Variable-Fidelity Modeling of Antenna Input Characteristics Using Domain Confinement and Two-Stage Gaussian Process Regression Surrogates

J. Pieter Jacobs[*] and Slawomir Koziel[#,$]

[*]Centre for Electromagnetism, Department of Electrical, Electronic and Computer Engineering, University of Pretoria, Pretoria, 0002, South Africa
jpjacobs@postino.up.ac.za

[#]Engineering Optimization & Modeling Center, Department of Applied Engineering, Reykjavik University, Menntavegur 1, 101 Reykjavik, Iceland
koziel@ru.is

[$]Faculty of Electronics, Telecommunications and Informatics
Gdansk University of Technology, Poland

Keywords: Antenna design, surrogate modeling, data-driven models, Gaussian process regression, variable-fidelity simulations.

Abstract

The major bottleneck of electromagnetic (EM)-driven antenna design is the high CPU cost of massive simulations required by parametric optimization, uncertainty quantification, or robust design procedures. Fast surrogate models may be employed to mitigate this issue to a certain extent. Unfortunately, the curse of dimensionality is a serious limiting factor, hindering the construction of conventional data-driven models valid over wide ranges of the antenna parameters and operating conditions. This paper proposes a novel surrogate modeling approach that capitalizes on two recently proposed frameworks: the nested kriging approach and two-stage Gaussian process regression (GPR). In our methodology, the first-level surrogate of nested kriging is applied to define the confined domain of the model in which the final surrogate is constructed using two-stage GPR. The latter permits blending information from a sparsely-sampled high-fidelity EM simulation model and a densely-sampled low-fidelity (or coarse-mesh) model. This combination enables significant computational savings in terms of training data acquisition while retaining excellent predictive power of the surrogate. At the same time, the proposed framework inherits all the benefits of nested kriging, including ease of uniform sampling of the confined domain as well as straightforward generation of a good initial design for surrogate model optimization. Comprehensive benchmarking carried out using two antenna examples demonstrates superiority of our technique over conventional surrogates (unconfined domain), and standard GPR applied to the confined domain. Application examples for antenna optimization are also provided.

1. Introduction

Design of modern antennas has become a challenging endeavor with respect to almost every aspect of the process. On one hand, the conceptual development requires taking into account multiple performance specifications imposed on the electrical and field characteristics, implementation of various functionalities (broadband [1] or multi-band [2] operation, circular polarization [3], polarization/pattern diversity [4]), but also maintaining a small physical size of the antenna [5], [6]. Handling these requirements normally leads to topologically complex structures parameterized by a large number of variables and requiring full-wave electromagnetic (EM) analysis for their reliable evaluation. On the other hand, a precise parameter tuning has become mandatory to boost the performance as much as possible [7] and to find a usable trade-off between the conflicting design goals [8], [9]. Having in mind its complexity (multiple objectives and constraints and a typically highly dimensional parameter space), the design closure is nowadays often carried out through rigorous numerical optimization, as an alternative to still widespread parameter sweeping.

Clearly, non-negligible cost of individual EM analysis becomes a serious issue when performing EM-driven design tasks that require massive simulations, notably parametric optimization (both local [10] and global [11]), uncertainty quantification [12], or tolerance-aware design [13]. The necessity of extending the computational domain through incorporation of connectors, radomes, feeding structures, or other (coupled) radiators, only aggravates the problem. A number of techniques have been developed to improve the computational efficiency of simulation-based procedures, including adjoints sensitivities [14], [15], gradient-based search with sparse sensitivity updates [16]-[18], machine learning

methods [19], [20], as well as surrogate-assisted frameworks (space mapping [21], [22], response correction techniques [23], [24], feature-based optimization [25]).

Utilization of fast surrogate models as replacements for expensive EM analysis is a potential way of reducing the computational burden whenever numerous simulations are required. By far, the most popular class of surrogates are approximation models constructed exclusively from sampled high-fidelity (EM simulation) data. Some widely used techniques include polynomial regression [26], kriging [27], radial basis function interpolation [28], Gaussian process regression (GPR) [29], neural networks [30], polynomial chaos expansion [31], and support vector regression [32]. All of these methods are affected by the curse of dimensionality, i.e., a rapid growth of the number of training data samples required to render a usable model as a function of the number of system parameters and their ranges. In the case of antennas, typically characterized by highly nonlinear responses, conventional surrogates can be constructed for structures described by a few parameters and within narrow ranges thereof. Clearly, this is largely insufficient for creating design-ready models of contemporary antennas. The methods such as high-dimensional model representation (HDMR) [33] or orthogonal matching pursuit (OMP) [34] may alleviate these difficulties to certain extent but are not applicable for general-purpose modeling of nonlinear antenna characteristics.

In [35] an alternative approach to handling dimensionality issues has been proposed with the surrogate model established within a small region of the parameter space, determined by a set of reference designs pre-optimized with respect to the performance figures pertinent to the structure at hand. Due to the confined domain being dramatically smaller than the conventional space (delimited by the lower/upper bounds for the

parameters), a reliable surrogate could be constructed without formally restricting neither the dimensionality nor the parameter ranges [35]. This concept has been further developed in [36] to permit an arbitrary number of performance figures, and in the nested kriging method of [37] that enables arbitrary allocation of the reference designs, uniform sampling, and straightforward surrogate model optimization.

This paper proposes an alternative surrogate modeling approach, which falls into the category of performance-driven modeling but offers further advantages by incorporating variable-fidelity EM simulations into the nested kriging framework of [37]. Handling of variable-fidelity models is realized using two-stage GPR [38]. Under the first stage of two-stage GPR, a full-wave simulator is used to generate a low-fidelity (coarse) training data set of m points, and $m_{aux} << m$ points of the corresponding (computationally expensive) high-fidelity (fine) training set. A model is then trained that maps low-fidelity training targets (e.g., $\mathrm{Re}\{S_{11}\}$), along with the design vector and frequency, to the corresponding high-fidelity ones. This model is then used to predict the remaining $m - m_{aux}$ high-fidelity targets that were not simulated. The m_{aux} simulated high-fidelity targets and the $m - m_{aux}$ predicted ones – together with the input vectors–yield the m-point "approximate" high-fidelity training set. The second stage entails the construction of a final (i.e., main) GPR surrogate model using the latter training set. In this work, both first- and second-stage models are constructed within the confined domain rendered by the nested kriging framework.

The proposed methodology brings in some important technical novelties and contributions: (i) enhancing performance-driven modeling by employment of variable-fidelity EM simulation models, (ii) development of a rigorous modeling framework that combines nested kriging with two-stage GPR, (iii) further reduction of the computational

cost of surrogate model construction beyond what was offered so far in the literature, (iv) demonstration of the applicability of the proposed models for solving EM-based design tasks, specifically, parametric optimization. The presented framework is illustrated using two microstrip antennas: in each case, significant computational savings in terms of (equivalent) high-fidelity model evaluations is demonstrated, while still covering wide ranges of operating conditions. Benchmarking against conventional (unconstrained) GPR as well as single-fidelity nested GPR is also provided.

2. Nested Kriging Modeling

This section gives a brief overview of the nested kriging modeling framework, primarily utilized here to establish a confined domain for the two-stage GPR surrogate.

2.1. Parameter Space and Objective Space. Design Optimality

The modeling process involves two spaces. The first is the antenna (geometry) parameter space X defined as an interval $[\boldsymbol{l}, \boldsymbol{u}]$, where $\boldsymbol{l} = [l_1 \ldots l_n]^T$ and $\boldsymbol{u} = [u_1 \ldots u_n]^T$ are the lower and upper bounds, respectively. The parameter vectors are denoted as $\boldsymbol{x} = [x_1 \ldots x_n]^T$. The second space is the objective (or operating condition space) F consisting of vectors $\boldsymbol{f} = [f_1 \ldots f_N]^T$, with f_k being the figures of interest pertinent to the design process (e.g., operating frequencies of a multi-band antenna). The space F is also an interval defined by $f_{k.\min} \le f_k \le f_{k.\max}$, $k = 1, \ldots, N$.

The design optimality with respect to the objective vector $\boldsymbol{f}$, denoted as $U_f(\boldsymbol{f})$, is understood as the solution to the minimization problem [37]

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

where U is a scalar objective function encoding the performance requirements.

According to the performance-driven modeling approach [35]-[37], the surrogate model is to be established in the vicinity of the N-dimensional manifold $U_f(F)$ that contains the optimum designs for all $\boldsymbol{f} \in F$. Operating outside $U_f(F)$, in particular, allocating the training data and constructing the surrogate model, is a waste of resources as these regions only contain uninteresting designs from the point of view of the considered performance figures.

2.2. Reference Designs. First-Level Surrogate

In the nested kriging method [37], the manifold $U_f(F)$ is approximated using the reference designs $\boldsymbol{x}^{(j)} \in U_f(F), j = 1, \ldots, p$, optimized w.r.t. selected vectors $\boldsymbol{f}^{(j)} = [f_1^{(j)} \ldots f_N^{(j)}]$ (also, using the notation of (1), we have $\boldsymbol{x}^{(j)} = U_f(\boldsymbol{f}^{(j)})$). In particular, the data set $\{\boldsymbol{f}^{(j)}, \boldsymbol{x}^{(j)}\}, j = 1, \ldots, p$, is employed to construct the first-level surrogate $\boldsymbol{s}_I(\boldsymbol{f})$ mapping the space F into the parameter space X. In [37] (and in the present work), $\boldsymbol{s}_I$ is is implemented using kriging [27]. The image $\boldsymbol{s}_I(F)$ of F provides the best approximation of $U_f(F)$ one can obtain using limited information contained in the reference designs.

2.3. Model Domain

The domain X_S of the surrogate model is established by an extending $\boldsymbol{s}_I(F) \subset X$ into its orthogonal directions. This is to ensure that the extended set contains $U_f(F)$ (or at least a majority of it). Given $\{\boldsymbol{v}_n^{(k)}(\boldsymbol{f})\}, k = 1, \ldots, n - N$, an orthonormal basis of vectors normal to $\boldsymbol{s}_I(F)$ at $\boldsymbol{f}$, the extension coefficients $\boldsymbol{a}(\boldsymbol{f}) = [a_1(\boldsymbol{f}) \ldots a_{n-N}(\boldsymbol{f})]^T$ are defined as

$$\boldsymbol{a}(\boldsymbol{f}) = \frac{D}{2}\left|\boldsymbol{d}_x \boldsymbol{V}_n(\boldsymbol{f})\right|^{\mathrm{T}} = \frac{D}{2}\left[\left|\boldsymbol{d}_x \boldsymbol{v}_n^{(1)}(\boldsymbol{f})\right| \ \ldots \ \left|\boldsymbol{d}_x \boldsymbol{v}_n^{(n-N)}(\boldsymbol{f})\right|\right]^{\mathrm{T}} \tag{2}$$

where $\boldsymbol{d}_x = \boldsymbol{x}_{\max} - \boldsymbol{x}_{\min}$ is the range of antenna parameter variation within $\boldsymbol{s}_I(F)$, with $\boldsymbol{x}_{\max} = \max\{\boldsymbol{x}^{(k)}, k = 1, \ldots, p\}$ and $\boldsymbol{x}_{\min} = \min\{\boldsymbol{x}^{(k)}, k = 1, \ldots, p\}$; D is a user-defined parameter

determining the domain "thickness". The matrix $\boldsymbol{V}_n$ is defined as $\boldsymbol{V}_n(\boldsymbol{f}) = [\boldsymbol{v}_n^{(1)}(\boldsymbol{f}) \ \ldots \ \boldsymbol{v}_n^{(n-N)}(\boldsymbol{f})]$.

The coefficients $a_k(\boldsymbol{f})$ determine the boundaries of the domain X_S, which is allocated between the manifolds M_- and M_+ with

$$M_{\pm} = \left\{ \boldsymbol{x} \in X : \boldsymbol{x} = \boldsymbol{s}_I(\boldsymbol{f}) \pm \sum_{k=1}^{n-N} a_k(\boldsymbol{f}) \boldsymbol{v}_n^{(k)}(\boldsymbol{f}) \right\} \tag{3}$$

Formally, the surrogate model domain is defined as

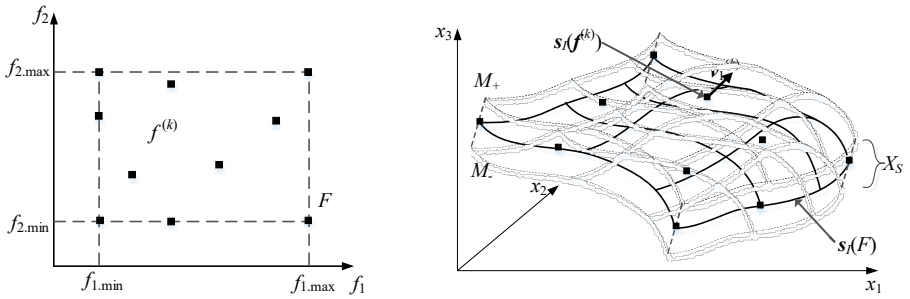
$$X_S = \left\{ \begin{array}{c} \boldsymbol{x} = \boldsymbol{s}_I(\boldsymbol{f}) + \sum_{k=1}^{n-N} \lambda_k a_k(\boldsymbol{f}) \boldsymbol{v}_n^{(k)}(\boldsymbol{f}) : \boldsymbol{f} \in F, \\ -1 \le \lambda_k \le 1, k = 1, \ldots, n-N \end{array} \right\} \tag{4}$$

i.e., it contains all points of the form $\boldsymbol{x} = \boldsymbol{s}_I(\boldsymbol{f}) + \sum_{k=1,\ldots,n-N} \lambda_k a_k(\boldsymbol{f}) \boldsymbol{v}_n^{(k)}(\boldsymbol{f})$ with $\boldsymbol{f} \in F$ and $-1 \le \lambda_k \le 1$, for $k = 1, \ldots, n - N$.

Figure 1 illustrates the objective space F, the allocation of the reference designs, as well as the defining manifolds $\boldsymbol{s}_I(F)$, M_- and M_+. The thickness parameter D determines the lateral size of the surrogate model domain in relation to the tangential size of $\boldsymbol{s}_I(F)$.

2.4. Second-Level Surrogate

In [37], the second-level (or the actual) surrogate was a kriging model established in the domain X_S. The procedure for uniform allocation of the training data samples $\{\boldsymbol{x}_B^{(k)}, \boldsymbol{R}(\boldsymbol{x}_B^{(k)})\}_{k=1,\ldots,NB}$ ($\boldsymbol{R}$ being the EM-simulation model of the antenna), based on the surjective mapping between a uniform interval $[0,1]^n$, can be found in [37]. In this work, the surrogate is set up using variable-fidelity models and two-stage GPR as described in the next section.



(a) (b)

Fig. 1. Surrogate model definition according to the nested kriging approach [37] (here, using two-dimensional objective space F and three-dimensional variable space X): (a) reference designs and objective space F; (b) the image $\boldsymbol{s}_I(F)$ of the first-level surrogate model and the normal vector $\boldsymbol{v}_1^{(k)}$ at $\boldsymbol{f}^{(k)}$; the manifolds M_- and M_+ as well as the surrogate model domain X_S defined as the orthogonal extension of $\boldsymbol{s}_I(F)$.

3. Two-Stage Gaussian Process Regression (GPR) Models

3.1. Standard GPR

A Gaussian process is a mathematical set comprised of an infinite number of random variables; any subset of these is jointly Gaussian distributed as well [39]. A Gaussian process $g(\boldsymbol{w})$ can be defined as $g(\boldsymbol{w}) \sim GP(r(\boldsymbol{w}), \mathrm{cov}(\boldsymbol{w}, \boldsymbol{w}'))$, with $\boldsymbol{w}$ and $\boldsymbol{w}'$ any pair of inputs in R^P space; the mean function $r(\boldsymbol{w})$ and covariance function $\mathrm{cov}(\boldsymbol{w},\boldsymbol{w}')$ are given by $r(\boldsymbol{w})=E[g(\boldsymbol{w})]$ and $\mathrm{cov}(\boldsymbol{w},\boldsymbol{w}') = E[g(\boldsymbol{w})-r(\boldsymbol{w}))(g(\boldsymbol{w}')-r(\boldsymbol{w}'))]$, where $E[Y]$ denotes the expectation of the random variable Y.

It follows that any collection of m function values $\boldsymbol{g} = [g_1 \ldots g_m]^T = [g(\boldsymbol{w}_1) \ldots g(\boldsymbol{w}_m)]^T$ has a jointly Gaussian distribution with mean vector $[r(\boldsymbol{w}_1) \ldots r(\boldsymbol{w}_m)]^T$ and an $m \times m$ covariance matrix with entries $K_{ij} = \mathrm{cov}(\boldsymbol{w}_i,\boldsymbol{w}_j)$, i.e. specified by the covariance function.

Suppose that a training data set of m noise-free function observations is available, namely $C = \{(\boldsymbol{w}_i, g_i) \mid i = 1, \ldots, m\}$, where the inputs $\boldsymbol{w}_i$ are vectors of dimension P, while the output observations/targets g_i are scalars. It is desired to predict the function values $\left[g_1^* \ldots g_{m^*}^*\right]^T = \left[g\left(\boldsymbol{w}_1^*\right) \ldots g\left(\boldsymbol{w}_{m^*}^*\right)\right]^T$, where $\boldsymbol{w}_1^* \ldots \boldsymbol{w}_{m^*}^*$ are test input vectors. To make

predictions, a jointly Gaussian distribution of zero mean is assumed over the m training outputs (vector $\boldsymbol{g}$), and the m* unknown test outputs (vector $\boldsymbol{g}$*) (the prior distribution):

$$\begin{bmatrix} \boldsymbol{g} \\ \boldsymbol{g}* \end{bmatrix} \sim N\left(\mathbf{0}, \begin{bmatrix} K(W,W) & K(W,W*) \\ K(W*,W) & K(W*,W*) \end{bmatrix} \right) \tag{5}$$

Here, $K(W,W*)$ is the $m \times m*$ matrix of covariances evaluated between all possible pairs of m training and m* test outputs, where W and W* are matrices containing the training and test input vectors respectively (other sub-matrices are defined in a similar manner); $N(\boldsymbol{z}, H)$ denotes a multivariate Gaussian distribution with mean vector $\boldsymbol{z}$ and covariance matrix H.

The posterior distribution, i.e. involving the test outputs conditioned on the known training outputs $\boldsymbol{g}$, can then be expressed as $\boldsymbol{g}*|W*,W, \boldsymbol{g} \sim N(\boldsymbol{r}, \Sigma)$, with mean vector $\boldsymbol{r}$ and covariance matrix Σ, given by $\boldsymbol{r} = K(W*,W)K(W,W)^{-1}\boldsymbol{g}$ and $\Sigma = K(W*,W*) - K(W*,W)K(W,W)^{-1}K(W,W*)$ respectively [39]. The predictive mean $\boldsymbol{r}$ contains the most likely values of the test outputs associated with the test input vectors in W*, while the diagonal of the covariance matrix Σ gives the corresponding predictive variances. Prior to conditioning, the hyperparameters of the covariance function are optimised by minimising the negative log marginal likelihood with respect to the hyperparameters [39, eq. (2.29)].

3.2. Two-stage GPR

1) First stage: Consider a computationally expensive finely-discretized training data set $C_{fine} = \{(\boldsymbol{w}_i, y_{fine,i}) \mid i = 1,...,m\}$ with input vectors

$$\boldsymbol{w}_i = [\boldsymbol{x}_i^T \; q_{oi}]^T = [x_{1i} \; x_{2i} \ldots x_{Mi} \; q_{oi}]^T \tag{6}$$

of dimension $M+1$, and scalar targets $y_{fine,i} = \mathrm{Re}\{S_{11}\}_{fine,i}$. The design vector $\boldsymbol{x}_i = [x_{1i} \, x_{2i} \ldots x_{Mi}]^T$ consists of M adjustable antenna design variables, while q_{oi} is a frequency value within the band of interest (the target values could also be $\mathrm{Im}\{S_{11}\}$ or $|S_{11}|$; for the sake of

brevity we refer to $\mathrm{Re}\{S_{11}\}$ below). In general, obtaining C_{fine} via EM simulation is assumed to be computationally prohibitively expensive.

The purpose of the first stage is to "approximate" C_{fine} by an inexpensive data set $C_{fine,approx}$ of the same size. This can be accomplished by using a separate auxiliary model $\boldsymbol{R}_{aux}$ trained on a specially constructed training data set C_{aux}. In order to construct C_{aux}, the above m geometries are first simulated cheaply using a coarse mesh, yielding the data set $C_{coarse}=\{(\boldsymbol{w}_i, y_{coarse,i}) \mid i=1,...,m\}$, with $\boldsymbol{w}_i$ as for C_{fine}, and $y_{\mathrm{coarse},i} = \mathrm{Re}\{S_{11}\}_{\mathrm{coarse},i}$. Next, a small randomly selected subset of C_{fine}, consisting of $m_{aux} < m$ points, is simulated (i.e., at high fidelity). Using the latter subset of C_{fine} as well as C_{course}, the training set C_{aux} can then be compiled as

$$C_{aux} = \left\{ (\boldsymbol{w}_{aux,k}, y_{fine,k}) \mid k = 1, ..., m_{aux} \right\} \tag{7}$$

where the $(M+2)$-dimensional training input vector

$$\boldsymbol{w}_{aux,k} = [x_{1k} \;\; x_{2k} \ldots x_{Mk} \;\; q_{ok} \;\; \mathrm{Re}\{S_{11}\}_{coarse,k}]^T \tag{8}$$

is of the form of the input vectors of C_{fine} augmented by the associated coarse $\mathrm{Re}\{S_{11}\}$ target value from C_{coarse}. The target $y_{fine,k}$ is the corresponding $\mathrm{Re}\{S_{11}\}$ value from the above small subset of C_{fine} (note that C_{coarse} and C_{fine} share the same input vectors; the difference lies in the meshing density with which the targets have been simulated). Hence $\boldsymbol{R}_{aux}$ learns a mapping between coarse and fine $\mathrm{Re}\{S_{11}\}$ simulations using training data that correspond to m_{aux} specific instances of sets of design variables and frequency (the initial $M+1$ components of $\boldsymbol{w}_{aux,k}$ uniquely identifies the $\mathrm{Re}\{S_{11}\}$ values). (This mapping embodies the correlations between the coarse and fine model responses.)

After training, $\boldsymbol{R}_{aux}$ is used to predict the $m - m_{aux}$ fine $\mathrm{Re}\{S_{11}\}$ values that were not simulated, from their coarsely simulated counterparts. The predicted targets are denoted by

$y_{pred,k} = \mathrm{Re}\{S_{11}\}_{pred,k}$, $k = (m_{aux}+1), ..., m$. The m_{aux} full-wave-simulated fine $\mathrm{Re}\{S_{11}\}$ target values and the $m - m_{aux}$ predicted ones then yield (along with input vectors consisting of geometry parameters and frequency of the form (6)) an m-point "approximate" fine training data set for $\boldsymbol{R}_s$ (the second-stage, i.e. final model):

$$C_{fine,approx} = \left\{ \begin{array}{c} (\boldsymbol{w}_k, y_{fine,k}) \mid k = 1, ..., m_{aux} \\ (\boldsymbol{w}_k, y_{pred,k}) \mid k = (m_{aux}+1), ..., m \end{array} \right\} \tag{8}$$

The substantial savings in computational costs that can be achieved within the nested kriging framework by obtaining the targets $y_{pred,k}$ via model predictions—as opposed to direct full-wave simulations—are described below.

2) Second stage: Using standard GPR, $C_{fine,approx}$ is used to train $\boldsymbol{R}_s$, the final surrogate that maps design variables and frequency to $\mathrm{Re}\{S_{11}\}$.

3. Proposed Modeling Framework: Variable-Fidelity Nested Kriging-GPR Surrogates

This section summarizes the proposed methodology. The modeling process, as explained using the flowchart shown in Fig. 2, starts by defining the surrogate model domain based on the nested kriging method of [37] (cf. Section 2). The reference designs involved at this stage may be already available from the previous design work with the same structure or obtained specifically for building the surrogate.

The actual surrogate is constructed by blending the low- and high-fidelity EM simulation data using the two-stage GPR procedure of [38] (cf. Section 3). The advantage brought by incorporating variable-fidelity computational models is a further reduction of the CPU cost of setting up the surrogate as compared to [38]. This is demonstrated in the next section.

4. Verification Examples and Benchmarking

The proposed modeling technique is validated using two antenna structures: a dual-band uniplanar dipole antenna (Antenna I) shown in Fig. 3(a) [40], and a ring slot antenna (Antenna II) shown in Fig. 3(b) [35]. Antenna I is implemented on a RO4350 substrate (ε_r = 3.5, h = 0.76 mm) and fed using a 50 ohm coplanar waveguide (CPW). The design variables are: $\boldsymbol{x} = [l_1\ l_2\ l_3\ w_1\ w_2\ w_3]^T$; l_0 = 30, w_0 = 3, s_0 = 0.15 and o = 5 are fixed (all dimensions in mm).

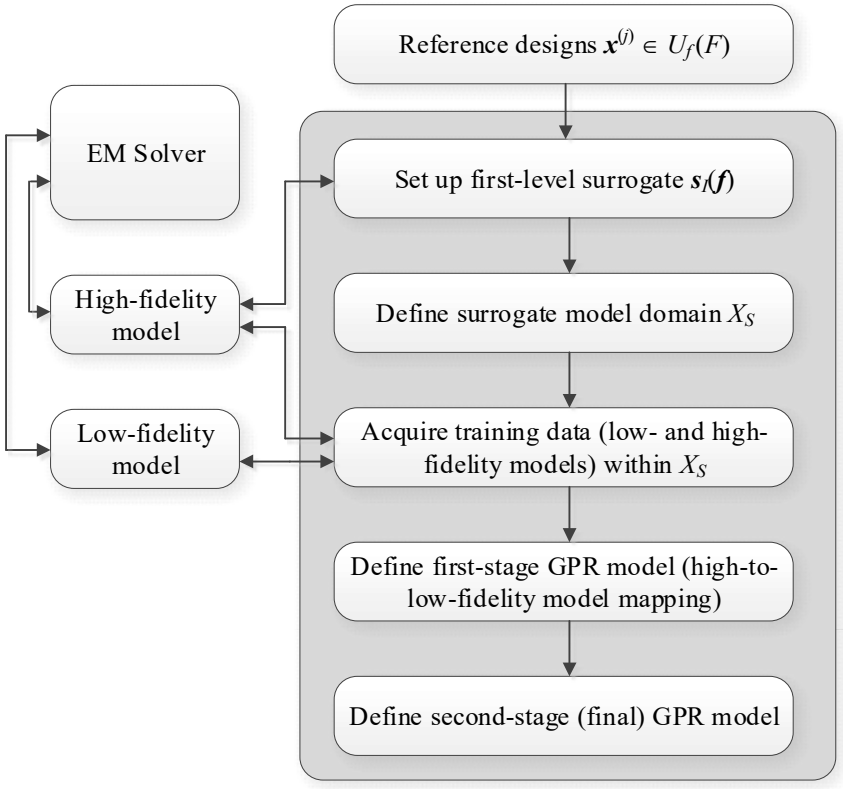


Fig. 2. Flowchart of the proposed surrogate modeling procedure. The surrogate model domain is defined according to the nested kriging modeling of [37] using a set of reference designs (cf. Sections 2.2 and 2.3). The two-stage GPR surrogate involving sparsely sampled high-fidelity EM model and densely sampled low-fidelity model is rendered in the domain X_S as described in Section 3.

Antenna II is implemented on 0.76-mm-thick substrate and described by eight parameters $\boldsymbol{x} = [l_f\ l_d\ w_d\ r\ s\ s_d\ o\ g]^T$. The relative permittivity ε_r of the substrate is an additional variable for the modeling process. The feed line width w_f is calculated for any given ε_r to ensure 50 ohm input impedance.

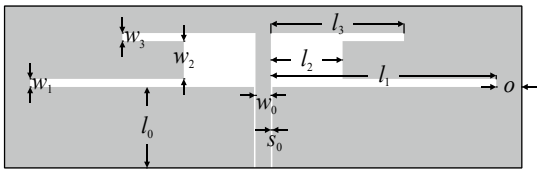
The computational models are implemented in CST Microwave Studio. We have, for Antenna I:

- High-fidelity model $\boldsymbol{R}_f$: ~100,000 mesh cells, simulation time 60 s;
- Low-fidelity model $\boldsymbol{R}_c$: ~40,000 mesh cells, 20 s;

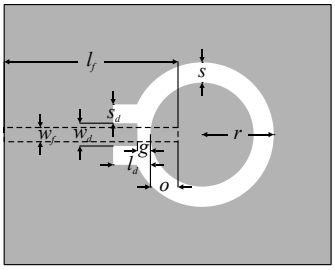
and for Antenna II:

- High-fidelity model $\boldsymbol{R}_f$: ~300,000 mesh cells, 90 s;
- Low-fidelity model $\boldsymbol{R}_c$: ~90,000 mesh cells, 22 s.

Note that the low-fidelity model on the whole is three to four times faster than the high-fidelity one.



(a)



(b)

Fig. 3. Antenna geometries: (a) dual-band uniplanar dipole antenna (Antenna I) [39], (b) ring slot antenna (Antenna II) [35].

For Antenna I, the modeling objective is to construct a surrogate valid for the following ranges of operating frequencies: 2.0 GHz $\leq f_1 \leq$ 3.0 GHz (lower band), and 4.0 GHz $\leq f_2 \leq$ 5.5 GHz (upper band). For Antenna II, the goal is to construct a surrogate model for the operating frequencies f within the range 2.5 GHz $\leq f \leq$ 6.5 GHz, and substrate permittivity ε_r within the range of $2.0 \leq \varepsilon_r \leq 5.0$. The allocation of the reference designs and the lower/upper bounds determining the parameter space can be found in [37].

The surrogate model domains have been established according to the nested kriging approach (cf. Section 2.1 to 2.3) using the thickness parameters $D = 0.05$ and $D = 0.1$. Subsequently the two-stage GPR surrogates were constructed using the following training sets:

- Antenna I: 200 antenna geometries, each coarsely evaluated at 10 random frequencies ($m = 10$); one of the 10 frequencies was also simulated at high fidelity ($m_{aux} = 1$; cf. Eq. (7)); $m_{aux}/m = 10\%$.
- Antenna II: 400 antenna geometries, each coarsely evaluated at 12 random frequencies ($m = 12$), with one of the 12 frequencies ($m_{aux} = 1$) also simulated at high fidelity; $m_{aux}/m = 8.3\%$.

Given the time evaluation ratios between the high- and low-fidelity EM models, the average number of equivalent high-fidelity simulations is 867 for Antenna I (note that Antenna I required 10 low-fidelity simulations and one high-fidelity simulation per geometry, that there were 200 geometries, and that the low-fidelity model was three times faster than the high-fidelity model – hence $200\times(1+10/3) \cong 867$ equivalent high-fidelity simulations). For Antenna II, the number of equivalent high-fidelity simulations is 1600.

Two-stage GPR models were constructed separately for the real and imaginary parts of the reflection response: because of their smoothness, these responses are more naturally modelled by GPR than $|S_{11}|$. Predictive results of the two models were then combined to yield $|S_{11}|$. Throughout, the Matérn covariance function with $\nu = 5/2$ was used [39].

For comparison purposes, standard (i.e. not two-stage) GPR models were trained on the above two-stage surrogate model (confined) domains; training data were comprised of the same geometries/frequencies as in the case of the two-stage models, but simulated at high fidelity only (e.g., 200 geometries × 10 frequencies for Antenna I). Standard GPR models were also trained on the original, unconfined domains – this constitutes conventional GPR (as noted above, the bounds delimiting the original design space can be found in [34]).

The numerical results have been gathered in Table 1 (predictive RMS errors are normalized to the ranges of the test target values for each antenna [38]). These are compared to conventional surrogates (i.e. where the domains are not confined), as well as the nested kriging model [37] but with GPR (instead of kriging) used as second-level surrogate (referred to as nested GPR in Table 1). It can be observed that the variable-fidelity GPR surrogates exhibit similar levels of predictive performance to the conventional nested-GPR surrogates, but at greatly reduced cost: in the case of Antenna I, the savings in terms of high-fidelity simulations is 57%; in the case of Antenna II, it is 67%.

It can be expected that the computational benefits of the confined two-stage GPR would be even more pronounced if the low-fidelity model was faster as compared to the high-fidelity one. In our examples, the high-to-low-fidelity time evaluation ratio was only three for Antenna I and four for Antenna II.

Figure 4 shows the antenna responses for some selected test designs. The visual agreement between the surrogate and EM-simulated characteristics can be considered satisfactory for practical purposes.

Table 1. Predictive results for $|S_{11}|$ for Antennas I and II

Antenna	Number of high-fidelity training samples[$]	Normalized RMS Error			Number of equivalent high-fidelity training samples[$]	Normalized RMS Error	
		Conventional (unconfined) GPR	Nested GPR Model			Nested two-stage GPR Model	
			$D = 0.05$	$D = 0.1$		$D = 0.05$	$D = 0.1$
I	2000	9.97%	3.74 %	4.95 %	867	5.05%	6.19%
II	4800	18.01%	2.91%	4.5%	1600	4.27%	5.60%

[$] Number of high-fidelity training samples calculated as the number of antenna geometries multiplied by the number of randomly selected frequencies per geometry.
[#]Number of equivalent high-fidelity training samples calculated as the number of geometries multiplied by the number of equivalent high-fidelity simulations used per geometry (see text) to set up the surrogate.

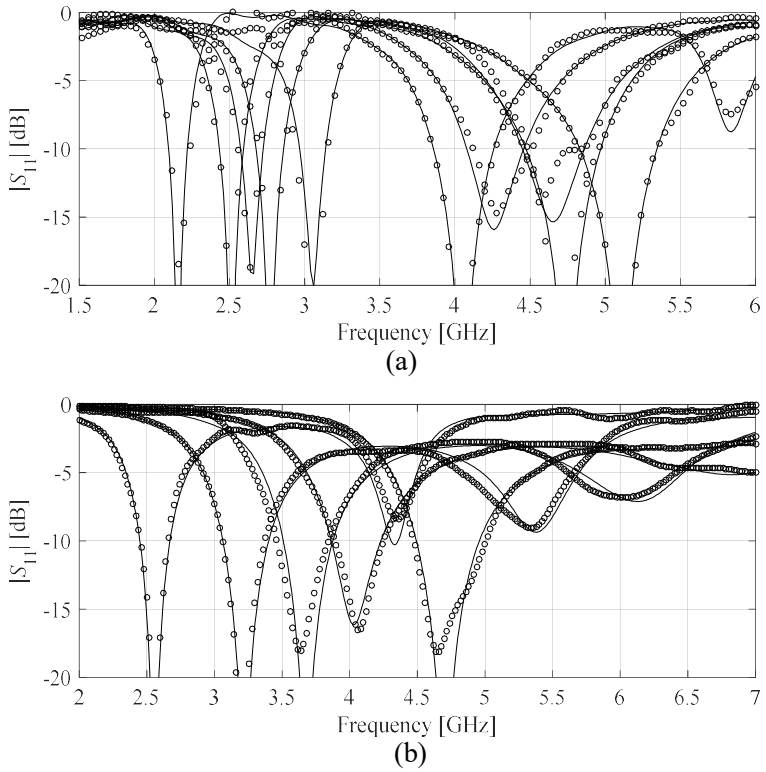


Fig. 4. Responses of the antennas of Fig. 3 at the selected test designs: EM model (—), confined two-stage GPR surrogate (o): (a) Antenna I, (b) Antenna II.

The two-stage GPR surrogates (obtained for $D = 0.05$) were also used for antenna optimization. Figure 5 shows the results obtained for Antenna I for the following pairs of target operating frequencies: {2.2,4.5}, {3.0,5.0}, {2.45,4.8}, and {2.7,5.3} (frequencies in GHz). Figure 6 shows the results for Antenna II for four pairs of operating frequency and substrate dielectric constant (frequency in GHz): {4.8,2.2}, {5.3,3.5}, {2.7,2.5}, and {4.8,4.4}. It can be observed that the two-stage GPR models yields good results which are in agreement with the EM simulations. This confirms usability of the models for design purposes.

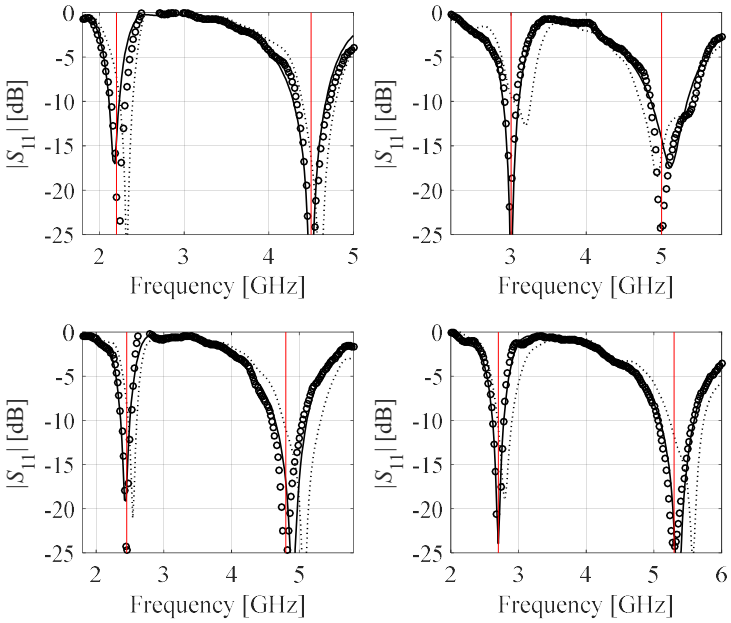


Fig. 5. Optimization results of Antenna I using the two-stage GPR surrogate for the four pairs of operating frequencies, from top-left to bottom-right: {2.2,4.5}, {3.0,5.0}, {2.45,4.8}, and {2.7,5.3} (frequencies in GHz). Shown are: initial design obtained from the first-level kriging model (····), response of the optimized two-stage GPR surrogate (o), EM-simulated antenna response at the GPR model optimum (—).

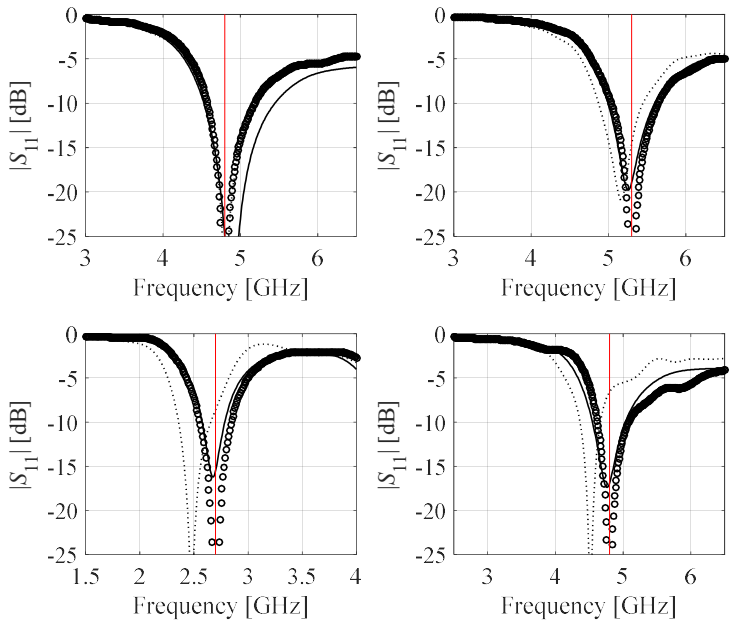


Fig. 6. Optimization results of Antenna II using the two-stage GPR surrogate for the four pairs of operating frequency and substrate permittivity, from top-left to bottom-right: {4.8,2.2}, {5.3,3.5}, {2.7,2.5}, and {4.8,4.4} (frequencies in GHz). Shown are: initial design obtained from the first-level kriging model (····), response of the optimized two-stage GPR surrogate (o), EM-simulated antenna response at the GPR model optimum (—).

5. Conclusion

The paper proposed a novel technique for surrogate modeling of antenna input characteristics. It incorporates variable-fidelity EM simulations into the recently reported nested kriging framework by means of two-stage GPR. Our methodology has been validated using two planar antennas and benchmarked against conventional surrogates as well as high-fidelity-only nested GPR. The numerical results indicate that considerable computational savings can be obtained despite the fact that the time evaluation ratio between the high- and low-fidelity models is quite limited in the considered cases. At the same time, the surrogates produced using the proposed approach can be effectively

employed for design purposes as comprehensively demonstrated through parametric optimization of the respective antenna structures.

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 206606051 and by National Science Centre of Poland Grant 2018/31/B/ST7/02369.

References

[1] H. Huang, Y. Liu, and S. Gong, "A broadband dual-polarized base station antenna with sturdy construction," *IEEE Ant. Wireless Prop. Lett.*, vol. 16, pp. 665-668, 2017.

[2] C. Goswami, R. Ghatak, and D.R. Poddar, „Multi-band bisected Hilbert monopole antenna loaded with multiple subwavelength split-ring resonators," *IET Microwaves Ant. Prop.*, vol. 12, no. 10, pp. 1719-1727, 2018.

[3] D. Feng, H. Zhai, L. Xi, S. Yang, K. Zhang, and D. Yang, "A broadband low-profile circular polarization antenna on an AMC reflector," *IEEE Ant. Wireless Prop. Lett.*, vol. 16 pp. 2840-2843, 2017.

[4] G. Wolosinski, V. Fusco, U. Naeem, and P. Rulikowski, "Pre-matched eigenmode antenna with polarization and pattern diversity," *IEEE Trans. Ant. Prop.*, vol. 67, no. 8, pp. 5145-5153, 2019.

[5] A.A. Omar and Z. Shen, "A compact and wideband vertically polarized monopole antenna," *IEEE Trans. Ant. Prop.*, vol. 67, no. 1, pp. 626-631, 2019.

[6] J. Wu and K. Sarabandi, "Compact omnidirectional circularly polarized antenna," *IEEE Trans. Ant. Prop.*, vol. 65, no. 4, pp. 1550-1557, 2017.

[7] N. Nguyen-Trong and C. Fumeaux, "Tuning range and efficiency optimization of a frequency-reconfigurable patch antenna," *IEEE Ant. Wireless Prop. Lett.*, vol. 17, no. 1, pp. 150-154, 2018.

[8] S. Koziel and A. Bekasiewicz, "Multi-objective design of antennas using surrogate models," World Scientific, 2016.

[9] A. Lakbakhsh, M.U. Afzal, and K.P. Esselle, "Multiobjective particle swarm optimization to design a time-delay equalizer metasurface for an electromagnetic band-gap resonator antenna," *IEEE Ant. Wireless Prop. Lett.*, vol. 16, pp. 915-915, 2017.

[10] J. Nocedal and S. Wright, *Numerical Optimization*, 2nd edition, Springer, New York, 2006.

[11] A. Darvish and A. Ebrahimzadeh, "Improved fruit-fly optimization algorithm and its applications in antenna array synthesis," *IEEE Trans. Ant. Prop.*, vol. 66, no. 4, pp. 1756-1766, 2018.

[12] A.S.O. Hassan, H.L. Abdel-Malek, A.S.A. Mohamed, T.M. Abuelfadl, and A.E. Elqenawy, "Statistical design centering of RF cavity linear accelerator via non-derivative trust region optimization," *IEEE Int. Conf. Numerical EM Multiphysics Modeling Opt. (NEMO)*, pp. 1-3, 2015.

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

[14] E. Hassan, D. Noreland, R. Augustine, E. Wadbro, and M. Berggren, "Topology optimization of planar antennas for wideband near-field coupling," *IEEE Trans. Ant. Prop.*, vol. 63, no. 9, pp. 4208-4213, 2015.

[15] S. Koziel and A. Bekasiewicz, "Rapid design optimization of antennas using variable-fidelity EM models and adjoint sensitivities," *Eng. Comp.*, vol. 33, no. 7, pp. 2007-2018, 2016.

[16] 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.

[17] A. Pietrenko-Dabrowska and S. Koziel, „Computationally-efficient design optimization of antennas by accelerated gradient search with sensitivity and design change monitoring," *IET Microwaves Ant. Prop.*, vol. 14, no. 2, pp. 165-170, 2020.

[18] J.A. Tomasson, S. Koziel, and A. Pietrenko-Dabrowska, "Quasi-global optimization of antenna structures using principal components and affine subspace-spanned surrogates," *IEEE Access*, vol. 8, no. 1, pp. 50078-50084, 2020.

[19] A.M. Alzahed, S.M. Mikki, and Y.M.M. Antar, "Nonlinera mutual coupling compensation operator design using a novel electromagnetic machine learning paradigm," *IEEE Ant. Wireless Prop. Lett.*, vol. 18, no. 5, pp. 861-865, 2019.

[20] J. Joung, "Machine learning-based antenna selection in wireless communications," *IEEE Comm. Lett.*, vol. 20, no. 11, pp. 2241-2244, 2016.

[21] I.A. Baratta, C.B. de Andrade, R.R. de Assis, and E.J. Silva, "Infinitesimal dipole model using space mapping optimization for antenna placement," *IEEE Ant. Wireless Prop. Lett.*, vol. 17, no. 1, pp. 17-20, 2018.

[22] S. Koziel, and A. Bekasiewicz, "Reliable low-cost surrogate modeling and design optimization of antennas using implicit space mapping with substrate segmentation," *IET Microwaves Ant. Prop.*, vol. 11, no. 14, pp. 2066-2070, 2017.

[23] S. Koziel and L. Leifsson, "Simulation-driven design by knowledge-based response correction techniques," Springer, 2016.

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

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

[26] J.L. Chávez-Hurtado and J.E. Rayas-Sánchez, "Polynomial-based surrogate modeling of RF and microwave circuits in frequency domain exploiting the multinomial theorem," *IEEE Trans. Microwave Theory Tech.*, vol. 64, no. 12, pp. 4371-4381, 2016.

[27] T.W. Simpson, J.D. Pelplinski, P.N. Koch, and J.K. Allen, "Metamodels for computer-based engineering design: survey and recommendations", *Engineering with Computers*, vol. 17, pp. 129-150, 2001.

[28] S. Mishra, R.N. Yadav, and R.P. Singh, "Directivity estimations for short dipole antenna arrays using radial basis function neural networks," *IEEE Ant. Wireless Prop. Lett.*, vol. 14, pp.

[29] 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.

[30] T.N. Kapetanakis, I.O. Vardiambasis, M.P. Ioannidou, and A. Maras, "Neural network modeling for the solution of the inverse loop antenna radiation problem," *IEEE Trans. Ant. Prop.*, vol. 66, no. 11, pp. 6283-6290, 2018.

[31] J. Zhang, C. Zhang, F. Feng, W. Zhang, J. Ma, and Q.J. Zhang, "Polynomial chaos-based approach to yield-driven EM optimization," *IEEE Trans. Microwave Theory Tech.*, vol. 66, no. 7, pp. 3186-3199, 2018.

[32] D.R. Prado, J.A. Lopez-Fernandez, M. Arrebola, and G. Goussetis, "Support vector regression to accelerate design and crosspolar optimization of shaped beam reflectarray for space applications," *IEEE Trans. Ant. Prop.*, vol. 67, no. 3, pp. 1659-1668, 2019.

[33] 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.

[34] X. Li, "Finding deterministic solution from underdetermined equation: large-scale performance modeling of analog/RF circuits," *IEEE Trans. on Computer-Aided Design of Integrated Circuits and Systems* (*TCAD*), vol. 29, no. 11, pp. 1661-1668, 2010.

[35] S. Koziel and A. Bekasiewicz, "On reduced-cost design-oriented constrained surrogate modeling of antenna structures," *IEEE Ant. Wireless Prop. Lett.*, vol. 16, pp. 1618-1621, 2017.

[36] S. Koziel and A.T. Sigurdsson, "Triangulation-based constrained surrogate modeling of antennas," *IEEE Trans. Ant. Prop.*, vol. 66, no. 8, pp. 4170-4179, 2017.

[37] 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.

[38] 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.

[39] C. E. Rasmussen and C. K. I. Williams, *Gaussian processes for machine learning*, MIT Press, Cambridge, MA, 2006.

[40] 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.