k)}\right]^2, \quad j = 1, \ldots, n \tag{4}$$

where $x_{r.j}^{(k)}$ is the jth component of the observable vector $\boldsymbol{x}_r^{(k)}$. The weighting factors w_k are introduced to discriminate between "good" and "poor" observables. This is decided upon using the supplementary vectors $\boldsymbol{p}_r^{(j)} = [p_{r.1}^{(j)} \ \ldots \ p_{r.N}^{(j)}]^{\mathrm{T}}$, also extracted from EM-simulated antenna outputs. The entries of $\boldsymbol{p}_r^{(j)}$ contain relevant data concerning the design quality. For example, in the discussed case of a dual band antenna, these could be the reflection levels at the operating frequencies, the lower the better. The weighing factors are assigned as

$$w_k = \left[M - \max\{p_1(\boldsymbol{x}^{(j)}), ..., p_N(\boldsymbol{x}^{(j)})\} \right]^2, \qquad k = 1, \ldots, N_r \tag{5}$$

where $M = \max\{k = 1, \ldots, N_r, j = 1, \ldots, N : p_j^{(k)}\}$. Here, it is assumed that $p_j^{(k)}$ is nonnegative and lower values correspond to a better design. Again, the previously considered example of reflection levels at the operating frequencies conform to this assumption, as their moduli may change between zero and one.

The idea behind the weighting factors is to put more emphasis on high-quality observables, while still incorporating information contained in the remaining vectors. In particular, the designs characterized by higher weights are allocated closer to the optimum design manifold, therefore, they should determine the inverse surrogate in a more significant manner than the lower-quality ones. The aforementioned knowledge-based assessment of the quality of the observables allows for more efficient and reliable rendition of the inverse surrogate. A graphical illustration of the inverse regression model, corresponding to the example of Fig. 2, has been shown in Fig. 3.

2.3. Two-Stage Modeling: Stage Two – Domain Definition and Final Surrogate

The image $s_r(F)$ of the inverse model provides a rough approximation of the spatial allocation of the optimum design manifold. In order to encapsulate most of $U_F(F)$ into the domain of the surrogate model, $s_r(F)$ needs to be extended, which is realized using the procedure similar to the one developed for the nested kriging framework [64]. Let $\{\boldsymbol{v}_n^{(k)}(\boldsymbol{f})\}$, $k = 1, \ldots, n - N$, denote the orthonormal basis of vectors normal to $s_r(F)$ at the objective vector $\boldsymbol{f}$. Further, let $\boldsymbol{T} = [T_1 \ \ldots \ T_n]^{\mathrm{T}}$ be a vector of positive real numbers determining the amount of extension for the antenna parameters. Using these, we calculate the extension coefficients

$$\boldsymbol{\alpha}(\boldsymbol{f}) = [\alpha_1(\boldsymbol{f}) \ldots \alpha_{n-N}(\boldsymbol{f})]^{\mathrm{T}} = \left[|\boldsymbol{T}\boldsymbol{v}_n^{(1)}(\boldsymbol{f})| \ \ldots \ |\boldsymbol{T}\boldsymbol{v}_n^{(n-N)}(\boldsymbol{f})| \right]^{\mathrm{T}} \tag{6}$$

The domain X_S of the final surrogate model is then defined similarly as in [64]

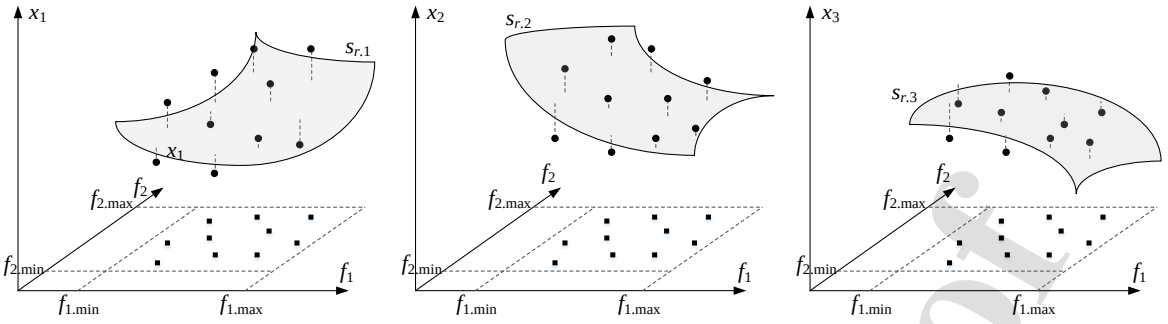


Fig. 3. Graphical illustration of the knowledge-based inverse regression model s_r established using the pre-selected observables $\boldsymbol{x}_r^{(j)}$ and the corresponding objective vectors $\boldsymbol{f}_r^{(j)}$. The components $s_{r.j}$ of the inverse model are obtained as in (3)-(5), and their images are visualized as the grey-shaded surfaces for antenna parameters x_1 (left), x_2 (middle), and x_3 (right), respectively.

$$X_S = \left\{ \begin{array}{c} \boldsymbol{x} = s_r(\boldsymbol{f}) + \sum_{k=1}^{n-N} \lambda_k \alpha_k(\boldsymbol{f}) \boldsymbol{v}_n^{(k)}(\boldsymbol{f}) : \boldsymbol{f} \in F, \\ -1 \le \lambda_k \le 1, k = 1, \ldots, n-N \end{array} \right\} \tag{7}$$

which is a set of vectors of the form of (7) obtained for all $\boldsymbol{f} \in F$, and all $\lambda_k \in [-1, 1]$, $k = 1, \ldots, n - N$. It can be noted that the extension boundaries of the domain are obtained for the extreme values of the coefficients λ_k, which are the surfaces defined by $S_+ = \left\{ \boldsymbol{x} \in X : \boldsymbol{x} = s_r(\boldsymbol{f}) + \sum_{k=1}^{n-N} \alpha_k(\boldsymbol{f}) \boldsymbol{v}_n^{(k)}(\boldsymbol{f}) \right\}$ and $S_- = \left\{ \boldsymbol{x} \in X : \boldsymbol{x} = s_r(\boldsymbol{f}) - \sum_{k=1}^{n-N} \alpha_k(\boldsymbol{f}) \boldsymbol{v}_n^{(k)}(\boldsymbol{f}) \right\}$.

It should be emphasized that the extension factors T_j (the entries of the vector $\boldsymbol{T}$) are individualized for all geometry parameters (e.g., in the nested kriging framework of [64], it was a scalar coeffficient). Also, the amount of extension can be estimated using the available observable set (in [64], it was arbitrarily chosen by the user). The knowledge-based assessment of the extension factors is arranged as follows. Let, for any observable pair $\{\boldsymbol{x}_r^{(j)}, \boldsymbol{f}_r^{(j)}\}$, the vector $P_k(\boldsymbol{x}_r^{(j)})$ be the element of the space $[l_k\ u_k] \times F$ (recall that l_k and u_k are parameter-space-defining lower and upper bounds for the kth parameter) that minimizes the distance between $[x_{r.k}^{(j)}\ (\boldsymbol{f}_r^{(j)})^{\mathrm{T}}]^{\mathrm{T}}$ and $[s_{r.k}(\boldsymbol{f})\ \boldsymbol{f}^T]^T$, $\boldsymbol{f} \in F$ (here, $x_{r.k}^{(j)}$ is the kth component of the observable vector $\boldsymbol{x}_r^{(j)}$), i.e.,

$$P_k(\boldsymbol{x}_r^{(j)}) = \arg\min_{\boldsymbol{f}\in F} \| [x_{r.k}^{(j)} \ (\boldsymbol{f}_r^{(j)})^{\mathrm{T}}]^{\mathrm{T}} - [s_r(\boldsymbol{f}) \ \boldsymbol{f}^{\mathrm{T}}]^{\mathrm{T}} \| \tag{8}$$

In other words, $P_k(\boldsymbol{x}_r^{(j)})$ determines the orthogonal projection of $[x_{r.k}^{(j)} \ (\boldsymbol{f}_r^{(j)})^{\mathrm{T}}]^{\mathrm{T}}$ onto the image of the kth component of the inverse regression model within the space $[l_k \ u_k] \times F$. Then,

$$d_{r.k}(\boldsymbol{x}_r^{(j)}) = \| [x_{r.k}^{(j)} \ (\boldsymbol{f}_r^{(j)})^{\mathrm{T}}]^{\mathrm{T}} - [s_r(P(\boldsymbol{x}_r^{(j)})) \ P(\boldsymbol{x}_r^{(j)})^{\mathrm{T}}]^{\mathrm{T}} \| \tag{9}$$

is the minimum distance between $[x_{r.k}^{(j)} \ (\boldsymbol{f}_r^{(j)})^{\mathrm{T}}]^{\mathrm{T}}$ and the said image. Graphically, $d_{r.k}$ can be interpreted as the distances between the respective observable components and the grey-colored surfaces in Fig. 3.

Using the above, the extension factor T_k can be then defined as

$$T_k = \frac{1}{2N_r} \sum_{j=1}^{N_r} d_{r.k}(\boldsymbol{x}_r^{(j)}) \tag{10}$$

which is half of the average distance between the observable component and the respective inverse regression model surface. It should be recalled that high-quality observables are allocated close to the optimum design manifold $U_F(F)$ (they also contribute more to the determination of the inverse model through the associated weighting factors w_k, cf. (4)), whereas poor observable are farther away from $U_F(F)$. Consequently, the average distance between the inverse regression model surface and the observables gives an overly pessimistic estimate of the necessary domain extension (i.e., if used, the domain would also contain many designs that are of not-so-good quality). To account for this, the definition (10) uses half of the average distance instead. As indicated in Section 3, this is sufficient to ensure design utility of the surrogate.

Having the domain X_S, the final surrogate $\boldsymbol{s}(\boldsymbol{x})$ is generated therein using kriging interpolation [71]. The training data set is denoted as $\{\boldsymbol{x}_B^{(k)}, \boldsymbol{R}(\boldsymbol{x}_B^{(k)})\}_{k=1,\ldots,N_B}$, where $\boldsymbol{x}_B^{(k)} \in X_S$, whereas $\boldsymbol{R}$ represents the relevant antenna characteristics obtained through EM analysis. Figure 4 presents graphical illustration of the design of experiments procedure similar to that proposed in [64], which capitalizes on the one-to-one transformation of the unit interval $[0,1]^n$ onto X_S.

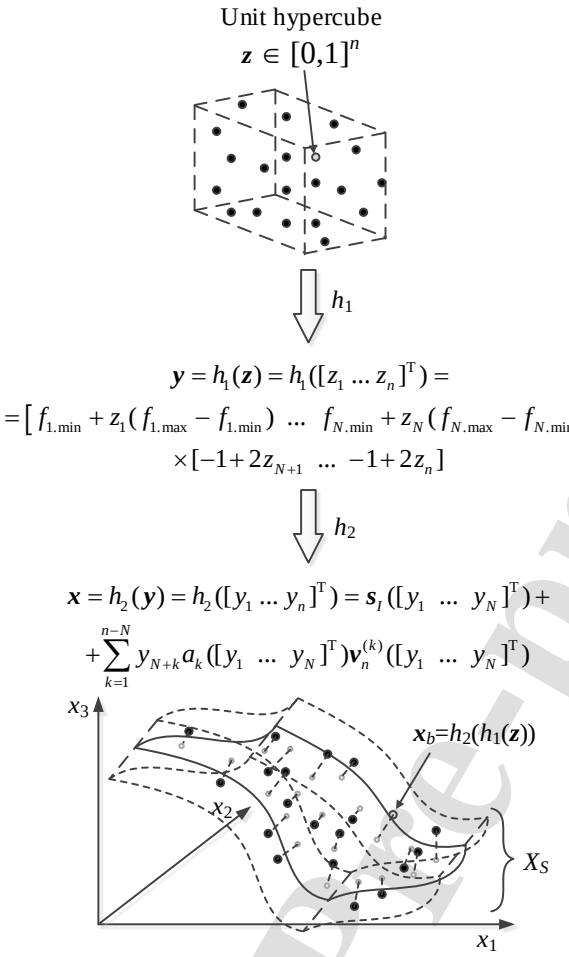


Fig. 4. Conceptual illustration of the employed design experiments procedure: (i) samples allocated using Latin Hypercube Sampling (LHS) [72] (top), (ii) transformation with the use of function h_1 onto the Cartesian product of the objective space F and $[-1,1]^{n-N}$ (middle), (iii) samples (black circles) mapped onto the constrained domain X_S using transformation h_2 (bottom); whereas the samples allocated onto the image of F before orthogonal shift are marked with light gray circles.

Observe that the data samples $\boldsymbol{x}_B^{(k)} = h_2(h_1(\boldsymbol{z}^{(k)}))$ allocated in X_S are uniformly distributed with respect to the objective space F, and not necessarily with respect to the parameter space X. Here, $\{\boldsymbol{z}^{(k)}\}$, $k = 1, \ldots, N_B$, denote a uniformly distributed set of samples in $[0,1]^n$ (obtained using Latin Hypercube Sampling, LHS [72]). Note that the same mapping can be also employed to carry out design procedures (e.g., parametric optimization) using the proposed two-stage surrogate model. Although a straightforward application of the model (i.e., operating with the domain X_S) might be challenging due to a relatively complex geometry of the domain, all operations can be executed within the unity interval $[0, 1]^n$, with the antenna parameters

mapped to X_S for the purpose of surrogate model evaluation. Furthermore, the vector $s_r(\boldsymbol{f})$ provides a good initial design for any target objective vector $\boldsymbol{f} \in F$.

The computational benefits of restricting the model domain as described in this section are expected to be similar to those of other performance-driven procedures [65], in particular, the nested kriging framework. By focusing on a volume-reduced region of the space, the surrogate can be constructed using small training data sets without formally limiting range of validity of the model, both in terms of the ranges of antenna geometry parameters and the operating conditions. On the other hand, the approach proposed here does not require any pre-optimized reference designs, which has a significant impact on the model setup cost. This will be demonstrated in detail in Section 3. An additional benefit is that defining the surrogate domain using random observables (cf. Section 2.2), provides a convenient way to determine its lateral dimensions (cf. (8)-(10)), which is a non-trivial problem for prior performance-driven techniques.

2.4. Complete Modeling Framework

This section provides a concise summary of the modeling procedure proposed in this work. The only user-defined control parameters are the number of (accepted) random observables N_r (typically 50 to 100) and the number N_B of training data samples for rendering the final surrogate model. The input data consists of the definition of the parameter space X, and the objective space F. The modeling process works as follows:

1. Generate random observables $\boldsymbol{x}_r^{(j)} \in X$ until N_r samples are found such that their corresponding objective vectors $\boldsymbol{f}_r^{(j)}$ are in the assumed objective space F (cf. Section 2.2); for these samples, also evaluate supplementary performance vectors $\boldsymbol{p}_r^{(j)}$;

2. Construct the inverse regression model s_r using $\{\boldsymbol{x}_r^{(j)},\boldsymbol{f}_r^{(j)}\}_{j = 1,\ldots,Nr}$, as the training data set (cf. (3)) and supplementary vectors $\boldsymbol{p}_r^{(j)}$ to compute the weighting factors for the observables (cf. (5));
3. Calculate the extension vector $\boldsymbol{T}$ as in (8)-(10) as well as the extension coefficients $\boldsymbol{\alpha}$ (cf. (6));
4. Define the surrogate model domain X_S using (7);
5. Perform design of experiments (see Fig. 4) to find N_B training data samples $\{\boldsymbol{x}_B^{(k)},\boldsymbol{R}(\boldsymbol{x}_B^{(k)})\}_{k = 1, \ldots, NB}$;
6. Identify the final surrogate model $\boldsymbol{s}$ using kriging interpolation of the training samples obtained in Step 5 extended by the observable set $\{\boldsymbol{x}_r^{(k)},\boldsymbol{R}(\boldsymbol{x}_r^{(k)})\}_{k = 1, \ldots, Nr}$;

Figure 5 shows the flow diagram of the modeling process. It should be noted that adding the observable set to the training data is expected to considerably improve the predictive power of the proposed model for small values of N_B, which is an additional advantage of the presented approach.

3. Demonstration Case Studies and Benchmarking

This section demonstrates application of the proposed two-stage procedure to modeling of several microstrip antenna structures. The selected structures cover a variety of antenna response types (narrow-band, multi-band, broadband, along with modeling of different characteristics, i.e., reflection and gain). Furthermore, considered modeling tasks are challenging, especially as compared to what is normally reported in the literature, both with respect to the dimensionality of the parameters space, and—more importantly—the parameter ranges. For the sake of benchmarking, our technique is compared to conventional surrogates, as well as the recent performance-driven approach, specifically, the nested kriging framework. The chosen benchmark set contains the methods representative for conventional modelling

approaches, as well as performance-driven ones, with the nested-kriging method viewed as a direct predecessor of the procedure introduced in this work (cf. Section 2). Design utility is validated by employing the two-stage surrogate to parameter tuning of the considered antennas.

3.1. Case I: Ring-Slot Antenna

The first verification case is the ring-slot antenna [73] shown in Fig. 6. Table 2 provides the details on the antenna design variables, its computational model, as well as the design objectives. The relative permittivity ε_r of the substrate is one of the figures of interest (cf. Section 2.1), i.e., one of the components of the objective space F. The feed line width w_f is adjusted for a given ε_r to ensure 50 ohm input impedance.

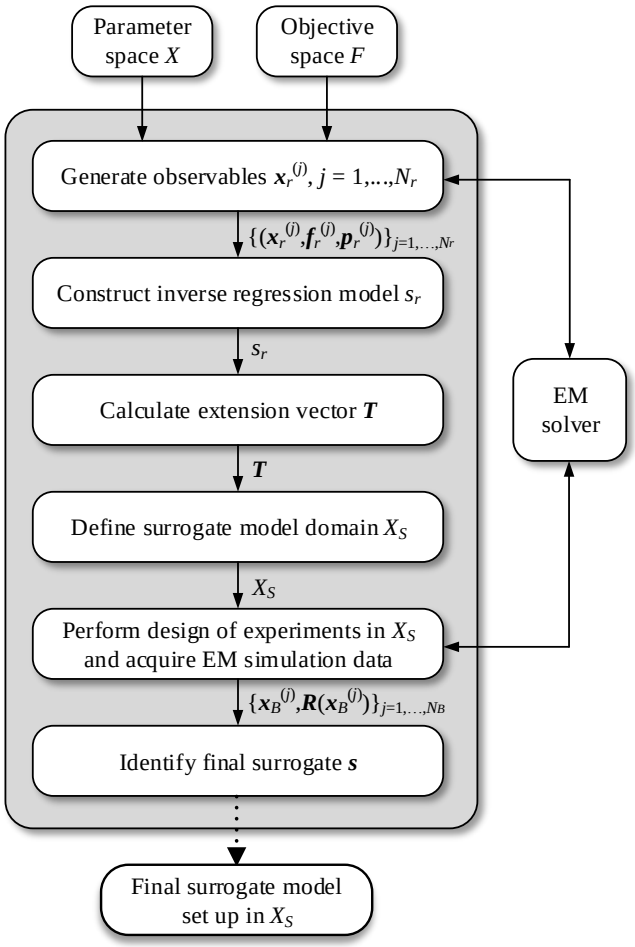


Fig. 5. Flow diagram of the proposed two-stage knowledge-based modeling procedure involving inverse regression surrogates and domain confinement.

The surrogate model of the antenna reflection coefficient is supposed to be valid over the entire objective space as given in Table 2. It should be noted that the modeling task is challenging not only due to the number of parameters but primarily due to their broad ranges (the average upper-to-lower bound ratio is around five).

The proposed modeling procedure has been validated with the number of random (accepted) observables (cf. Section 2) set to N_r = 50. These were obtained using 106 antenna simulations. The surrogates were constructed for N_B = 50, 100, 200, 400, and 800 training samples so that the scalability of the modeling error can be investigated as well. Furthermore, these numbers are considered typical for modelling of high-frequency components in engineering practice. The training data was allocated using modified Latin Hypercube Sampling (LHS) [72].

The following methods were used as benchmark:

- Conventional kriging interpolation model (set up in the space X);
- Conventional RBS model (set up in the space X);
- Nested kriging model [64] with the thickness parameter of T = 0.1 (domain X_S);

In the case of nested kriging, ten reference designs were employed. The computational cost of their acquisition is 864 EM simulations of the antenna (even when using feature-based optimization procedure [47]), which adds to the overall cost of the surrogate model setup.

Table 2. Case I: Ring-Slot Antenna

Parameters	
Substrate thickness	h = 0.76 mm
Designable parameters	$\boldsymbol{x} = [l_f\ l_d\ w_d\ r\ s\ s_d\ o\ g]^T$
Lower bounds	$\boldsymbol{l} = [22.0\ 3.5\ 0.3\ 6.5\ 3.0\ 0.5\ 3.5\ 0.2]^T$
Upper bounds	$\boldsymbol{u} = [27.0\ 8.0\ 2.3\ 16.0\ 7.0\ 5.5\ 6.0\ 2.3]^T$
Computational model	~300,000 cells
Simulation time	90 s
Design objective ranges	
Substrate permittivity	$2.0 \le \varepsilon_r \le 5.0$
Operating frequency	2.5 GHz $\le f_0 \le$ 6.5 GHz

Table 3 provides the numerical results, whereas Fig. 7 visualizes the inverse regression model surfaces along with the random observables for selected antenna parameters. It can be observed that observables are well correlated for the parameters that are dominant in terms of determining the antenna operating conditions, and spread for other variables. Figure 8 compares EM-simulated and surrogate-model-predicted antenna reflection characteristics at the selected test locations. The accuracy is estimated using a relative RMS error, calculated at the design $\boldsymbol{x}$ as $||\boldsymbol{R}_s(\boldsymbol{x}) - \boldsymbol{R}_f(\boldsymbol{x})||/||\boldsymbol{R}_f(\boldsymbol{x})||$, where $\boldsymbol{R}_s$ and $\boldsymbol{R}_f$ stand for the surrogate and EM-simulated reflection characteristics, respectively. The testing set included 100 samples allocated using LHS. It should be noted that utilization of the split-sample validation methods may result in a relatively large variance of modeling error estimation, especially for smaller training data sets, where a specific alloction of both training and testing samples may influence the error value in a noticeable manner. Notwithstanding, the results presented in Table 3 (as well as the result tables for other test cases) are consistent in the sense of the monotonicity of the model predictive power as a function of the training data set cardinality.

It can be observed that the accuracy of the proposed surrogate is considerably better than for the conventional models, both kriging and RBF. At the same time, it is comparable to that of the nested kriging framework. As a matter of fact, it is noticeably better for low values of N_B (here, 50 and 100), which is because the random observables are included into the training set (unlike in nested kriging). However, the most important advantage of the two-stage surrogate is no need for reference designs. Acquisition of this set required almost nine hundred EM simulations in the case of nested kriging, whereas the cost of generating observable within the proposed approach is only about one hundred simulations. Thus, while ensuring similar predictive power, the setup cost is considerably lower, e.g., by over 80 percent for $N_B = 50$, and as much as 45 percent for $N_B = 800$.

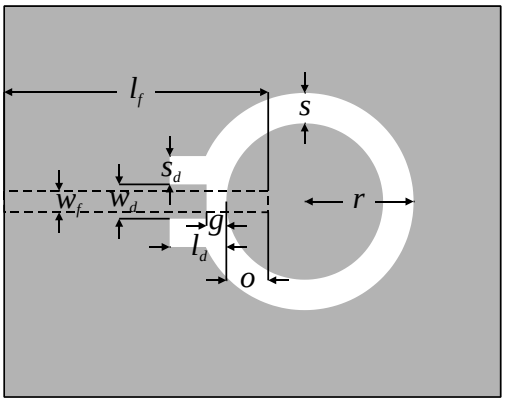
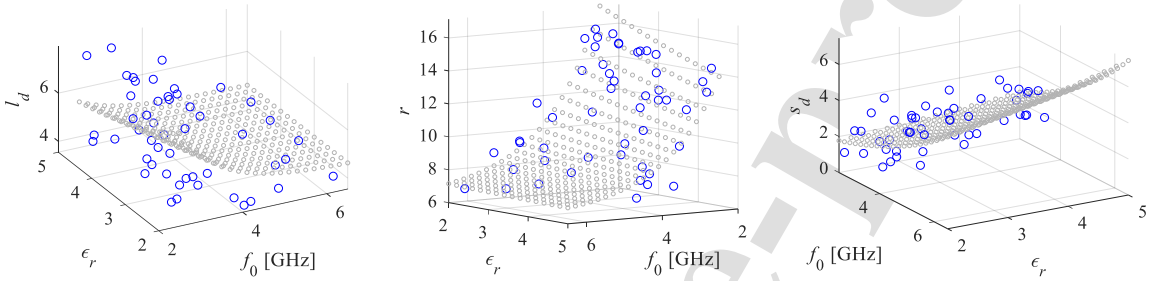


Fig. 6. Ring slot antenna geometry [73]. Microstrip feed is shown using a dashed line.



(a) (b) (c)

Fig. 7. Ring slot antenna: inverse regression model for selected antenna parameters: (a) l_d, (b) r, (c) s_d. Inverse model surfaces and random observables shown using grey and blue circles, respectively.

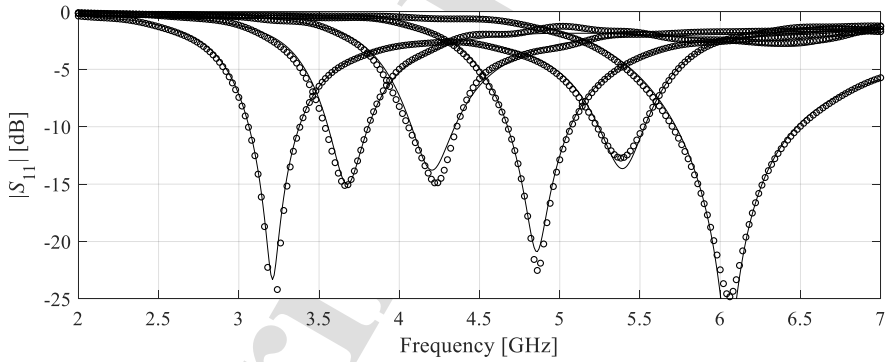


Fig. 8. Ring-slot antenna: reflection characteristics at the selected test designs: EM model (—), and the proposed two-stage surrogate (o). The surrogate set up using N_B = 400 training samples.

Design utility of the proposed two-stage surrogate has been demonstrated through parameter tuning of the antenna for several target operating frequencies f_0 and the substrate permittivity values ε_r, the antenna is to be implemented on. The results shown in Fig. 9 (see also Table 3) indicate that the model is well suited for the task. It does not only allow for identifying satisfactory designs across the range of considered operating frequencies and

permittivity values, but the agreement between the antenna response predicted by the surrogate and EM simulation data at the optimized design is excellent. Furthermore, the initial designs generated by the inverse regression model s_r are of high quality, i.e., the corresponding antenna operating frequencies are relatively close to the assumed targets.

Table 3. Ring-slot antenna of Fig. 6: Modeling results and benchmarking

Number of training samples	Modeling technique							
	Kriging		RBF		Nested kriging [64]		Two-stage surrogate (this work)	
	Modeling error	Model setup cost	Modeling error	Model setup cost	Modeling error	Model setup cost[$]	Modeling error	Model setup cost
50	56.9 %	50	61.0 %	50	19.4 %	914	13.4 %	156
100	50.8 %	100	53.2 %	100	12.9 %	964	9.9 %	206
200	35.8 %	200	37.9 %	200	7.7 %	1,064	6.9 %	306
400	31.5 %	400	34.1 %	400	5.1 %	1,264	5.4 %	506
800	25.6 %	800	27.2 %	800	3.7 %	1,664	4.4 %	906

[$]The cost includes acquisition of the reference designs, which is 864 EM simulations of the antenna when using feature-based optimization [47] as listed in the table. Conventional (minimax) optimization required 1,012 simulations.
[#]The cost includes generation of random observables, here, 106 simulations in total to yield N_r = 50 accepted samples.

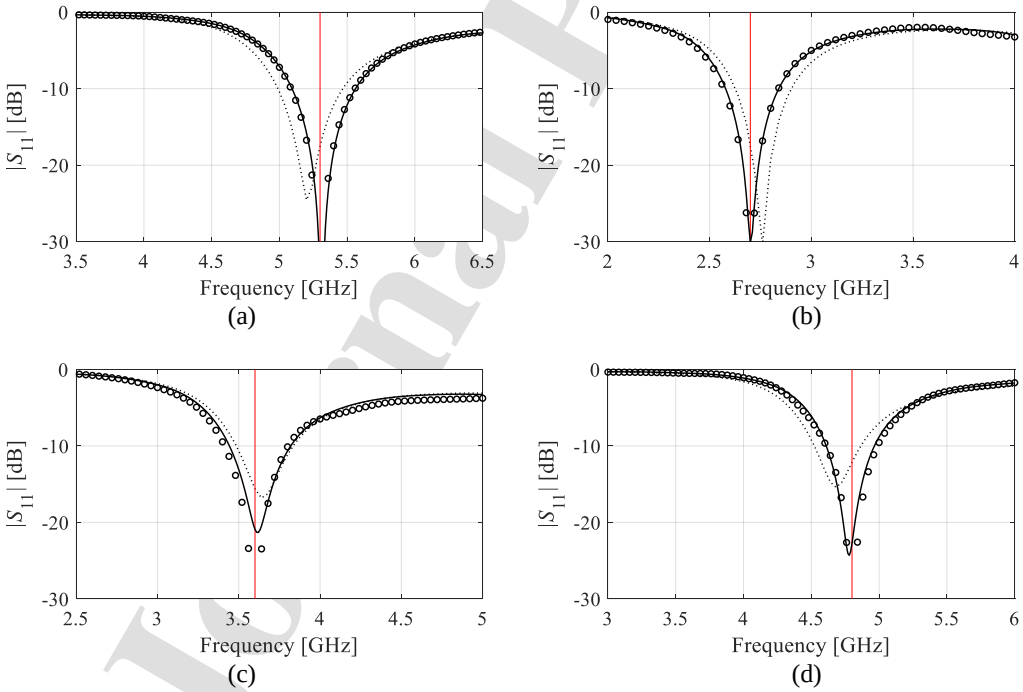


Fig. 9. Ring-slot antenna: EM-simulated reflection responses at the initial design (····) obtained using the inverse regression model, as well as surrogate (o) and EM-simulated response (—) at the design obtained by optimizing the proposed two-stage surrogate set up using N_B = 800 training samples. The designs obtained for the following objective vectors: (a) f_0 = 5.3 GHz, ε_r = 3.5, (b) f_0 = 2.7 GHz, ε_r = 4.3, (c) f_0 = 3.6 GHz, ε_r = 2.2, (d) f_0 = 4.8 GHz, ε_r = 4.3. The vertical line denotes the target operating frequency.

Table 4. Ring-slot antenna: Optimization results using the proposed two-stage surrogate

Target operating conditions		Geometry parameter values [mm]							
f_0 [GHz]	ε	l_f	l_d	w_d	r	s	s_d	o	g
5.3	3.5	24.3	4.67	1.15	7.60	5.10	4.70	5.31	1.77
2.7	4.3	24.9	5.60	1.43	14.0	5.26	2.42	4.65	1.09
3.6	2.2	23.7	5.19	1.10	11.6	4.68	2.40	4.59	0.86
4.8	4.3	24.7	4.93	1.22	7.98	5.27	4.57	5.26	1.71

3.2. Case II: Dual-Band Dipole Antenna

For the second example, let us consider a dual-band uniplanar dipole antenna shown in Fig. 10 [74], is implemented on RO4350 substrate. The details on the antenna designable and fixed parameters, the antenna computational model, as well as the design objectives are given in Table 5. In this case, the objective space is defined by the operating frequencies of the antenna at the lower and upper band, f_1 and f_2, respectively. Note that the average upper-to-lower bound ratio is around three with the maximum of seven for the last variable.

The proposed two-stage modeling framework has been validated in a similar way as for the first case study. The assumed number of random observables is N_r = 50, here, obtained using 230 EM simulations of the antenna structure. The surrogates were constructed for N_B = 50, 100, 200, 400, and 800 training samples. The benchmark methods are the same as in Section 3.1: kriging interpolation and RFB (both set up in the space X), as well as the nested kriging surrogate [64] with the thickness parameter of T = 0.1 (established in the domain X_S). The nested kriging model utilizes ten reference designs, the cost of their acquisition is 930 EM simulations of the antenna using feature-based optimization [47]. This cost is added to the overall expenses pertinent to the surrogate model setup.

The numerical results are provided in Table 6 for the proposed and the benchmark methods. Figure 11 shows the inverse regression model for selected antenna parameters. A

comparison of EM-simulated and surrogate-predicted responses for selected test points can be found in Fig. 12. The results are consistent with those obtained in Section 3.1.

On the one hand, the predictive power of the two-stage surrogate is significantly better than for the conventional models. On the other hand, it is comparable with the nested kriging for larger training data sets but considerably better for smaller ones (50 and 100 samples), which is due to exploiting the information contained in the random observables. Similarly as for the previous example, the computational cost of setting up the proposed model is much lower than for nested kriging because of not using any reference designs. The savings are as high as 72 percent for N_B = 50, and 41 percent for N_B = 800.

Table 5. Case I: Dual-Band Dipole Antenna

Parameters	
Substrate thickness	h = 0.76 mm
Substrate permittivity	ε_r = 3.5
Designable parameters	$\boldsymbol{x} = [l_1\ l_2\ l_3\ w_1\ w_2\ w_3]^T$
Other parameters [mm]	l_0 = 30, w_0 = 3, s_0 = 0.15, o = 5
Lower bounds	$\boldsymbol{l} = [29\ 5.0\ 17\ 0.2\ 1.5\ 0.5]^T$
Upper bounds	$\boldsymbol{u} = [42\ 12\ 25\ 0.6\ 5.2\ 3.5]^T$
Computational model	~100,000 cells
Simulation time	60 s
Design objective ranges	
Operating frequency (lower band)	2.0 GHz ≤ f_1 ≤ 3.0 GHz
Operating frequency (upper band)	4.0 GHz ≤ f_2 ≤ 5.5 GHz

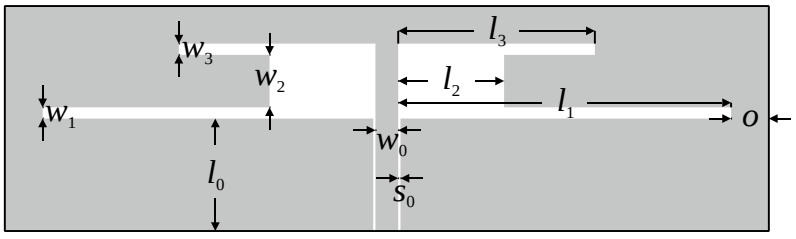
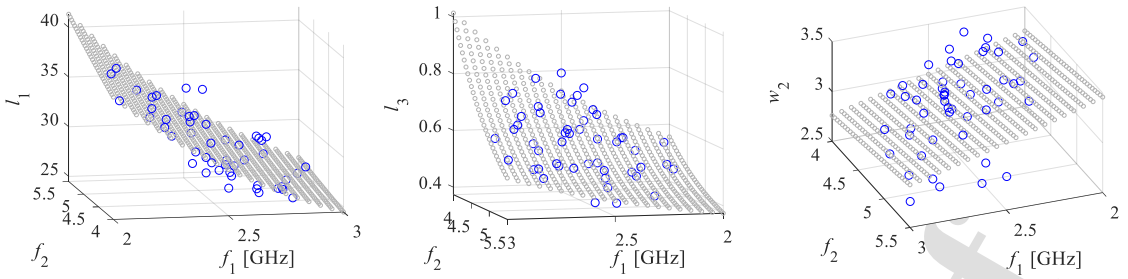


Fig. 10. Dual-band dipole antenna geometry [74].



(a) (b) (c)

Fig. 11. Dual-band antenna: inverse regression model for selected antenna parameters: (a) l_1, (b) l_3, (c) w_2. Inverse model surfaces and random observables shown using grey and blue circles, respectively.

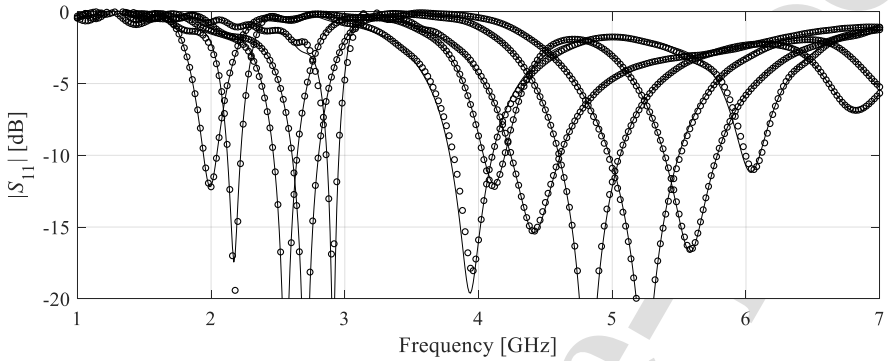


Fig. 12. Dual-band antenna: reflection characteristics at the selected test designs: EM model (—), and the proposed two-stage surrogate (o). The surrogate set up using N_B = 400 training samples.

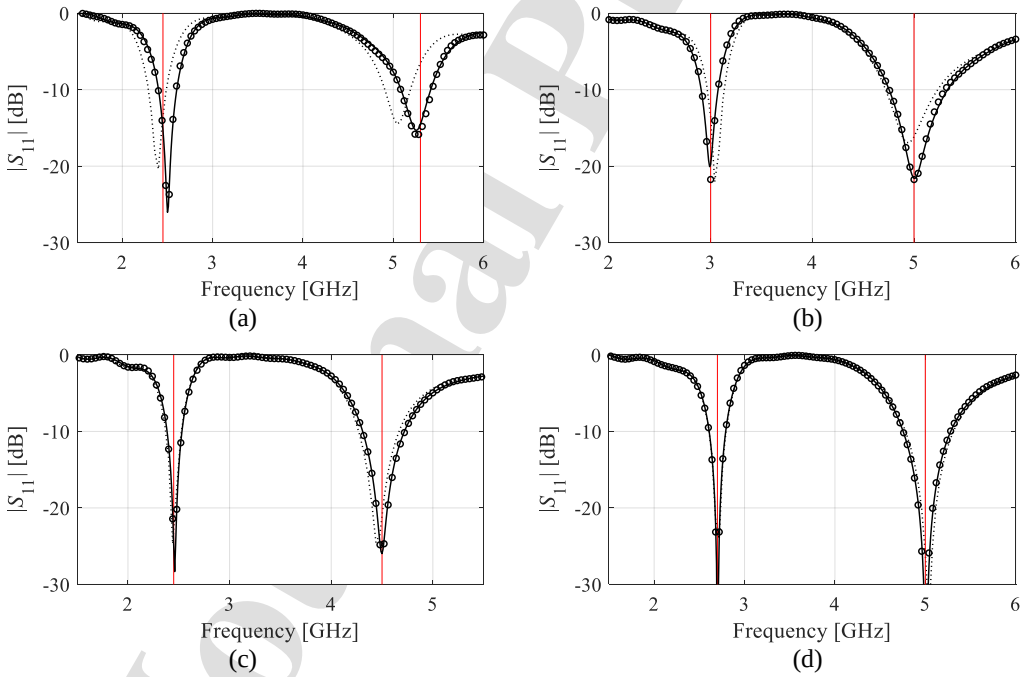


Fig. 13. Dual-band dipole antenna: EM-simulated reflection responses at the initial design (····) obtained using the inverse regression model, as well as surrogate (o) and EM-simulated response (—) at the design obtained by optimizing the proposed two-stage surrogate set up using N_B = 800 training samples. The designs obtained for the following objective vectors: (a) f_1 = 2.45 GHz, f_2 = 5.3 GHz, (b) f_1 = 3.0 GHz, f_2 = 5.0 GHz, (c) f_1 = 2.45 GHz, f_2 = 4.5 GHz, (d) f_1 = 2.7 GHz, f_2 = 5.0 GHz. The vertical lines denote the target operating frequencies.

Table 6. Dual-band dipole antenna of Fig. 10: Modeling results and benchmarking

Number of training samples	Modeling technique							
	Kriging		RBF		Nested kriging [64]		Two-stage surrogate (this work)	
	Modeling error	Model setup cost	Modeling error	Model setup cost	Modeling error	Model setup cost[$]	Modeling error	Model setup cost
50	21.7 %	50	24.9 %	50	9.9 %	980	7.3 %	280
100	17.3 %	100	19.8 %	100	6.4 %	1,030	5.1 %	330
200	12.6 %	200	14.3 %	200	4.4 %	1,130	3.8 %	430
400	9.3 %	400	10.5 %	400	3.8 %	1,330	3.1 %	630
800	7.2 %	800	8.7 %	800	3.4 %	1,730	2.5 %	1,030

[$]The cost includes acquisition of the reference designs, which is 930 EM simulations of the antenna when using feature-based optimization [47] as listed in the table. Conventional (minimax) optimization required 1,201 simulations.
[#]The cost includes generation of random observables, here, 230 simulations in total to yield $N_r = 50$ accepted samples.

To verify the design utility of the model, it has been employed to parameter tuning of the antenna for several pairs of target operating frequencies, as indicated in Table 6 and Fig. 13. It can be observed that the resonant frequencies of the antenna are properly allocated in all considered cases, and the initial design produced using the inverse regression model is of good quality, which makes local search sufficient to successfully conclude the optimization process. Furthermore, the agreement between the surrogate-predicted and EM-simulated antenna characteristics at the final designs is excellent.

3.3. Case III: Quasi-Yagi Antenna

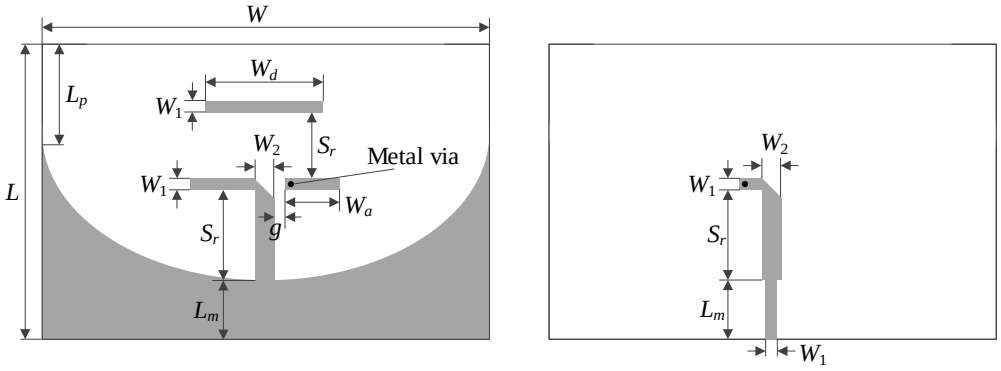
The third example is a quasi-Yagi antenna with a parabolic reflector [75], shown in Fig. 14. Table 8 gives the details on the antenna design variables, its computational model, as well as the design objective ranges. The feed line width W_1 is computed for a given substrate permittivity to ensure 50-ohm input impedance. The computational model is implemented in CST Microwave Studio. Just as for the first case study (Section 3.1), the permittivity is a part of the objective space F. The surrogate model is to represent the antenna reflection and realized gain characteristics over F. Note that the dimensionality of X is ten, which makes this case the most challenging from the point of view of surrogate modeling.

Table 7. Dual-band dipole antenna: Optimization results using the proposed two-stage surrogate

Target operating conditions		Geometry parameter values [mm]					
f_1 [GHz]	f_2 [GHz]	l_1	l_2	l_3	w_1	w_2	w_3
2.45	5.30	30.7	3.91	0.45	0.86	3.15	2.28
3.00	5.00	26.0	4.37	0.61	0.96	2.95	0.98
2.45	4.50	30.3	4.00	0.57	0.92	3.08	1.84
2.70	5.00	28.4	3.98	0.54	0.89	2.96	1.54

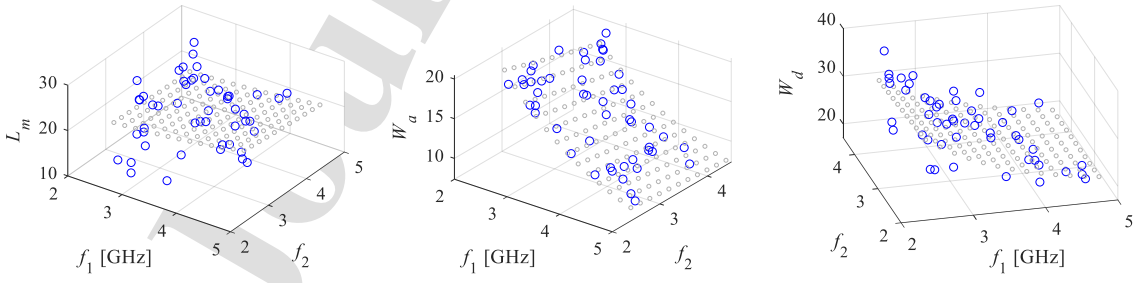
Table 8. Case I: Quasi-Yagi antenna

Parameters	
Substrate thickness	h = 1.5 mm
Designable parameters	$\boldsymbol{x} = [W\ L\ L_m\ L_p\ S_d\ S_r\ W_2\ W_a\ W_d\ g]^T$
Lower bounds	$\boldsymbol{l} = [22.0\ 3.5\ 0.3\ 6.5\ 3.0\ 0.5\ 3.5\ 0.2]^T$
Upper bounds	$\boldsymbol{u} = [27.0\ 8.0\ 2.3\ 16.0\ 7.0\ 5.5\ 6.0\ 2.3]^T$
Computational model	~300,000 cells
Simulation time	90 s
Design objective ranges	
Substrate permittivity	$2.0 \le \varepsilon_r \le 5.0$
Operating frequency	2.5 GHz $\le f_0 \le$ 6.5 GHz



(a) (b)

Fig. 14. Quasi-Yagi antenna geometry [75]: (a) top layer, (b) bottom layer.



(a) (b) (c)

Fig. 15. Quasi-Yagi antenna: inverse regression model for selected antenna parameters: (a) L_m, (b) W, (c) W_d. Inverse model surfaces and random observables shown using grey and blue circles, respectively.

Table 9. Quasi-Yagi antenna of Fig. 14: Modeling results and benchmarking

Number of training samples	Modeling technique							
	Kriging		RBF		Nested kriging [64]		Two-stage surrogate (this work)	
	Modeling error	Model setup cost	Modeling error	Model setup cost	Modeling error	Model setup cost[$]	Modeling error	Model setup cost
50	61.4 %	50	65.3 %	50	17.9 %	1,949	10.8 %	242
100	50.7 %	100	51.8 %	100	13.3 %	1,999	8.4 %	292
200	39.8 %	200	43.2 %	200	7.5 %	2,099	7.1 %	392
400	32.8 %	400	37.1 %	400	5.4 %	2,299	5.9 %	592
800	31.8 %	800	33.6 %	800	4.5 %	2,699	5.0 %	992

[$]The cost includes acquisition of the reference designs, which is 1,899 EM simulations of the antenna.
[#]The cost includes generation of random observables, here, 192 simulations in total to yield N_r = 50 accepted samples.

The experimental setup is the same as for the first two problems. We use N_r = 50 random observables which are obtained using 192 EM simulations of the antenna structure. The surrogates were constructed for N_B = 50, 100, 200, 400, and 800 training samples. The benchmark methods include kriging interpolation and RBF (both set up in the space X), and the nested kriging surrogate with the thickness parameter of T = 0.05 (in the domain X_S). The nested kriging model utilizes eight reference designs, the cost of their acquisition is 1,899 EM simulations of the antenna. This cost is much higher than for the previous examples due to problem dimensionality as well as more involved formulation of the optimization task [76].

Table 9 shows the numerical results for the two-stage surrogate and the benchmark techniques. Visualization of the inverse regression model for selected antenna parameters can be found in Fig. 15, whereas Fig. 16 shows a comparison of EM-simulated and surrogate-predicted responses for selected test points. The results indicate that the predictive power of the two-stage surrogate is considerably better than for the conventional models, and it is comparable with the nested kriging for training data set of 200 samples and larger, but it is superior for N_B = 50 and 100. The aforementioned performance is consistent with the previous verification cases discussed in Sections 3.1 and 3.2. However, for this antenna, the computational savings with respect to nested kriging are even more pronounced due to high

cost of reference design acquisition for the latter (1,899 EM simulations). For example, the cost reduction of the surrogate model setup is 88 percent for N_B = 50, and 63 percent for N_B = 800.

The design utility of the model has been demonstrated by optimizing the antenna for the target operating frequency f_0 assuming a given substrate permittivity ε_r (both within the assumed objective space). The optimization process aimed at ensuring at least eight-percent fractional bandwidth (symmetric w.r.t. f_0) and maximizing the average realized gain within the same bandwidth. The results are shown in Table 10 and Fig. 17. It can be noticed that the operating bandwidth of the antenna is well controlled, and the inverse regression model yields good initial designs for further tuning. Also, the agreement between the EM-simulated antenna responses and the predictions of the two-stage surrogate is excellent.

Table 10. Quasi-Yagi antenna: Optimization results using the proposed two-stage surrogate

Target operating conditions		Geometry parameter values [mm]									
f_0 [GHz]	ε_r	W	L	L_m	L_p	S_d	S_r	W_2	W_a	W_d	g
3.6	3.0	118.4	71.8	22.4	22.3	13.5	13.7	3.96	13.3	23.7	0.69
4.1	3.0	118.5	71.8	22.9	22.3	14.1	13.3	3.72	11.2	21.8	0.68
2.7	4.4	114.3	66.3	19.4	21.4	13.8	16.6	3.96	16.6	28.6	0.83
4.2	4.4	117.2	65.0	24.0	22.4	15.8	12.4	3.86	8.22	18.8	0.69

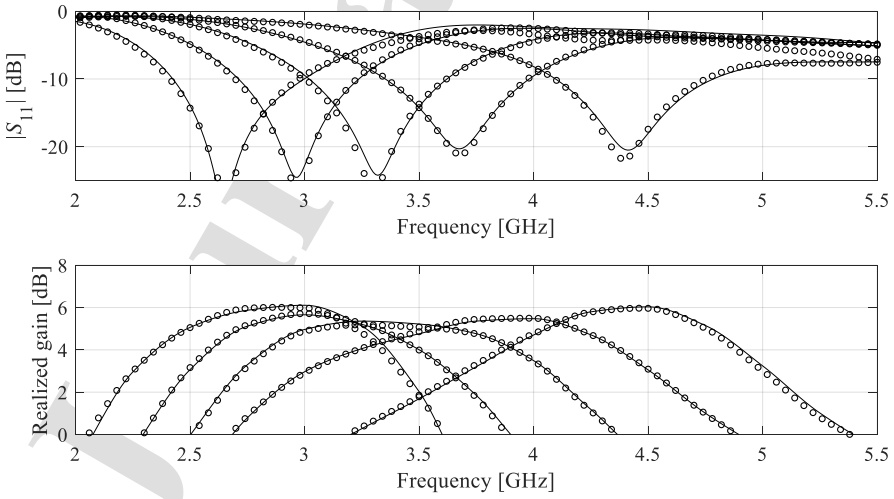


Fig. 16. Quasi-Yagi antenna: reflection (top) and realized gain (bottom) characteristics at the selected test designs: EM model (—), and the proposed two-stage surrogate (o). The surrogate set up using N_B = 400 training samples.

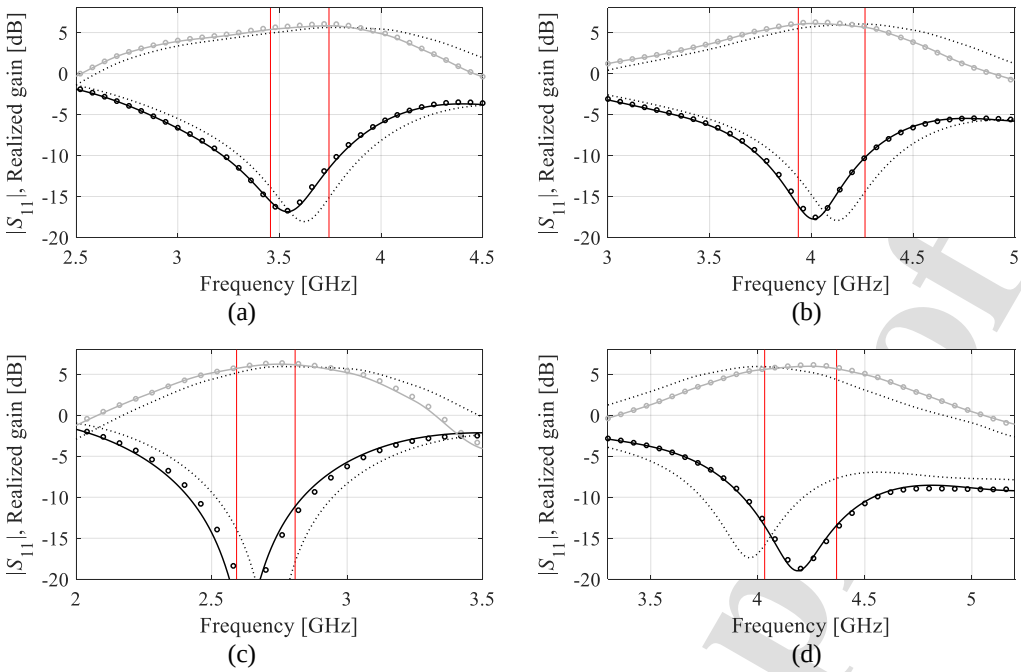


Fig. 17. Quasi-Yagi antenna: EM-simulated reflection responses at the initial design (····) obtained using the inverse regression model, as well as surrogate (o) and EM-simulated response (—) at the design obtained by optimizing the proposed two-stage surrogate set up using N_B = 800 training samples. The designs obtained for the following objective vectors: (a) f_0 = 3.6 GHz, ε_r = 3.0, (b) f_0 = 4.1 GHz, ε_r = 3.0, (c) f_0 = 2.7 GHz, ε_r = 4.4, (d) f_0 = 4.2 GHz, ε_r = 4.4. Vertical lines denote the target operating band of the antenna. Reflection and realized gain characteristics shown using black and gray lines, respectively.

4. Conclusion

The paper proposed a novel approach to low-cost surrogate modeling of antenna structures. Our two-stage knowledge-based methodology capitalizes on the performance-driven modeling concept. The confined domain of the final surrogate is defined using an auxiliary inverse regression model. At this stage of the modeling process, unlike the traditional constrained modeling methods, the procedure presented here does not require any pre-optimized reference designs. Instead, the inverse model is rendered based on a problem-specific knowledge in the form of a set of random observables generated within the assumed parameter space. This has a profound effect on the computational cost of the surrogate model setup, which is reduced by up to eighty percent, as compared to the nested kriging framework, the recent performance-driven approach. As demonstrated using three antenna examples, the

proposed technique ensures superior predictive power of the surrogate while using limited numbers of training data samples. The accuracy-related advantages are particularly pronounced for smaller data sets as the random observables are also taken into account for model identification. The knowledge-based definition of the confined domain also provides means for straightforward uniform sampling as well as model optimization. Finally, the definition of the model domain permits a convenient determination of its lateral size by exploiting to an even larger extent the knowledge about the system under design, which has been a non-trivial problem for the previous performance-driven modeling approaches. Design utility of the two-stage approach has been corroborated through application case studies (antenna optimization).

The future work will include investigation of the scalability properties for high-dimensional parameter spaces (beyond ten parameters), as well as the incorporation of response feature technology (i.e., handling the modeling task at the level of the system response features or characteristic points, rather than the frequency characteristics in their entirety), which should lead to further computational savings. Furthermore, dimensionality reduction methods (primarily based on spectral analysis of the observable set) will be employed to span the model domain along the most important directions only and achieve additional cost reduction and predictive power improvements. Finally, the presented technique will be applied to construct replacement models of other types of high-frequency components.

Acknowledgement

The authors would like to thank Dassault Systemes, France, for making CST Microwave Studio available. This work is partially supported by the Icelandic Centre for Research (RANNIS) Grant 217771 and by National Science Centre of Poland Grant 2018/31/B/ST7/02369.

References

[1] K. Lu and K. W. Leung, "On the circularly polarized parallel-plate antenna," *IEEE Trans. Ant. Propag.,* vol. 68, no. 1, pp. 3-12, 2020.

[2] O. J. Famoriji and Z. Xu, "Antenna feed array synthesis for efficient communication systems," *IEEE Sensors J.*, vol. 20, no. 24, pp. 15085-15098, 2020.

[3] B. L. G. Jonsson, S. Shi, L. Wang, F. Ferrero, and L. Lizzi, "On methods to determine bounds on the Q-factor for a given directivity," *IEEE Trans. Ant. Propag.*, vol. 65, no. 11, pp. 5686-5696, 2017.

[4] U. Ullah and S. Koziel, "A broadband circularly polarized wide-slot antenna with a miniaturized footprint," *IEEE Ant. Wireless Prop. Lett.*, vol. 17, no. 12, pp. 2454-2458, 2018.

[5] L. Y. Nie, X. Q. Lin, Z. Q. Yang, J. Zhang, and B. Wang, "Structure-shared planar UWB MIMO antenna with high isolation for mobile platform," *IEEE Trans. Ant. Propag.*, vol. 67, no. 4, pp. 2735-2738, 2019.

[6] D. Wen, Y. Hao, M. O. Munoz, H. Wang, and H. Zhou, "A compact and low-profile MIMO antenna using a miniature circular high-impedance surface for wearable applications," *IEEE Trans. Ant. Propag.*, vol. 66, no. 1, pp. 96-104, 2018.

[7] A. Krishna, A. F. Abdelaziz, and T. Khattab, "Patch antenna array designs for wireless communication applications inside jet engines," *IEEE Trans. Ant. Propag.*, vol. 67, no. 2, pp. 971-979, 2019.

[8] S. Lei, Y. Yang, H. Hu, Z. Zhao, B. Chen, and X. Qiu, "Power gain optimization method for wide-beam array antenna via convex optimization," *IEEE Trans. Ant. Propag.*, vol. 67, no. 3, pp. 1620-1629, 2019.

[9] J. Cui, A. Zhang, and X. Chen, "An omnidirectional multiband antenna for railway application," *IEEE Ant. Wireless Prop. Lett.*, vol. 19, no. 1, pp. 54-58, 2020.

[10] J. Wang, H. Wong, Z. Ji, and Y. Wu, "Broadband CPW-fed aperture coupled metasurface antenna," *IEEE Ant. Wireless Prop. Lett.*, vol. 18, no. 3, pp. 517-520, 2019.

[11] G. Das, A. Sharma, R. K. Gangwar, and M. S. Sharawi, "Performance improvement of multiband MIMO dielectric resonator antenna system with a partially reflecting surface," *IEEE Ant. Wireless Prop. Lett.*, vol. 18, no. 10, pp. 2105-2109, 2019.

[12] Y. Feng, J. Li, B. Cao, J. Liu, G. Yang, and D. Wei, "Cavity-backed broadband circularly polarized cross-dipole antenna," *IEEE Ant. Wireless Prop. Lett.*, vol. 18, no. 12, pp. 2681-2685, 2019.

[13] J. Liu, Z. Weng, Z. -Q. Zhang, Y. Qiu, Y. -X. Zhang, and Y. -C. Jiao, "A wideband pattern diversity antenna with a low profile based on metasurface," *IEEE Ant. Wireless Prop. Lett.*, vol. 20, no. 3, pp. 303-307, 2021.

[14] N. Liu, L. Zhu, Z. Liu, Z. Zhang, G. Fu, and Y. Liu, "Cross-polarization reduction of a shorted patch antenna with broadside radiation using a pair of open-ended stubs," *IEEE Trans. Ant. Propag.*, vol. 68, no. 1, pp. 13-20, 2020.

[15] T. T. Le, H. H. Tran, and H. C. Park, "Simple-structured dual-slot broadband circularly polarized antenna," *IEEE Ant. Wireless Prop. Lett.*, vol. 17, no. 3, pp. 476-479, 2018.

[16] J. Duan and L. Zhu, "An EH0-mode microstrip leaky-wave antenna with transversal single beam via periodical loading of shorting pins and U-shaped slots," *IEEE Ant. Wireless Prop. Lett.*, vol. 19, no. 12, pp. 2187-2191, 2020.

[17] K. Wei, J. Y. Li, L. Wang, R. Xu, and Z. J. Xing, "A new technique to design circularly polarized microstrip antenna by fractal defected ground structure," *IEEE Trans. Ant. Propag.*, vol. 65, no. 7, pp. 3721-3725, 2017.

[18] M. Kovaleva, D. Bulger, and K. P. Esselle, "Comparative study of optimization algorithms on the design of broadband antennas," *IEEE J. Multiscale Multiphysics Comp. Techn.*, vol. 5, pp. 89-98, 2020.

[19] J. A. Easum, J. Nagar, P. L. Werner and D. H. Werner, "Efficient multiobjective antenna optimization with tolerance analysis through the use of surrogate models," *IEEE Trans. Ant. Propag.*, vol. 66, no. 12, pp. 6706-6715, 2018.

[20] A. Pietrenko-Dabrowska, S. Koziel, and M. Al-Hasan, „Expedited yield optimization of narrow- and multi-band antennas using performance-driven surrogates," *IEEE Access*, pp. 143104-143113, 2020.

[21] Y. Liu, M. Li, R. L. Haupt, and Y. J. Guo, "Synthesizing shaped power patterns for linear and planar antenna arrays including mutual coupling by refined joint rotation/phase optimization," *IEEE Trans. Ant. Propag.*, vol. 68, no. 6, pp. 4648-4657, 2020.

[22] Q. Li, Q. Chu, Y. Chang, and J. Dong, "Tri-objective compact log-periodic dipole array antenna design using MOEA/D-GPSO," *IEEE Trans. Ant. Propag.*, vol. 68, no. 4, pp. 2714-2723, 2020.

[23] D. Z. Zhu, P. L. Werner, and D. H. Werner, "Design and optimization of 3-D frequency-selective surfaces based on a multiobjective lazy ant colony optimization algorithm," *IEEE Trans. Ant. Propag.*, vol. 65, no. 12, pp. 7137-7149, 2017.

[24] S. E. Hosseininejad, N. Komjani, and A. Mohammadi, "Accurate design of planar slotted SIW array antennas," *IEEE Ant. Wireless Prop. Lett.*, vol. 14, pp. 261-264, 2015.

[25] C. Nadia, M. Tomader, and M. Benbrahim, "Array antenna characteristics improvement: Parasitic patches (two disposals) and multilayer substrate techniques," *Int. Conf. Opt. Appl. (ICOA)*, Mohammedia, Morocco, pp. 1-6, 2018.

[26] J. Du and C. Roblin, "Stochastic surrogate models of deformable antennas based on vector spherical harmonics and polynomial chaos expansions: application to textile antennas," *IEEE Trans. Ant. Prop.*, vol. 66, no. 7, pp. 3610-3622 , 2018.

[27] A. Kouassi, N. Nguyen-Trong, T. Kaufmann, S. Lalléchère, P. Bonnet, and C. Fumeaux, "Reliability-aware optimization of a wideband antenna," *IEEE Trans. Ant. Prop.*, vol. 64, no. 2, pp. 450-460, 2016.

[28] L. S. Kalantari and M. H. Bakr, "Wideband cloaking of objects with arbitrary shapes exploiting adjoint sensitivities," *IEEE Trans. Ant. Prop.*, vol. 64, no. 5, pp. 1963-1968, 2016.

[29] F. Feng, J. Zhang, W. Zhang, Z. Zhao, J. Jin, and Q. Zhang, "Coarse- and fine-mesh space mapping for EM optimization incorporating mesh deformation," *IEEE Microwave Wireless Comp. Lett.*, vol. 29, no. 8, pp. 510-512, 2019.

[30] S. Koziel and A. Pietrenko-Dabrowska, "Reduced-cost electromagnetic-driven optimization of antenna structures by means of trust-region gradient-search with sparse Jacobian updates," *IET Microwaves Ant. Prop.*, vol. 13, no. 10, pp. 1646-1652, 2019.

[31] S. Koziel and A. Pietrenko-Dabrowska, "Variable-fidelity simulation models and sparse gradient updates for cost-efficient optimization of compact antenna input characteristics," *Sensors*, vol. 19, no. 8, 2019.

[32] J. Wang, X.S. Yang, and B.Z. Wang, "Efficient gradient-based optimization of pixel antenna with large-scale connections," *IET Microwaves Ant. Prop.*, vol. 12, no. 3, pp. 385-389, 2018.

[33] J. Xu, M. Li, R. Chen, "Space mapping optimisation of 2D array elements arrangement to reduce the radar cross-scattering," *IET Microwaves Ant. Propag.*, vol. 11, no. 11, pp. 1578-1582, 2017.

[34] Y. Su, J. Li, Z. Fan, and R. Chen, „Shaping optimization of double reflector antenna based on manifold mapping," *Int. Applied Comp. Electromagnetics Soc. Symp. (ACES)*, Suzhou, China, pp. 1-2, 2017.

[35] S. Koziel and S. D. Unnsteinsson, "Expedited design closure of antennas by means of trust-region-based adaptive response scaling," *IEEE Ant. Wireless Propag. Lett.*, vol. 17, no. 6, pp. 1099–1103, June 2018.

[36] S. Koziel, S. Ogurtsov, Q.S. Cheng, and J.W. Bandler, "Rapid EM-based microwave design optimization exploiting shape-preserving response prediction and adjoint sensitivities," *IET Microwaves, Ant. Prop.*, vol., 8, no. 10, pp. 775-781, 2014.

[37] M. Salucci, L. Tenuti, G. Oliveri and A. Massa, "Efficient prediction of the EM response of reflectarray antenna elements by an advanced statistical learning method," *IEEE Trans. Ant. Prop.*, vol. 66, no. 8, pp. 3995-4007, 2018.

[38] Q. Zhou, Y. Wang, P. Jiang, X. Shao, S.-K. Choi, J. Hu, L. Cao, X. Meng, "An active learning radial basis function modeling method based on self-organization maps for simulation-based design problems," *Knowledge-Based Systems*, vol. 131, pp. 10-27, 2017.

[39] J. Dong, W. Qin, and M. Wang, "Fast multi-objective optimization of multi-parameter antenna structures based on improved BPNN surrogate model," *IEEE Access*, vol. 7, pp. 77692-77701, 2019.

[40] J.P. Jacobs, "Characterization by Gaussian processes of finite substrate size effects on gain patterns of microstrip antennas," *IET Microwaves Ant. Prop.*, vol. 10, no. 11, pp. 1189-1195, 2016.

[41] J. Cai, J. King, C. Yu, J. Liu, and L. Sun, "Support vector regression-based behavioral modeling technique for RF power transistors," *IEEE Microwave and Wireless Comp. Lett.*, vol. 28, no. 5, pp. 428-430, 2018.

[42] Q. Gu, Q. Wang, X. Li, X. Li, "A surrogate-assisted multi-objective particle swarm optimization of expensive constrained combinatorial optimization problems," *Knowledge-Based Systems*, In-press, 107049, 2021.

[43] Y. Zhao, C. Sun, J. Zeng, Y. Tan, G. Zhang, "A surrogate-ensemble assisted expensive many-objective optimization," Knowledge-Based Syst., vol. 211, art. no. 106520, 2021.

[44] S. Xiao, G.Q. Liu, K.L. Zhang, Y.Z. Jing, J.H. Duan, P. Di Barba, and J.K. Sykulski, "Multi-objective Pareto optimization of electromagnetic devices exploiting kriging with Lipschitzian optimized expected improvement," *IEEE Trans. Magn.*, vol. 54, no. 3, paper ID 7001704, 2018.

[45] D. Spina, F. Ferranti, G. Antonini, T. Dhaene, and L. Knockaert, "Efficient variability analysis of electromagnetic systems via polynomial chaos and model order reduction," *IEEE Trans. Comp. Packaging Manufacturing Techn.*, vol. 4, no. 6, pp. 1038-1051, 2014.

[46] S. Koziel and A. Pietrenko-Dabrowska, "Expedited feature-based quasi-global optimization of multi-band antennas with Jacobian variability tracking," *IEEE Access*, vol. 8, pp. 83907-83915, 2020.

[47] S. Koziel, "Fast simulation-driven antenna design using response-feature surrogates," *Int. J. RF & Micr. CAE*, vol. 25, no. 5, pp. 394-402, 2015.

[48] A. Pietrenko-Dabrowska and S. Koziel, „Fast design closure of compact microwave components by means of feature-based metamodels," *Electronics*, vol. 10, 2021.

[49] C. Zhang, F. Feng, V. Gongal-Reddy, Q. J. Zhang, and J. W. Bandler, "Cognition-driven formulation of space mapping for equal-ripple optimization of microwave filters," *IEEE Trans. Microwave Theory Techn.*, vol. 63, no. 7, pp. 2154-2165, 2015.

[50] C. Fu, P. Wang, L. Zhao, X. Wang, "A distance correlation-based Kriging modeling method for high-dimensional problems," Knowledge-Based Systems, vol. 206, art. no. 106356, 2020.

[51] R. Yao, Y. Zhang, S. Wang, N. Qi, N. I. Miridakis, and T. A. Tsiftsis, "Deep neural network assisted approach for antenna selection in untrusted relay networks," *IEEE Wireless Comm. Lett.*, vol. 8, no. 6, pp. 1644-1647, 2019.

[52] S. Skaria, A. Al-Hourani, M. Lech, and R. J. Evans, "Hand-gesture recognition using two-antenna doppler radar with deep convolutional neural networks," *IEEE Sensors J.*, vol. 19, no. 8, pp. 3041-3048, 2019.

[53] L. -Y. Xiao, W. Shao, F. -L. Jin, B. -Z. Wang, and Q. H. Liu, "Radial basis function neural network with hidden node interconnection scheme for thinned array modeling," *IEEE Ant. Wireless Propag. Lett.*, vol. 19, no. 12, pp. 2418-2422, 2020.

[54] A. Petrocchi, A. Kaintura, G. Avolio, D. Spina, T. Dhaene, A. Raffo, and D.M.P.-P. Schreurs, „Measurement uncertainty propagation in transistor model parameters via polynomial chaos expansion," *IEEE Microwave Wireless Comp. Lett.*, vol. 27, no. 6, pp. 572-574, 2017.

[55] D. R. Prado, J. A. López-Fernández, M. Arrebola, M. R. Pino, and G. Goussetis, "Wideband shaped-beam reflectarray design using support vector regression analysis," *IEEE Ant. Wireless Propag. Lett.*, vol. 18, no. 11, pp. 2287-2291, 2019.

[56] D. Gorissen, K. Crombecq, I. Couckuyt, T. Dhaene, and P. Demeester, „A surrogate modeling and adaptive sampling toolbox for computer based design," *J. Machine Learning Research*, vol. 11, pp. 2051-2055, 2010.

[57] S. Marelli and B. Sudret, „UQLab: a framework for uncertainty quantification in Matlab," *2nd Int. Conf. on Vulnerability and Risk Analysis and Management (ICVRAM 2014)*, University of London, UK, July 13-15, pp. 2554-2563, 2014.

[58] A.C. Yücel, H. Bağcı, and E. Michielssen, "An ME-PC enhanced HDMR method for efficient statistical analysis of multiconductor transmission line networks," *IEEE Trans. Comp. Packaging and Manufacturing Techn.*, vol .5, no. 5, pp. 685-696, 2015.

[59] R. Hu, V. Monebhurrun, R. Himeno, H. Yokota, and F. Costen, „An adaptive least angle regression method for uncertainty quantification in FDTD computation," *IEEE Trans. Ant. Prop.*, vol. 66, no. 12, pp. 7188-7197, 2018.

[60] J.P. Jacobs and S. Koziel, "Two-stage framework for efficient Gaussian process modeling of antenna input characteristics," *IEEE Trans. Antennas Prop.*, vol. 62, no. 2, pp. 706-713, 2014.

[61] M.C. Kennedy and A. O'Hagan, "Predicting the output from complex computer code when fast approximations are available", *Biometrika*, vol. 87, pp. 1-13, 2000.

[62] S. Koziel, "Low-cost data-driven surrogate modeling of antenna structures by constrained sampling," *IEEE Antennas Wireless Prop. Lett.*, vol. 16, pp. 461-464, 2017.

[63] S. Koziel and A.T. Sigurdsson, "Triangulation-based constrained surrogate modeling of antennas," *IEEE Trans. Ant. Prop.*, vol. 66, no. 8, pp. 4170-4179, 2018.

[64] S. Koziel and A. Pietrenko-Dabrowska, "Performance-based nested surrogate modeling of antenna input characteristics," *IEEE Trans. Ant. Prop.*, vol. 67, no. 5, pp. 2904-2912, 2019.

[65] S. Koziel and A. Pietrenko-Dabrowska, *Performance-driven surrogate modeling of high-frequency structures*, Springer, New York, 2020.

[66] A. Pietrenko-Dabrowska and S. Koziel, "Antenna modeling using variable-fidelity EM simulations and constrained co-kriging," *IEEE Access*, vol. 8, no. 1, pp. 91048-91056, 2020.

[67] A. Pietrenko-Dabrowska and S. Koziel, "Reliable surrogate modeling of antenna input characteristics by means of domain confinement and principal components," *Electronics*, vol. 9, no. 5, pp. 1-16, 2020.

[68] A. Pietrenko-Dabrowska, S. Koziel, and M. Al-Hasan, „Cost-efficient bi-layer modeling of antenna input characteristics using gradient kriging surrogates," *IEEE Access*, vol. 8, pp. 140831-140839, 2020.

[69] M. Abdullah and S. Koziel, "A novel versatile decoupling structure and expedited inverse-model-based re-design procedure for compact single-and dual-band MIMO antennas," *IEEE Access*, vol. 9, pp. 37656-37667, 2021.

[70] S. Koziel and A. Pietrenko-Dabrowska, "On computationally-efficient reference design acquisition for reduced-cost constrained modeling and re-design of compact microwave passives," *IEEE Access*, vol. 8, pp. 203317-203330, 2020.

[71] A.I.J. Forrester and A.J. Keane, "Recent advances in surrogate-based optimization," *Prog. Aerospace Sci.*, vol. 45, pp. 50-79, 2009.

[72] B. Beachkofski and R. Grandhi, "Improved distributed hypercube sampling," *American Institute of Aeronautics and Astronautics*, paper AIAA 2002-1274, 2002.

[73] S. Koziel and A. Pietrenko-Dabrowska, "Design-oriented modeling of antenna structures by means of two-level kriging with explicit dimensionality reduction," *AEU Int. J. Electronics Comm.*, vol. 127, pp. 1-12, 2020.

[74] Y.-C. Chen, S.-Y. Chen, and P. Hsu, "Dual-band slot dipole antenna fed by a coplanar waveguide," *IEEE Int. Symp. Ant. Prop.*, pp. 3589-3592, 2006.

[75] Z. Hua, G. Haichuan, L. Hongmei, L. Beijia, L. Guanjun, and W. Qun, "A novel high-gain quasi-Yagi antenna with a parabolic reflector," *Int. Symp. Ant. Prop. (ISAP)*, Hobart, Australia, 2015.