Two-Stage Framework for Efficient Gaussian Process Modeling of Antenna Input Characteristics

J. P. Jacobs and S. Koziel

***Abstract*—A two-stage approach based on Gaussian process regression that achieves significantly reduced requirements for computationally expensive high-fidelity training data is presented for the modeling of planar antenna input characteristics. Our method involves variable-fidelity electromagnetic simulations. In the first stage, a mapping between electromagnetic models (simulations) of low and high fidelity is learned, which allows us to substantially reduce (by 80% or more) the computational effort necessary to set up the high-fidelity training data sets for the actual surrogate models (second stage), with negligible loss in predictive power. We illustrate our method by modeling the input characteristics of three antenna structures with up to seven design variables. The accuracy of the two-stage method is confirmed by the successful use of the surrogates within a space-mapping-based optimization/design framework.**

***Index Terms*—Gaussian processes, microwave antennas, modeling, optimization.**

I. Introduction

CONTEMPORARY microwave engineering relies heavily on full-wave electromagnetic simulations as they permit highly accurate evaluation of microwave structures, including antennas. Unfortunately, high-fidelity simulations are computationally expensive. While this is not a problem for design verification, the use of accurate electromagnetic simulations to solve various design tasks involving numerous analyses, such as statistical analysis, yield-driven design, or parametric design optimization, might become infeasible under certain conditions. For example, global optimization using metaheuristics (e.g., genetic algorithms) [1], [2] might require thousands of full-wave analyses of candidate geometries of the structure to be optimized. In situations like these, the use of fast and yet accurate models of the microwave structures under consideration (so-called surrogates) become indispensable. Identified on a training set consisting of a limited number of input–output pairs, these models by virtue of their ability to generalize over the input space make it possible to quickly obtain the desired performance characteristics for inputs not previously presented to the model.

A powerful approach for constructing surrogate models of antenna structures is Gaussian process regression (GPR) [3]. GPR has been shown to be a particularly effective modeling technique for performance characteristics such as the input reflection coefficient response against frequency [4], [5]. (Other kernel-based machine learning methods that have been used in this context have included standard support vector regression, e.g., [6], and the more expressive, GPR-based Bayesian support vector regression [7].)

A Gaussian process is a stochastic process that entails the generalization of the Gaussian probability distribution to functions. The Gaussian nature of the distribution leads to tractable—relatively simple—calculations when learning and inference need to be performed. Under suitable conditions, Gaussian processes can be considered equivalent to neural networks, but Gaussian processes are generally easier to implement and interpret—one reason is that training of far fewer parameters (in the order of the dimension of the input vectors) is required compared to the number of weights in a neural network (e.g., multi-layer perceptron).

Probably the most serious limitation of approximation-based modeling methods, including GPR, is the high initial cost of gathering the fine-discretization data necessary to train the model so as to ensure that it has sufficient predictive accuracy. In the present study, we address this problem by using variable-fidelity electromagnetic simulations. More specifically, we use a two-stage modeling scheme: in the first stage, we generate by full-wave simulation a low-fidelity (coarse) training data set of n points, and $n_{\rm aux} < n$ points of the corresponding (computationally expensive) high-fidelity (fine) training set. We then learn a model that maps low-fidelity training targets ($\mathrm{Re}\{S_{11}\}$ or $\mathrm{Im}\{S_{11}\}$) to the high-fidelity ones, and use it to predict the remaining $n - n_{\rm aux}$ high-fidelity targets (i.e., that were not simulated). The $n_{\rm aux}$ simulated high-fidelity targets and the $n - n_{\rm aux}$ predicted ones (together with the input vectors) then yield the n-point "approximate" high-fidelity training set. In the second stage, a final GPR model is constructed using the latter training set. We show below that considerable reduction of the model setup cost without compromising its accuracy is possible by exploiting the knowledge embedded in the low-fidelity simulations in this manner.

Previously, optimal data selection for microwave modeling problems has been attempted through various adaptive sampling techniques that aim, within optimization contexts, to reduce the number of samples necessary to ensure the desired modeling accuracy by iterative identification of the model and adding new

This work was supported in part by the Icelandic Centre for Research (RANNIS) under Grants 120016021 and 13045051.

J. P. Jacobs is with the Centre for Electromagnetism, Department of Electrical, Electronic and Computer Engineering, University of Pretoria, Pretoria 0002, South Africa (e-mail: jpjacobs@postino.up.ac.za).

S. Koziel is with the School of Science and Engineering, Reykjavik University, Reykjavik 101, Iceland (e-mail: koziel@ru.is).

training samples based on the actual model error at selected locations (e.g., [8]) or expected error values (statistical infill criteria, e.g., [9]). [8], [9] are local/trust region models; in contrast, our focus is on global or library type models that give accurate predictions over the entire input space, and that can be used for a variety of applications (e.g., optimization, statistical analysis). Another approach to reducing the computational cost of surrogate model setup was presented in [10], where only the support vectors of an initial (global) BSVR model trained on low-fidelity data were selected for fine-discretization simulation. A 31-to-48 percent reduction in model setup costs could be achieved without compromising predictive capability.

The methodology described in this paper is novel in that it formally maps the correlations between physically-related coarse and fine simulation models of the same antenna via an auxiliary model, ultimately blending full-wave simulation data at two fidelity levels into one final surrogate model (via training on the above "approximate" high-fidelity data set). The essentially parameter-less nature of the model enhancement also constitutes novelty compared to earlier methods. Our method constitutes a fundamental advance over the antenna modeling of [4], [5]: here conventional GPR models were trained on data sets obtained in full from expensive high-fidelity data; no attempt was made to reduce the costs associated with acquiring this data, even though this contributed overwhelmingly to the model setup costs.

Our approach is demonstrated using three examples of antennas with highly nonlinear $|S_{11}|$ responses as a function of tunable geometry parameters and frequency: a narrowband coplanar waveguide (CPW)-fed slot dipole antenna with two design variables, an ultrawideband (UWB) CPW-fed T-shaped slot antenna with four design variables, and a dielectric resonator antenna with seven design variables. We furthermore evaluate the accuracy of our GPR surrogates by using them within a surrogate-based optimization framework involving output space mapping [11]–[13].

The paper is organized as follows. Section II provides a brief theoretical overview of GPR along the lines of [3], and describes the two-stage modeling approach. Section III provides comprehensive numerical verification of the proposed methodology. In particular, it describes how GPR models involving increasingly fewer high-fidelity simulations were set up for each of the above antennas, and gives their predictive accuracies. These results clearly indicate the computational advantages of the two-stage modeling approach, demonstrating that dramatic reduction of the CPU overhead related to model setup can be obtained without compromising predictive power—unlike the conventional GPR modeling of [4], [5] which required training data obtained in full through high-fidelity simulations in order to ensure accurate predictions. In Section IV, the models obtained in Section III are used as basis for antenna optimization using a space mapping algorithm [11]. Conclusions are presented in Section V.

II. Gaussian Process Regression Modeling

A. Fundamentals of Gaussian Process Regression

A Gaussian process describes a distribution over functions. It is a mathematical set consisting of an infinite number of random variables, of which any subset is jointly Gaussian. As such it is a natural extension of a jointly Gaussian distribution to the case where the mean vector is infinitely long and the covariance matrix is of infinite by infinite dimension. A GP can be notated as $f(\mathbf{u}) \sim GP(m(\mathbf{u}), k(\mathbf{u}, \mathbf{u}'))$, with $\mathbf{u}$ and $\mathbf{u}'$ positions in R^P space, and $m(\mathbf{u})$ and $k(\mathbf{u}, \mathbf{u}')$ its mean and covariance functions, respectively, (defined as [3, (2.13)]). The GP encapsulates all possible functions in the vast space of functions that subscribe to $m(\mathbf{u})$ and $k(\mathbf{u}, \mathbf{u}')$. The model is semi-parametric in the sense that any sample function is not specified in terms of a finite number of parameters (such as weights in the case of a linear model), but directly in the space of functions.

Consider for example a finite (practical) training data set of n observations, $D = \{(\mathbf{u}_i, y_i)|i = 1, \ldots, n\}$. The inputs $\mathbf{u}_i$ are column vectors of dimension P, while the output targets y_i are scalars. The corresponding Gaussian process $f(\mathbf{u})$ in this case would be implemented as the collection of random variables $f_i = f(\mathbf{u}_i)$, with any n-dimensional point under their jointly Gaussian distribution representing n values of a sample function with index set the set of inputs $\{\mathbf{u}_i\}$.

The only parameterization that takes place is the specification of hyperparameters which determine the properties of the mean and covariance functions. The present study uses the squared-exponential covariance function

$$k(\mathbf{u}, \mathbf{u}') = \sigma_f^2 \exp(-0.5(\mathbf{u} - \mathbf{u}')^T M(\mathbf{u} - \mathbf{u}')) \qquad (1)$$

which gives the covariance between the output random variables $f(\mathbf{u})$ and $f(\mathbf{u}')$. The matrix $M = \mathrm{diag}(\mathbf{L})^{-2}$, with $\mathbf{L}$ the vector of positive characteristic length-scale parameters corresponding to the elements of the input vectors, and σ_f^2 is the signal variance (length-scale parameters are indicative of how quickly change occurs along the corresponding dimensions of the input space). Together, $\mathbf{L}$ and σ_f^2 constitute the hyperparameters of the covariance function. The hyperparameters may be found through a structured methodology which involves a process similar to Bayesian model selection. It entails finding the hyperparameters for which the negative log marginal likelihood is a minimum. The log marginal likelihood in the noise-free case is given by [3]

$$\log p(\mathbf{y}|X) = -\frac{1}{2}\mathbf{y}^T K^{-1}\mathbf{y} - \frac{1}{2}\log|K| - \frac{n}{2}\log 2\pi. \qquad (2)$$

In the above, $K = (X, X)$ is the $n \times n$ matrix of covariances evaluated between all possible pairs of n training outputs using the covariance function, X is the $P \times n$ matrix of training input (column) vectors $\mathbf{u}_i$, $|K|$ is the determinant of K, and $\mathbf{y}$ is the training target (column) vector.

In order to carry out predictions, a jointly Gaussian (normal) distribution of zero mean is assumed over the n random variables that represent the training outputs and are contained in column vector $\mathbf{f}$, and the n_* random variables representing the test outputs contained in $\mathbf{f}_*$—this is the prior distribution:

$$\begin{bmatrix} \mathbf{f} \\ \mathbf{f}_* \end{bmatrix} \sim N\left(\mathbf{0}, \begin{bmatrix} K(X,X) & K(X,X_*) \\ K(X_*,X) & K(X_*,X_*) \end{bmatrix}\right). \qquad (3)$$

In the above, $K(X, X_*)$ is the $n \times n_*$ matrix of covariances evaluated between all possible pairs of n training and n_* test outputs, with X_* a matrix containing the test input vectors (other sub-matrices are similarly defined).

The distribution of the test outputs conditioned on the known training outputs $\mathbf{y}$, or the posterior distribution (again multivariate Gaussian), can then be expressed as $\mathbf{f}_*|X_*, X, \mathbf{y} \sim N(\mathbf{m}, \Sigma)$ [3], with mean vector $\mathbf{m}$ and covariance matrix Σ given by

$$\mathbf{m} = K(X_*, X)K(X, X)^{-1}\mathbf{y} \tag{4}$$

$$\Sigma = K(X_*, X_*) - K(X_*, X)K(X, X)^{-1}K(X, X_*). \tag{5}$$

In the above, the predictive mean $\mathbf{m}$ contains the most likely values of the test outputs associated with the test input vectors in X_*, while the diagonal of the covariance matrix Σ gives the corresponding predictive variances. Conditioning on the known training data can be interpreted as retaining in the posterior distribution only functions that pass through the training data points. The computational requirement for GP regression is $O(n^3)$ due to the required inversion of $K(X, X)$ which is of dimension $n \times n$.

B. Two-Stage Modeling Approach

Suppose we would like to set up highly accurate GP surrogate models $\boldsymbol{R}_s$ that map geometry (design) variables and frequency to $\mathrm{Re}\{S_{11}\}$ or $\mathrm{Im}\{S_{11}\}$, from which $|S_{11}|$ can then be calculated. (We choose to model smooth $\mathrm{Re}\{S_{11}\}$ and $\mathrm{Im}\{S_{11}\}$ responses—rather than $|S_{11}|$—as they are better handled by the squared-exponential covariance function (1). For the sake of conciseness we will only refer to $\mathrm{Re}\{S_{11}\}$ in what follows.) For greatest accuracy we need to use an n-element set of high-fidelity training data (i.e., simulated using a finely discretized mesh),

$$D_{\mathrm{fine}} = \{(\mathbf{u}_i, y_{\mathrm{fine},i}) \,|\, i = 1, \ldots, n\} \tag{6}$$

with P-dimensional input vectors

$$\mathbf{u}_i = \left[\mathbf{x}_i^T \; f_{oi}\right]^T = [x_{1i} \; x_{2i} \; \ldots \; x_{Mi} \; f_{oi}]^T \tag{7}$$

and scalar targets $y_{\mathrm{fine},i} = \mathrm{Re}\{S_{11}\}_{\mathrm{fine},i}$. The design vector $\mathbf{x}_i = [x_{1i} \; x_{2i} \ldots x_{Mi}]^T$ consists of M adjustable antenna geometry variables and f_{oi} is a frequency value within the range of interest; hence, $P = M + 1$.

The cost of generating D_{fine} however may be prohibitive. In order to address this problem, we adopt a two-stage modeling approach. It aims at setting up a final GPR model that is based on a fraction of the high-fidelity simulations required to set up D_{fine} but is almost as accurate as a GPR model trained on the actual D_{fine}.

1) First Stage: The purpose of this stage is to "approximate" the expensive fine training data set D_{fine} by a relatively inexpensive data set $D_{\mathrm{fine,approx}}$ of the same size; this is accomplished by means of a separate auxiliary model $\boldsymbol{R}_{\mathrm{aux}}$ trained on a specially constructed training data set D_{aux}.

More specifically, instead of attempting to directly simulate the high-fidelity training data, we rather simulate—cheaply—the above n data points using a coarse discretization, yielding the data set

$$D_{\mathrm{coarse}} = \{(\mathbf{u}_i, y_{\mathrm{coarse},i}) \,|\, i = 1, \ldots, n\} \tag{8}$$

with $\mathbf{u}_i$ as before and $y_{\mathrm{coarse},i} = \mathrm{Re}\{S_{11}\}_{\mathrm{coarse},i}$. In addition, we simulate only a (small) randomly selected subset of D_{fine} consisting of $n_{\mathrm{aux}} < n$ points. Using this subset of D_{fine}, we construct a training set D_{aux} for $\boldsymbol{R}_{\mathrm{aux}}$ as follows:

$$D_{\mathrm{aux}} = \{(\mathbf{u}_{\mathrm{aux},k}, y_{\mathrm{fine},k}) \,|\, k = 1, \ldots, n_{\mathrm{aux}}\} \tag{9}$$

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

$$\mathbf{u}_{\mathrm{aux},k} = [x_{1k} \; x_{2k} \ldots x_{Mk} \; f_{ok} \; \mathrm{Re}\{S_{11}\}_{\mathrm{coarse},k}]^T \tag{10}$$

is of the form (7) augmented by the associated coarse $\mathrm{Re}\{S_{11}\}$ target value from D_{coarse}, and the target $y_{\mathrm{fine},k}$ is the corresponding $\mathrm{Re}\{S_{11}\}$ value from the above subset of D_{fine} (recall that D_{coarse} and D_{fine} share the same set of input vectors; the only difference lies in the meshing density with which the targets have been obtained). Hence we essentially learn a mapping between coarse and fine $\mathrm{Re}\{S_{11}\}$ simulations using training data that correspond to n_{aux} specific instances of sets of design variables and frequency (the first $M + 1$ elements of the input vector $\mathbf{u}_{\mathrm{aux},k}$ serves to uniquely identify the $\mathrm{Re}\{S_{11}\}$ values). The aforementioned mapping represents the correlations between the coarse and fine model responses. Due to the fact that both models are physically related (as evaluated using the same EM solver), the mapping learned for a limited number of fine training points is likely to be preserved across the entire design space.

After training, we use $\boldsymbol{R}_{\mathrm{aux}}$ to predict, from their coarsely simulated counterparts, the $n - n_{\mathrm{aux}}$ fine $\mathrm{Re}\{S_{11}\}$ values that were not simulated; we refer to these predicted targets as $y_{\mathrm{pred},k} = \mathrm{Re}\{S_{11}\}_{\mathrm{pred},k}, k = (n_{\mathrm{aux}} + 1), \ldots, n$. Taken together, the n_{aux} full-wave simulated fine $\mathrm{Re}\{S_{11}\}$ target values and the $n - n_{\mathrm{aux}}$ predicted ones yield—along with input vectors consisting of geometry parameters and frequency of the form (7)—an n-point "approximate" fine training data set for $\boldsymbol{R}_s$,

$$D_{\mathrm{fine,approx}} = \left\{ \begin{array}{c} (\mathbf{u}_k, y_{\mathrm{fine},k}) \,|\, k = 1, \ldots, n_{\mathrm{aux}} \\ (\mathbf{u}_k, y_{\mathrm{pred},k}) \,|\, k = (n_{\mathrm{aux}} + 1), \ldots, n \end{array} \right\}. \tag{11}$$

Obtaining the targets $y_{\mathrm{pred},k}$ via model predictions (as opposed to direct full-wave simulations) can result in very significant savings in computational costs, as will be outlined next.

2) Second Stage: Here we use $D_{\mathrm{fine,approx}}$ (instead of the full D_{fine}, which is not available) as training set for $\boldsymbol{R}_s$, the desired final surrogate that maps design variables and frequency to $\mathrm{Re}\{S_{11}\}$, using the "conventional" GPR of Section II-A. In Section IV, we show that these surrogates are sufficiently accurate to be used to good effect for optimization using space mapping.

It should be emphasized that the reduction in the number of simulated high-fidelity training points without compromising model accuracy is possible because the knowledge embedded in the low-fidelity model is exploited. In our approach the use of this knowledge is implemented through the mapping learned in the first stage that identifies correlations between the low- and high-fidelity simulation data.

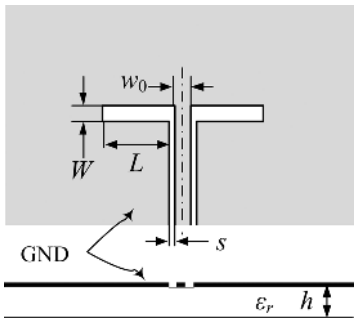

Fig. 1. Geometry of CPW-fed slot dipole antenna (Antenna 1). The ground plane (GND) is of infinite lateral extent.

III. Verification Examples

A. Slot Dipole Antenna (Antenna 1)

The geometry of a CPW-fed slot dipole antenna on a single-layer dielectric substrate is shown in Fig. 1. The design vector was $\mathbf{x} = [W\ L]^T$, and the design variable space was defined by the center vector $\boldsymbol{x}^0 = [7.5\ 39]^T$ mm and size vector $\boldsymbol{\delta} = [2.5\ 11]^T$ mm such that the variable ranges were $\boldsymbol{x}^0 \pm \boldsymbol{\delta}$ mm. Other dimensions/parameters were $w_0 = 4.0$ mm, $s = 0.5$ mm, $h = 1.6$ mm, and $\varepsilon_r = 4.4$. We were interested in S_{11} over the frequency band 2.0–2.7 GHz (visual inspection revealed that responses involving $\mathrm{Re}\{S_{11}\}$ and $\mathrm{Im}\{S_{11}\}$ against frequency over this band varied substantially throughout the design space).

In order to obtain training data input vectors for $\boldsymbol{R}_s$, 91 geometries were randomly selected from the input space using Latin hypercube sampling (LHS), with three frequencies per geometry uniformly randomly sampled from the above frequency range such that each geometry generally had a different set of frequencies. The total number of training input vectors for $\boldsymbol{R}_s$ was $n = 91 \times 3 = 273$; training input vectors had the form $\{\mathbf{u}_i = [\mathbf{x}_i^T\ f_{oi}]^T \,|\, i = 1, \ldots, n\}$, with f_{oi} a frequency value within the range of interest. Test input vectors were compiled from 100 new geometries, also obtained via LHS, with 71 equally-spaced frequencies per geometry ($n_* = 7100$).

Using CST Microwave Studio [14] on a dual-core 2.33 GHz Intel CPU with 2 GB RAM, we simulated the above training input vectors at a fine mesh density (∼130 000 mesh cells, simulation time 12 min) resulting in the (full) high-fidelity training data set D_{fine}, and at a coarse mesh density (∼5000 mesh cells, simulation time 30 s) yielding D_{coarse}. (We refer to the CST simulations at the fine mesh density as the high-fidelity model $\boldsymbol{R}_f$, and the simulations at the coarse density as the low-fidelity model $\boldsymbol{R}_c$). The test inputs were only simulated at the fine mesh density, yielding the test data set D_{test} used to evaluate the predictive capabilities of $\boldsymbol{R}_s$.

In carrying out the first stage of our approach we constructed training sets D_{aux} by randomly selecting n_{aux} data points from the above full D_{fine}, and then trained a model $\boldsymbol{R}_{\mathrm{aux}}$ as described in Section II.B that subsequently was used to estimate the rest of the high-fidelity target values in D_{fine}. This was repeated for $n_{\mathrm{aux}}/n \times 100\% \in \{70\%, 60\%, 50\%, 40\%, 30\%, 20\%, 10\%\}$, and the predictive errors of $\boldsymbol{R}_{\mathrm{aux}}$ on the remaining $n - n_{\mathrm{aux}}$ training points in D_{fine} are listed in Table I for each case. The results reveal that, even for the case $n_{\mathrm{aux}}/n \times 100\% = 10\%$,

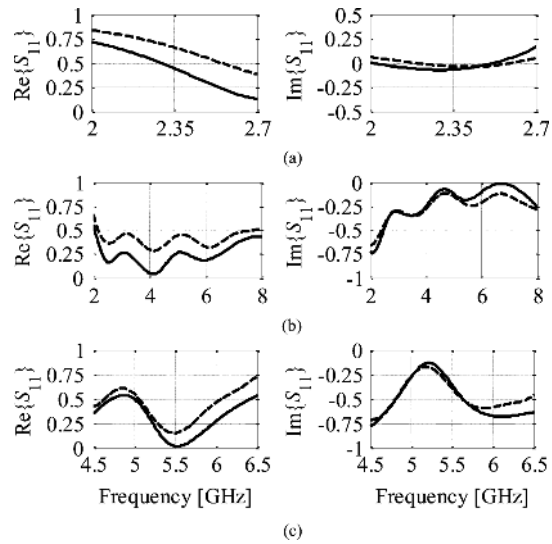

Fig. 2. Typical fine (—) and coarse (- - -) responses for $\mathrm{Re}\{S_{11}\}$ and $\mathrm{Im}\{S_{11}\}$ against frequency for (a) Antenna 1: $\mathbf{x} = [7.9107\ \ 35.4954]^T$ mm, (b) Antenna 2: $\mathbf{x} = [37.5046\ \ 25.5961\ \ 4.0234\ \ 14.8285]^T$ mm, and (c) Antenna (3): $\mathbf{x} = [7.7187\ \ 14.8472\ \ 8.8390\ \ 0.0149\ \ 1.9376\ \ 8.1535\ \ 8.9098]^T$ mm.

TABLE I
Predictive Errors[1] of Auxiliary Antenna Models $\boldsymbol{R}_{\mathrm{aux}}$ on Remaining $n - n_{\mathrm{aux}}$ Fine Training Data Points

n_{aux}/n ×100%	RMSE (%)					
	Antenna 1 ($n = 273$)		Antenna 2 ($n = 3348$)		Antenna 3 ($n = 1600$)	
	$\mathrm{Re}\{S_{11}\}$	$\mathrm{Im}\{S_{11}\}$	$\mathrm{Re}\{S_{11}\}$	$\mathrm{Im}\{S_{11}\}$	$\mathrm{Re}\{S_{11}\}$	$\mathrm{Im}\{S_{11}\}$
70	0.416	0.356	1.35	1.16	0.631	0.634
60	0.401	0.916	1.34	1.22	0.672	0.633
50	0.304	0.604	1.51	1.26	0.754	0.842
40	0.235	0.561	1.27	1.24	0.863	0.805
30	0.219	0.612	1.54	1.51	1.07	0.945
20	0.444	0.648	1.63	1.65	1.48	1.261
10	0.947	1.834	2.09	2.11	3.04	1.948

[1] root mean square error normalized to the target range, expressed as a percentage.

the remaining training targets could be predicted with reasonable accuracy by $\boldsymbol{R}_{\mathrm{aux}}$, likely due to the fact that values of $\mathrm{Re/Im}\{S_{11}\}_{\mathrm{coarse},k}$ in the training input vectors (10) were well correlated with the targets $\mathrm{Re/Im}\{S_{11}\}_{\mathrm{fine},k}$. Fig. 2(a) gives fine and coarse responses of $\mathrm{Re}\{S_{11}\}$ and $\mathrm{Im}\{S_{11}\}$ against frequency for a sample geometry; these responses are indicative of the typical discrepancy between these responses for Antenna 1. It is worth noticing that the overall "shapes" of the coarse and fine model responses (as functions of frequency) are similar. The major misalignment relates to the level of the responses. This indicates relatively good correlation between both models, and gives foundation for exploiting this correlation for coarse model enhancement even if a limited number of fine model training data points is utilized.

TABLE II
PREDICTIVE ERRORS OF SURROGATE ANTENNA MODELS $\boldsymbol{R}_s$ ON FINE TEST DATA

n_{aux}/n ×100%	RMSE (%)					
	ANTENNA 1 ($n = 273$)		ANTENNA 2 ($n = 3348$)		ANTENNA 3 ($n = 1600$)	
	$\text{Re}\{S_{11}\}$	$\text{Im}\{S_{11}\}$	$\text{Re}\{S_{11}\}$	$\text{Im}\{S_{11}\}$	$\text{Re}\{S_{11}\}$	$\text{Im}\{S_{11}\}$
100 ($\boldsymbol{R}_{s,\text{full}}$)	1.39	1.29	1.95	1.78	0.854	0.753
30	1.27	1.32	2.36	2.28	1.29	1.08
20	1.34	1.31	2.56	2.46	1.86	1.45
10	1.55	1.38	2.81	2.65	3.85	2.35

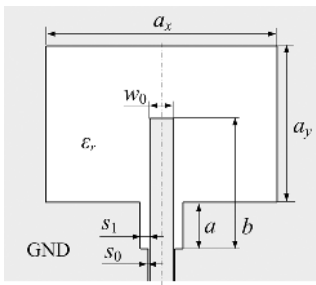

Fig. 3. Geometry of UWB CPW-fed T-shaped slot antenna (Antenna 2; top view). The ground plane (GND) is of infinite lateral extent.

In the next step, we constructed "approximate" fine training sets $D_{\text{fine,approx}}$ as described in Section II.B for cases where the savings in finely discretized training points were very significant, i.e., $n_{\text{aux}}/n \times 100\% \in \{30\%, 20\%, 10\%\}$, and trained GPR models $\boldsymbol{R}_s$ in each case. The predictive errors of these models on the test data set D_{test} are listed in Table II. Also given for comparison is the predictive error for the case where the full D_{fine} was used as training data, i.e., $n_{\text{aux}}/n \times 100\% = 100\%$—we refer to this model as $\boldsymbol{R}_{s,\text{full}}$. Again, predictive accuracies appeared to be good given the relatively small proportions of high-fidelity data present in the "approximate" fine training data sets.

B. UWB T-Shaped Slot Antenna (Antenna 2)

Fig. 3 shows the antenna layout [15]. The design vector was $\mathbf{x} = [a_x \ a_y \ a \ b]^T$, and the design space was delimited by the center vector $\boldsymbol{x}^0 = [40 \ 27.5 \ 7 \ 20]^T$ mm and size vector $\delta = [5 \ 7.5 \ 5 \ 10]^T$ mm such that the variable ranges were $\boldsymbol{x}^0 \pm \boldsymbol{\delta}$ mm (other dimensions were $w_0 = 4.0$ mm, $s_0 = 0.3$ mm, and $s_1 = 1.7$ mm; the single-layer substrate had height $h = 0.813$ mm and dielectric constant $\varepsilon_r = 3.38$). The frequency band of interest was 2–8 GHz (as before, visual inspection confirmed that S_{11}-against-frequency responses varied substantially throughout the design space).

Training data were comprised of 270 geometries obtained by LHS, with 12 frequencies per geometry, randomly selected as before ($n = 3348$). Test data were comprised of 50 new LHS geometries, with 121 equally-spaced frequencies per geometry.

Using CST Microwave Studio [14], we simulated the training input vectors at a fine mesh density (∼2 962 000 mesh cells, simulation time 21 min) resulting in D_{fine}, and at a coarse mesh density (∼44 500 mesh cells, simulation time 20 s) yielding D_{coarse}.

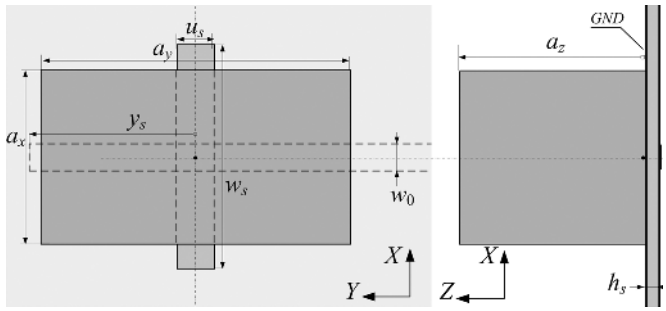

Fig. 4. Geometry of dielectric resonator antenna (Antenna 3): (a) top and (b) side views.

The models $\boldsymbol{R}_{\text{aux}}$ and $\boldsymbol{R}_s$ were set up in a manner similar to those for Antenna 1, and Tables I and II give the relevant predictive errors, again showing that both types of models had good predictive capabilities.

As in the case of Antenna 1, predictive accuracies for the $\boldsymbol{R}_s$ models appeared to be good given the comparatively small fractions of high-fidelity data present in the "approximate" fine training data sets. Fig. 2(b) gives fine and coarse responses of $\text{Re}\{S_{11}\}$ and $\text{Im}\{S_{11}\}$ against frequency for a sample geometry; these responses are representative of the fine/coarse discrepancies observed for this antenna.

C. Dielectric Resonator Antenna (Antenna 3)

The antenna geometry is shown in Fig. 4 [16]. The design vector was $\boldsymbol{x} = [a_x \ a_y \ a_z \ a_c \ u_s \ w_s \ y_s]^T$, where a_x, a_y, and a_z are dimensions of the dielectric resonator (DR) brick, a_c stands for the shift of the DR center in the Y-direction relative to the slot center, u_s is the slot width, w_s is the slot length, and y_s is the length of the microstrip stub. The relative dielectric constant and loss tangent of the DR were 10 and 10^{-4}, respectively. The substrate was 0.5 mm thick RO4003C material [16], and the metallization of the trace and ground was 50 μm copper. The design variable space was defined by the center vector $\boldsymbol{x}^0 = [8 \ 14 \ 8 \ 1 \ 2 \ 9 \ 8]^T$ mm and size vector $\boldsymbol{\delta} = [1 \ 1 \ 1 \ 1 \ 1 \ 1]^T$ mm. Other dimensions were $w_0 = 1.15$ mm and $h_s = 0.5$ mm. The frequency band of interest was 4.5–6.5 GHz.

Training data were 400 geometries obtained by LHS, with four randomly selected frequencies per geometry ($n = 1600$). Test data were comprised of 50 new LHS geometries with 121 equally-spaced frequencies per geometry.

Using CST Microwave Studio, we simulated the training input vectors at a fine mesh density (∼500 000 mesh cells, simulation time 12.5 min) resulting in D_{fine}, and at a coarse mesh density ∼15 000 mesh cells, simulation time 30 s) yielding D_{coarse}.

The models $\boldsymbol{R}_{\text{aux}}$ and $\boldsymbol{R}_s$ were set up in a manner similar to those for Antennas 1 and 2, and Tables I and II provide the relevant predictive errors that on the whole were good given that this antenna had seven design variables. Fig. 2(c) gives fine and coarse responses of $\text{Re}\{S_{11}\}$ and $\text{Im}\{S_{11}\}$ against frequency for a sample geometry that are typical of the fine/coarse discrepancy for this antenna.

IV. Application Examples: Antenna Optimizations

As a practical way to validate the proposed modeling methodology where our GPR surrogates $\boldsymbol{R}_s$ are trained using the "approximate" high-fidelity training set $D_{\text{fine,approx}}$ rather than the "original" training set D_{fine} obtained in full through direct simulation, we apply $\boldsymbol{R}_s$ within a space mapping algorithm aimed at optimizing the input characteristics of the antenna structures considered in the previous section. Optimization results are compared to results obtained by instead using $\boldsymbol{R}_{s,\text{full}}$ (i.e., $\boldsymbol{R}_s$ trained on D_{fine}).

It should be emphasized that the GPR surrogates considered in this paper are supposed to be multiple-purpose library models. Antenna optimization with respect to various sets of design specifications is one example of a typical application task. Another could be robust (yield-driven) optimization or statistical analysis.

Here, we consider antenna optimization with the initial design being the center of the region of interest $\boldsymbol{x}^{(0)}$. The design process starts from directly optimizing the GPR model. Because of nonzero modeling error, an iterative design refinement procedure is utilized that employs space mapping technology [11]

$$\boldsymbol{x}^{(i+1)} = \arg\min_{\boldsymbol{x}} U\left(\boldsymbol{R}_{\text{su}}^{(i)}(\boldsymbol{x})\right) \tag{12}$$

where $\boldsymbol{R}_{\text{su}}^{(i)}$ is a surrogate model obtained by output space mapping [11]. The surrogate model setup is performed using an evaluation of $\boldsymbol{R}_f$ at $\boldsymbol{x}^{(i)}$. U implements design specifications. For the sake of simplicity, we simply use the symbol $\boldsymbol{R}_{\text{co}}$ to denote either of $\boldsymbol{R}_{s.\text{full}}$ or $\boldsymbol{R}_s$, which can be considered the "coarse" models in the space mapping context. The surrogate model is defined as

$$\boldsymbol{R}_{\text{su}}^{(i)}(\boldsymbol{x}) = \boldsymbol{R}_{\text{co}}(\boldsymbol{x}) + \boldsymbol{d}^{(i)} \tag{13}$$

with

$$\boldsymbol{d}^{(i)} = \boldsymbol{R}_f(\boldsymbol{x}^{(i)}) - \boldsymbol{R}_{\text{co}}(\boldsymbol{x}^{(i)}). \tag{14}$$

The additive correction term $\boldsymbol{d}^{(i)}$ is calculated to ensure zero-order consistency (i.e., $\boldsymbol{R}_{\text{su}}^{(i)}(\boldsymbol{x}^{(i)}) = \boldsymbol{R}_f(\boldsymbol{x}^{(i)})$) between the surrogate and the high-fidelity model $\boldsymbol{R}_f$ [17] at the current design $\boldsymbol{x}^{(i)}$. In practice, because of good initial accuracy of the GPR surrogates, one or two iterations of the algorithm (12) are usually sufficient to yield an optimized design. It should be noted that the cost of each iteration (12) effectively corresponds to a single evaluation of the high-fidelity model: the cost of optimizing the surrogate itself can be neglected as compared to the evaluation of the high-fidelity model.

Fig. 5 shows, for all three antenna structures, the responses of models $\boldsymbol{R}_{s,\text{full}}$ and $\boldsymbol{R}_f$ (i.e., direct high-fidelity CST simulations) at the initial designs, as well as the response of $\boldsymbol{R}_f$ at the final designs. Likewise, Fig. 6 shows the responses of the GPR models $\boldsymbol{R}_s$ trained on the "approximate" high-fidelity data set $D_{\text{fine,approx}}$ (here with $n_{\text{aux}}/n \times 100\% = 20\%$) and the $\boldsymbol{R}_f$ model responses at $\boldsymbol{x}^{(0)}$, and the response of $\boldsymbol{R}_f$ at the final designs. The numerical results are summarized in Table III. It can be observed that the design quality and cost (expressed in terms of number of $\boldsymbol{R}_f$ evaluations) are very similar for the GPR

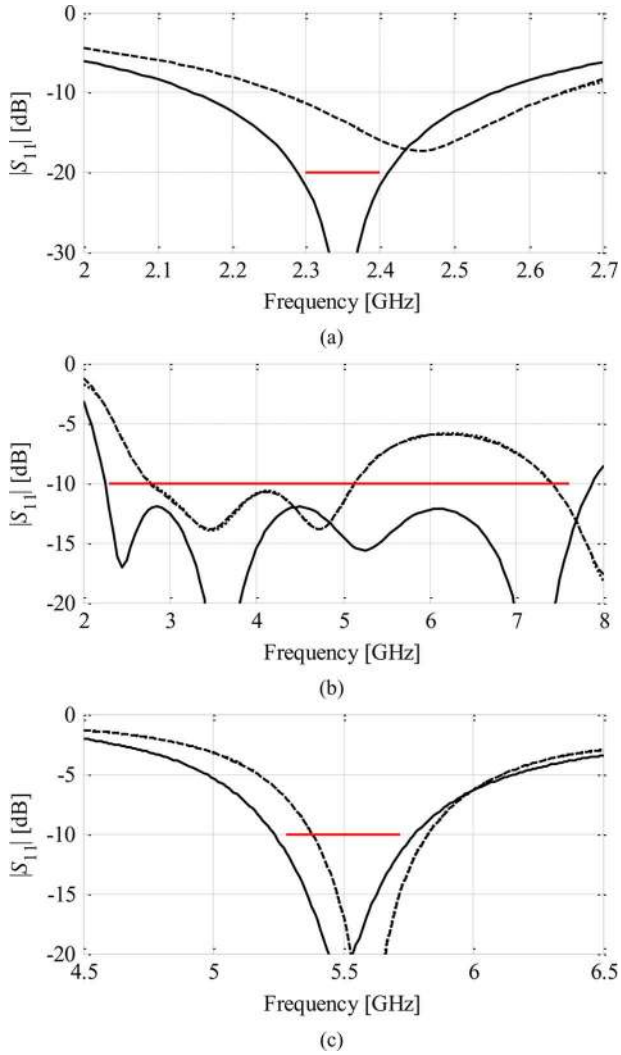

Fig. 5. Optimization results: responses of $\boldsymbol{R}_s$, full ($\cdots$) and $\boldsymbol{R}_f$(- - -) at the initial design, and $\boldsymbol{R}_f$ at the optimized design (—) for (a) Antenna 1, (b) Antenna 2, and (c) Antenna 3. Design specifications marked with horizontal solid line. Note that the GPR model responses are hardly distinguishable from the corresponding high-fidelity simulation ($\boldsymbol{R}_f$) response. (Note that the GPR responses were in fact computed from separate models for $\text{Re}\{S_{11}\}$ and $\text{Im}\{S_{11}\}$).

models obtained using the original and approximate high-fidelity training data sets. For Antennas 1 and 2, the optimization cost corresponds to three $\boldsymbol{R}_f$ evaluations. For Antenna 3, $\boldsymbol{R}_{s,\text{full}}$ exhibits better performance with only one refinement iteration necessary to yield an optimized design (three iterations for $\boldsymbol{R}_s$). Table III also contains the optimization results with the $\boldsymbol{R}_s$ models trained on $D_{\text{fine,approx}}$ where only 10% of the data were high-fidelity-simulated points. Despite the fact that these models are generally less accurate than the 20% versions (cf. Table II), they are still sufficiently reliable—in combination with the particular surrogate-based optimization technique (12)–(14)—to optimize our antenna structures: the quality of the final designs as well as the corresponding design costs are essentially the same for both 10%- and 20% high-fidelity-simulated-points GPR models.

For the sake of comparison, we also optimized the three antennas using a conventional (not surrogate-based) method, namely a state-of-the-art pattern-search algorithm [18], [19]

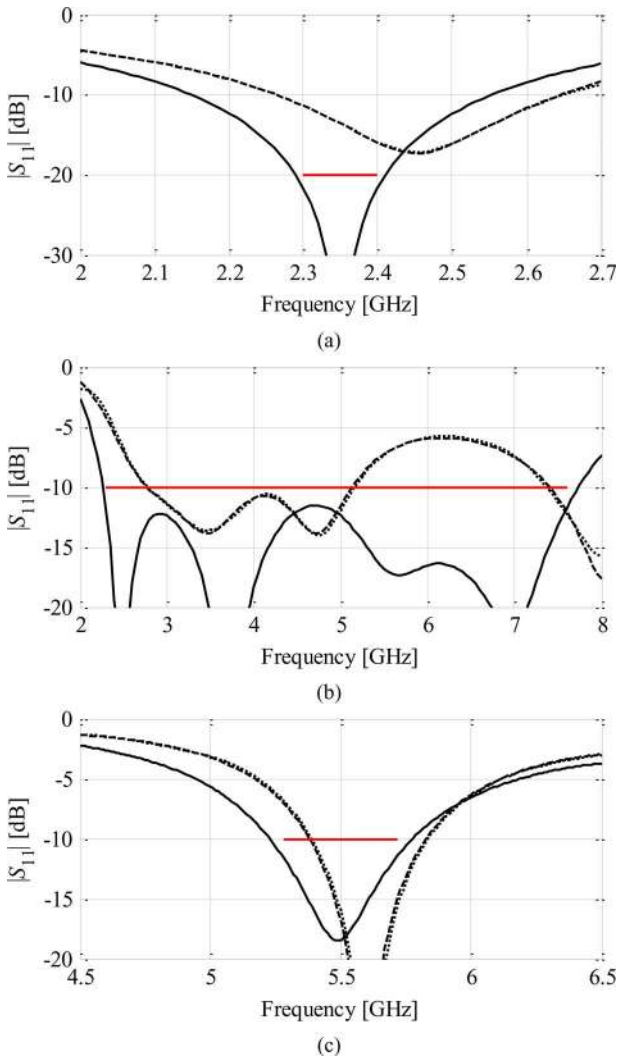

Fig. 6. Optimization results: responses of $\boldsymbol{R}_s(\cdots)$ and $\boldsymbol{R}_f(\text{- - -})$ at the initial design, and $\boldsymbol{R}_f$ at the optimized design (—) for (a) Antenna 1, (b) Antenna 2, and (c) Antenna 3. Design specifications marked with horizontal solid line. Note that the GPR model responses are very close to the high-fidelity model ($\boldsymbol{R}_f$) response. The $\boldsymbol{R}_s$ models represented here were trained on $D_{\text{fine,approx}}$ utilizing 20% data points simulated at high fidelity.

TABLE III
ANTENNA OPTIMIZATION RESULTS

Antenna	Model	max $\lvert S_{11}\rvert$ at Final Design[1]	Optimization Cost[2]
1	$\boldsymbol{R}_{s,full}$	–21.7 dB	3
	$\boldsymbol{R}_s$ (20%)[3]	–21.6 dB	3
	$\boldsymbol{R}_s$ (10%)[3]	–21.5 dB	3
2	$\boldsymbol{R}_{s,full}$	–12.0 dB	3
	$\boldsymbol{R}_s$ (20%)[3]	–11.5 dB	3
	$\boldsymbol{R}_s$ (10%)[3]	–11.4 dB	3
3	$\boldsymbol{R}_{s,full}$	–11.5 dB	2
	$\boldsymbol{R}_s$ (20%)[3]	–11.3 dB	4
	$\boldsymbol{R}_s$ (10%)[3]	–11.2 dB	4

[1] max $\lvert S_{11}\rvert$ at the frequency band of interest: 2.3 to 2.4 GHz (Antenna 1), 2.3 to 7.6 GHz (Antenna 2), and 5.28 to 5.72 GHz (Antenna 3). [2] Number of $\boldsymbol{R}_f$ evaluations including evaluation at the initial design. [3] 20% refers to the model trained on $D_{\text{fine,approx}}$ utilizing 20% data points that are actually simulated at high fidelity (10% accordingly).

that directly relied on fine-discretization full-wave simulations (i.e., $\boldsymbol{R}_f$) for its objective function evaluations. While maximum $|S_{11}|$ values at the final designs obtained for Antennas 1, 2, and 3 (-21.6 dB, -11.6 dB, and -10.4 dB, respectively) were similar to those obtained using our GPR models and the above space-mapping procedure, the computational expense for the conventional optimization was at least an order of magnitude larger (i.e., 40, 148, and 117 $\boldsymbol{R}_f$ evaluations for Antennas 1, 2, and 3 respectively). This indicates that fast and accurate surrogates are indispensable in the antenna design process, particularly, if they can be set up at relatively low computational cost. Here, the GPR models trained on $D_{\text{fine,approx}}$ are constructed at the cost of only 10 to 20 percent of that for GPR models based on the conventional approach.

V. CONCLUSION

A two-stage technique for Gaussian Process modeling of antenna input characteristics is presented. Our approach involves variable-fidelity electromagnetic simulations. By exploiting the knowledge embedded in the low-fidelity antenna model (simulations) utilized in the first modeling stage, where the mapping between the electromagnetic models of different fidelity is learned, it is possible to substantially reduce the number of actual high-fidelity simulations that need to be performed, without compromising the predictive power of the final surrogate. As demonstrated using three antenna examples, satisfactory results can be obtained even if the "approximate" high-fidelity training set contains only 10 to 20 percent targets obtained using fine-discretization simulations (the rest being predicted by the auxiliary model in the first stage of the procedure). As an additional verification, the above models were shown to be perfectly usable in a design/optimization context.

ACKNOWLEDGMENT

The authors would like to thank Computer Simulation Technology AG, Darmstadt, Germany, for making CST Microwave Studio available, and S. Ogurtsov for running the antenna simulations in CST.

REFERENCES

[1] R. L. Haupt, "Antenna design with a mixed integer genetic algorithm," *IEEE Trans. Antennas Propag.*, vol. 55, no. 3, pp. 577–582, Mar. 2007.
[2] M. F. Pantoja, P. Meincke, and A. R. Bretones, "A hybrid genetic algorithm space-mapping tool for the optimization of antennas," *IEEE Trans. Antennas Propag.*, vol. 55, no. 3, pp. 777–781, Mar. 2007.
[3] C. E. Rasmussen and C. K. I. Williams, *Gaussian Processes for Machine Learning*. Cambridge, MA, USA: MIT Press, 2006.
[4] J. P. De Villiers and J. P. Jacobs, "Gaussian process modeling of CPW-fed slot antennas," *Progress in Electromagn. Res.*, vol. 98, pp. 233–249, 2009.
[5] J. P. Jacobs and J. P. De Villiers, "Gaussian-process-regression-based design of ultrawide-band and dual-band CPW-fed slot antennas," *J. Electromagn. Waves Applicat.*, vol. 24, pp. 1763–1772, 2010.
[6] G. Angiulli, M. Cacciola, and M. Versaci, "Microwave devices and antennas modelling by support vector regression machines," *IEEE Trans. Magn.*, vol. 43, no. 4, pp. 1589–1592, Apr. 2007.
[7] J. P. Jacobs, "Bayesian support vector regression with automatic relevance determination kernel for modeling of antenna input characteristics," *IEEE Trans. Antennas Propag.*, vol. 60, no. 4, pp. 2114–2118, Apr. 2012.
[8] V. K. Devabhaktuni, M. C. E. Yagoub, and Q. J. Zhang, "A robust algorithm for automatic development of neural network models for microwave applications," *IEEE Trans. Microw. Theory Tech.*, vol. 49, no. 12, pp. 2282–2291, Dec. 2001.

[9] I. Couckuyt, F. Declercq, T. Dhaene, H. Rogier, and L. Knockaert, "Surrogate-based infill optimization applied to electromagnetic problems," *Int. J. RF Microw. CAE*, vol. 20, no. 5, pp. 492–501, 2010.
[10] J. P. Jacobs, S. Koziel, and S. Ogurtsov, "Computationally efficient multi-fidelity Bayesian support vector regression modeling of planar antenna input characteristics," *IEEE Trans. Antennas Propag.*, vol. 61, no. 2, pp. 980–984, Feb. 2013.
[11] S. Koziel, J. W. Bandler, and K. Madsen, "A space mapping framework for engineering optimization: Theory and implementation," *IEEE Trans. Microw. Theory Tech.*, vol. 54, no. 10, pp. 3721–3730, Oct. 2006.
[12] M. B. Yelten, T. Zhu, S. Koziel, P. D. Franzon, and M. B. Steer, "Demystifying surrogate modeling for circuits and systems," *IEEE Circuits Syst. Mag.*, vol. 12, no. 1, pp. 45–63, First Quarter, 2012.
[13] Q. S. Cheng, J. W. Bandler, S. Koziel, M. H. Bakr, and S. Ogurtsov, "The state of the art of microwave CAD: EM-based optimization and modeling," *Int. J. RF Microw. Comput.-Aided Eng.*, vol. 20, no. 5, pp. 475–491, 2010.
[14] CST Microwave Studio, Germany, ver. 2011, CST AG, Bad Nauheimer Str. 19, D-64289 Darmstadt, 2011.
[15] J.-J. Jiao, G. Zhao, F.-S. Zhang, H.-W. Yuan, and Y.-C. Jiao, "A broadband CPW-fed T-shape slot antenna," *Progress in Electromagn. Res.*, vol. 76, pp. 237–242, 2007.
[16] A. Petosa, *Dielectric Resonator Antenna Handbook*. Norwell, MA, USA: Artech House, 2007.
[17] N. M. Alexandrov and R. M. Lewis, "An overview of first-order model management for engineering optimization," *Optimiz. Eng.*, vol. 2, no. 4, pp. 413–430, 2001.
[18] T. G. Kolda, R. M. Lewis, and V. Torczon, "Optimization by direct search: New perspectives on some classical and modern methods," *SIAM Rev.*, vol. 45, no. 3, pp. 385–482, 2003.
[19] S. Koziel, "Multi-fidelity multi-grid design optimization of planar microwave structures with Sonnet," in *Int. Rev. Progr. Appl. Comput. Electromagn.*, Tampere, Finland, Apr. 26–29, 2010, pp. 719–724.

J. P. Jacobs received the B.Eng., M.Eng., and PhD degrees in electronic engineering from the University of Pretoria, Pretoria, South Africa, and a doctorate in music from Yale University, New Haven, CT, USA.

His research interests include modeling and optimization of microwave antennas and devices, computational electromagnetics, slot antennas and arrays, and pattern recognition in music.

Dr. Jacobs is a registered Professional Engineer in South Africa, and is a Senior Lecturer in the Department of Electrical, Electronic and Computer Engineering at the University of Pretoria.

S. Koziel received the M.Sc. and Ph.D. degrees in electronic engineering from Gdansk University of Technology, Gdansk, Poland, in 1995 and 2000, respectively, the M.Sc. degrees in theoretical physics and in mathematics, in 2000 and 2002, respectively, and the Ph.D. degree in mathematics in 2003, from the University of Gdansk.

He is currently a Professor with the School of Science and Engineering, Reykjavik University, Reykjavik, Iceland. His research interests include CAD and modeling of microwave circuits, simulation-driven design, surrogate-based optimization, space mapping, circuit theory, analog signal processing, evolutionary computation, and numerical analysis.