

Bayesian Support Vector Regression Modeling of Microwave Structures for Design Applications

J. Pieter Jacobs, Slawomir Koziel, and Leifur Leifsson

Abstract Fast and accurate models are indispensable in contemporary microwave engineering. Kernel-based machine learning methods applied to the modeling of microwave structures have recently attracted substantial attention; these include support vector regression and Gaussian process regression. Among them, Bayesian support vector regression (BSVR) with automatic relevance determination (ARD) proved to perform particularly well when modeling input characteristics of microwave devices. In this chapter, we apply BSVR to the modeling of microwave antennas and filters. Moreover, we discuss a more efficient version of BSVR-based modeling exploiting variable-fidelity electromagnetic (EM) simulations, where coarse-discretization EM simulation data is used to find a reduced number of fine-discretization training points for establishing a high-fidelity BSVR model of the device of interest. We apply the BSVR models to design optimization. In particular, embedding the BSVR model obtained from coarse-discretization EM data into a surrogate-based optimization framework exploiting space mapping allows us to yield an optimized design at a low computational cost corresponding to a few evaluations of the high-fidelity EM model of the considered device. The presented techniques are illustrated using several examples of antennas and microstrip filters.

Keywords Bayesian support vector regression · Computer-aided design (CAD) · Electromagnetic (EM) simulation · Microwave engineering · Space mapping · Surrogate-based optimization · Surrogate modeling

J.P. Jacobs (✉)

Centre for Electromagnetism, Department of Electrical, Electronic and Computer Engineering,
University of Pretoria, Pretoria, 0002, South Africa
e-mail: jpjacobs@up.ac.za

S. Koziel · L. Leifsson

Engineering Optimization & Modeling Center, School of Science and Engineering, Reykjavik
University, Menntavegur 1, 101 Reykjavik, Iceland

S. Koziel

e-mail: koziel@ru.is

L. Leifsson

e-mail: leifurth@ru.is

1 Introduction

Full-wave electromagnetic (EM) simulations based on the method of moments and/or finite elements play a ubiquitous part in microwave engineering, as they permit highly accurate evaluation of microwave structures such as planar antennas and filters. Such simulations, however, are costly in computational terms, and their use for tasks requiring numerous analyses (e.g., statistical analysis and parametric design optimization) might become infeasible under certain conditions (for instance, a genetic algorithm optimization might necessitate thousands of full-wave analyses of candidate geometries of the structure to be optimized). Hence, surrogate models are used instead. Trained on a training set consisting of a limited number of input-output pairs (such as adjustable antenna geometry parameters and frequency as input, and the magnitude of the input reflection coefficient $|S_{11}|$ obtained from full-wave simulations as output), 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.

The kernel-based machine learning method most widely used for microwave modeling tasks has been support vector regression (SVR) utilizing an isotropic Gaussian kernel [1]. It has recently been shown [2] that Bayesian support vector regression (BSVR) [3] using a Gaussian kernel with automatic relevance determination (ARD) significantly outperforms the above standard SVR with an isotropic kernel in modeling $|S_{11}|$ versus the frequency of CPW-fed slot antennas with multiple tunable geometry variables. BSVR is in essence a version of Gaussian process regression (GPR) [4]; the Bayesian framework enables efficient training of the multiple hyperparameters of the ARD kernel by minimizing the negative log probability of the data given the hyperparameters. Such training of multiple hyperparameters is intractable under standard SVR, which employs a grid-search/cross-validation approach towards this end. In addition to its advantageous Bayesian-based features, BSVR also exhibits certain desirable properties of standard SVR, such as quadratic programming and sparseness of solutions, i.e., solutions that are fully characterized by the set of SVs, which is a subset of the training set.

In this chapter, we explore BSVR within both global and local modeling contexts. Global, or “library-type,” surrogate models aim at giving accurate predictions over the entire input space, and can be used for a variety of applications (e.g., optimization and statistical analysis). In contrast, local/trust region models only apply to a subregion of input space specified by the optimization algorithm within which the model usually is embedded.

Similar to many other global modeling methods, a drawback of BSVR is the high starting cost of gathering the fine-discretization full-wave simulation data necessary to train the model so that it has high predictive accuracy. We address this problem by exploiting the sparseness property of BSVR to reduce the amount of expensive high-fidelity data required for training (see Sects. 3–5). Earlier methods aimed at optimal data selection for microwave modeling problems include various adaptive sampling techniques that aim, within optimization contexts, to reduce the number of samples necessary to ensure the desired modeling accuracy. This is done by iterative

identification of the model and the addition of new training samples based on the actual model error at selected locations (e.g., [5]) or expected error values (statistical infill criteria, e.g., [6]); [5, 6] were local/trust region models.

Our approach entails first training an auxiliary BSVR model using fast, inexpensive coarse-discretization data selected by means of traditional experimental design procedures, and then taking the support vectors of this model simulated at a high mesh density as training data for the actual (high-fidelity) BSVR model. (A similar approach was adopted in [7], but only standard SVR with an isotropic kernel was used to model comparatively uncomplicated underlying functions.) The role of the auxiliary model can be viewed as locating regions of the design space where more samples are needed compared to other regions—for example, because the response is more variable with respect to the design and/or frequency variables. Our modeling approach is demonstrated using both planar antenna and microstrip filter examples (see Sects. 4 and 5, respectively). We also evaluate the accuracy of our reduced-data BSVR surrogates by using them within a space mapping (SM) optimization framework.

As to local BSVR modeling, we consider surrogates for variable-fidelity EM-driven optimization (see Sect. 6). In this approach, the optimization is carried out using SM, whereas the underlying coarse model is created by approximating coarse-discretization EM simulation data using BSVR. The high-fidelity EM simulation is only launched to verify the design produced by the space-mapped BSVR coarse model and obtain the data for its further correction. This allows us to significantly reduce the computational cost of the design optimization process as illustrated using two antenna examples.

The above sections are preceded by a short overview in Sect. 2 of the BSVR framework, and followed by some summary remarks (see Sect. 7).

2 Modeling Using Bayesian Support Vector Regression

In this section, we briefly give an overview of the formulation of Bayesian support vector regression (BSVR).

Consider a training data set of n observations, $\mathcal{D} = \{(\mathbf{u}_i, y_i) \mid i = 1, \dots, n\}$. The BSVR formulation, which is explained at length in [3], follows the standard Bayesian regression framework for GPR in which training targets y_i corresponding to input vectors \mathbf{u}_i are expressed as $y_i = f(\mathbf{u}_i) + \delta_i$, where δ_i are independent, identically distributed noise variables; and the underlying function f is a random field. If $\mathbf{f} = [f(\mathbf{u}_1) \ f(\mathbf{u}_2) \ \dots \ f(\mathbf{u}_n)]$, then Bayes's theorem gives the posterior probability of \mathbf{f} given the training data \mathcal{D} as

$$p(\mathbf{f}|\mathcal{D}) = \frac{p(\mathcal{D}|\mathbf{f})p(\mathbf{f})}{p(\mathcal{D})}, \quad (1)$$

with $p(\mathbf{f})$ the prior probability of \mathbf{f} , $p(\mathcal{D}|\mathbf{f})$ the likelihood, and $p(\mathcal{D})$ the evidence. The likelihood is given by

$$p(\mathcal{D}|\mathbf{f}) = \prod_{i=1}^n p(\delta_i), \quad (2)$$

where $p(\delta_i) \propto \exp(-\zeta L(\delta_i))$ with $L(\delta_i)$ the loss function, and ζ a constant. In standard GPR [4] the loss function is quadratic; the crucial point in the BSVR formulation is that a new loss function, the soft insensitive loss function, is used that combines advantageous properties of both the ε -insensitive loss function (sparseness of solutions) of standard SVR [8], and Huber's loss function (differentiability). It is defined as [3]:

$$L_{\varepsilon, \beta}(\delta) = \begin{cases} -\delta - \varepsilon; & \delta \in (-\infty, -(1 + \beta)\varepsilon), \\ \frac{(\delta + (1 - \beta)\varepsilon)^2}{4\beta\varepsilon}; & \delta \in [-(1 + \beta)\varepsilon, -(1 - \beta)\varepsilon], \\ 0; & \delta \in (-(1 - \beta)\varepsilon, (1 - \beta)\varepsilon), \\ \frac{(\delta - (1 - \beta)\varepsilon)^2}{4\beta\varepsilon}; & \delta \in [(1 - \beta)\varepsilon, (1 + \beta)\varepsilon], \\ \delta - \varepsilon; & \delta \in ((1 + \beta)\varepsilon, +\infty), \end{cases} \quad (3)$$

where $0 < \beta \leq 1$, and $\varepsilon > 0$.

Solving for the maximum a posteriori (MAP) estimate of the function values entails solving the primal problem [3, Eqs. (19)–(21)], with the corresponding dual problem given by

$$\begin{aligned} \min_{\boldsymbol{\alpha}, \boldsymbol{\alpha}^*} \quad & \frac{1}{2}(\boldsymbol{\alpha} - \boldsymbol{\alpha}^*)^\top \boldsymbol{\Sigma}(\boldsymbol{\alpha} - \boldsymbol{\alpha}^*) - \sum_{i=1}^n y_i(\alpha_i - \alpha_i^*) \\ & + \sum_{i=1}^n (\alpha_i + \alpha_i^*)(1 - \beta)\varepsilon + \frac{\beta\varepsilon}{C} \sum_{i=1}^n y_i(\alpha_i^2 + \alpha_i^{*2}) \end{aligned} \quad (4)$$

subject to $0 \leq \alpha_i, \alpha_i^* \leq C$, $i = 1, \dots, n$. In the above, $\boldsymbol{\Sigma}$ is an $n \times n$ matrix with $\Sigma_{ij} = k(\mathbf{u}_i, \mathbf{u}_j)$ and $k(\cdot)$ is the kernel function. In particular, the Gaussian kernel with ARD (used throughout in this work) is given by

$$k(\mathbf{u}_i, \mathbf{u}_j) = \sigma_f^2 \exp\left(-\frac{1}{2} \sum_{k=1}^D \frac{(u_{ik} - u_{jk})^2}{\tau_k^2}\right) + \kappa, \quad (5)$$

where u_{ik} and u_{jk} are the k th elements of the i th and j th training input vectors. The hyperparameter vector $\boldsymbol{\theta}$, which includes σ_f^2 , τ_k , κ , C , and ε , can be determined by minimizing the negative log probability of the data given the hyperparameters [3],

$$\begin{aligned} -\ln p(\mathcal{D}|\boldsymbol{\theta}) = & \frac{1}{2}(\boldsymbol{\alpha} - \boldsymbol{\alpha}^*)^\top \boldsymbol{\Sigma}(\boldsymbol{\alpha} - \boldsymbol{\alpha}^*) \\ & + C \sum_{i=1}^n L_{\varepsilon, \beta}(y_i - f_{\text{MP}}(x_i)) + \frac{1}{2} \ln \left| \mathbf{I} + \frac{C}{2\beta\varepsilon} \boldsymbol{\Sigma}_M \right| + n \ln Z_S, \end{aligned} \quad (6)$$

with Σ_M an $m \times m$ submatrix of Σ corresponding to the off-bound support vectors, \mathbf{I} the $m \times m$ identity matrix, $\mathbf{f}_{MP} = \Sigma(\boldsymbol{\alpha} - \boldsymbol{\alpha}^*)$, and Z_s defined as [3, Eq. (15)]. The length scale τ_k associated with the k th input dimension can be considered the distance that has to be traveled along that dimension before the output changes significantly [4]. The regression estimate at a test input \mathbf{u}^* can be expressed as

$$f(\mathbf{u}^*) = \sum_{i=1}^n k(\mathbf{u}_i, \mathbf{u}^*)(\alpha_i - \alpha_i^*). \quad (7)$$

Training points corresponding to $|\alpha_i - \alpha_i^*| > 0$ are the support vectors (SVs); of these, points corresponding to $0 < |\alpha_i - \alpha_i^*| < C$ are termed off-bound SVs. Usually, the lower the parameter β in the loss function, the smaller the number of SVs [3]; β determines the density function of the additive noise associated with training targets.

3 BSVR Modeling with Reduced Data Sets

In this section, we discuss a method of exploiting EM simulations of variable fidelity in order to reduce the computational cost of creating the BSVR model. The model response of interest typically is $|S_{11}|$ or $|S_{21}|$ over a specified frequency range for a particular antenna or filter geometry. A model input (column) vector \mathbf{u} consists of the set of adjustable geometry parameters \mathbf{u} and a frequency value f ; thus we have $\mathbf{u} = [\mathbf{x}^T f]^T$. The (scalar) response of a model at a specific frequency, for example, the model response \mathbf{R}_f , which is the fine-discretization full-wave simulated S -parameters, is denoted as $\mathbf{R}_f(\mathbf{u})$, or $\mathbf{R}_f(\mathbf{x}, f)$. Suppose now that a BSVR surrogate \mathbf{R}_s of the CPU-intensive high-fidelity model \mathbf{R}_f has to be constructed. As noted earlier, the computational cost of gathering sufficient data to train \mathbf{R}_s typically is high. To address this, first, an auxiliary BSVR model $\mathbf{R}_{s,\text{aux}}$ of the antenna (or filter) is set up with training data obtained from coarse-discretization full-wave simulations (these simulations are referred to as the low-fidelity full-wave model \mathbf{R}_c). The training set for $\mathbf{R}_{s,\text{aux}}$ consists of n input vectors \mathbf{x}_i , $i = 1, \dots, n$, and associated targets $y_i = \mathbf{R}_c(\mathbf{u}_i)$, where \mathbf{u}_i contains geometry parameters and a frequency value as noted above, and y_i is the corresponding simulated $|S_{11}|$ (or $|S_{21}|$) value. The SVs obtained from $\mathbf{R}_{s,\text{aux}}$ are then simulated at the (high) mesh density of \mathbf{R}_f , providing the reduced fine-discretization training set for \mathbf{R}_s .

As experience has shown that the coarsely simulated targets \mathbf{R}_c and finely simulated targets \mathbf{R}_f of microwave structures such as antennas or filters are usually reasonably well correlated (in Sects. 4 and 5 we give coarse and fine meshing densities for specific examples to indicate by how much they can differ), we assume that the regions of the input space which support the crucial variations in the coarse response surface will also support the crucial variations in the fine response surface. Hence, the SVs of the coarse model should largely capture the crucial variations in the fine data as well, and along with target values obtained through fine-discretization simulations should make an adequate reduced-data training set for a high-fidelity BSVR model, i.e., \mathbf{R}_s .

4 Modeling and Optimization of Antennas Using BSVR

In this section, we present examples illustrating how global BSVR models for the reflection coefficients of planar slot antennas can be set up based on reduced finely discretized data sets. We then use these models for design optimization. We consider 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, an ultra-wideband (UWB) CPW-fed T-shaped slot antenna, and a broadband probe-fed microstrip patch with two U-shaped parasitic elements. We furthermore evaluate the accuracy of our reduced-data BSVR surrogates by using them within a space mapping (SM) optimization framework [9–14].

4.1 Slot Dipole Antenna (Antenna 1)

Figure 1 shows the geometry of a CPW-fed slot dipole antenna on a single-layer dielectric substrate. The design variables were $\mathbf{x} = [W \ L]^T$ mm, and the input space was specified as $5 \leq W \leq 10$ mm and $28 \leq L \leq 50$ mm. Other dimensions/parameters were $w_0 = 4.0$ mm, $s = 0.5$ mm, $h = 1.6$ mm, and $\epsilon_r = 4.4$. We were concerned with $|S_{11}|$ over the frequency band 2.0–2.7 GHz (visual inspection revealed that $|S_{11}|$ -versus-frequency responses over this band varied substantially throughout the above geometry input space). Using CST Microwave Studio [15] on a dual-core 2.33 GHz Intel CPU with 2 GB RAM, we considered a high-fidelity model \mathbf{R}_f ($\sim 130,000$ mesh cells, simulation time 12 min) and a low-fidelity model \mathbf{R}_c ($\sim 5,000$ mesh cells, simulation time 30 s).

For training input data, 99 geometries were selected at random from the input space using Latin hypercube sampling (LHS), with three frequencies per geometry uniformly randomly sampled from the above frequency range such that, in general, each geometry had a different set of frequencies. The total number of training points was $n = 99 \times 3 = 297$; training input vectors had the form $\{\mathbf{u}_i = [\mathbf{x}_i^T \ f_i]^T = [W_i \ L_i \ f_i]^T | i = 1, \dots, n\}$, with W_i and L_i the design variables corresponding to the i th input vector, and f_i a frequency value within the range of interest. Test data consisted of 100 new geometries, also obtained via LHS, with 71 equally spaced frequencies per geometry. The training data were simulated at the \mathbf{R}_c mesh density, and used to train the BSVR model $\mathbf{R}_{s,\text{aux}}$ for three different values of β at the low end of its possible range ($\beta \in \{0.05, 0.15, 0.25\}$); as noted earlier, usually the smaller the value of β , the smaller the number of SVs). Each $\mathbf{R}_{s,\text{aux}}$ was used to predict the test data (also simulated at the \mathbf{R}_c mesh density). %RMSE (percentage root mean square error normalized to the target range) values were in the vicinity of 1.1 %; this high predictive accuracy confirmed that the training set was sufficiently large.

For each $\mathbf{R}_{s,\text{aux}}$ model, the SVs were identified and simulated at the \mathbf{R}_f mesh density. BSVR models fitted to these fine-discretization data gave the desired surrogate models \mathbf{R}_s . For comparison, surrogate models $\mathbf{R}_{s,\text{full}}$ trained on the full fine-discretization training data set ($n = 297$) were also set up. Table 1 gives, for each of

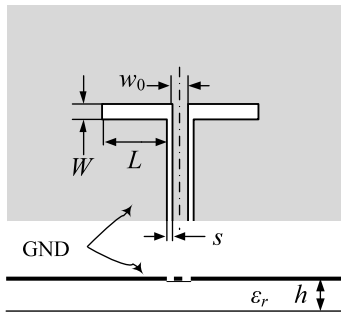


Fig. 1 Geometry of a CPW-fed slot dipole antenna (Antenna 1). The ground plane (GND) has infinite lateral extent

Table 1 Predictive errors of surrogate antenna models

Antenna	β	RMSE (%)			n_{SV}	n_{SV}/n (%)	Model cost ^a	
		$R_{s,aux}$	R_s	$R_{s,full}$			R_s ^b	$R_{s,full}$
1 ($n = 297$)	0.05	1.14	1.77	1.68	168	56	168	297
	0.15	1.13	1.71	1.72	176	59	176	
	0.25	1.15	1.68	1.68	206	69	206	
2 ($n = 3,528$)	0.05	3.37	4.14	3.83	1,819	52	1,819	3528
	0.15	3.14	4.11	3.52	2,010	57	2,010	
	0.25	3.13	3.89	3.50	2,225	63	2,225	
3 ($n = 6,400$)	0.05	6.59	5.77	5.53	3,658	57	3,658	6,400
	0.15	6.54	5.87	5.41	3,937	62	3,937	
	0.25	6.59	5.71	5.27	3,916	61	3,916	

^aCost is in terms of number of R_f evaluations, i.e., fine-discretization full-wave simulations, required to generate the training data

^bThe actual cost of setting up R_s is slightly (by 1 to 4 percent depending on the particular antenna example) higher due to the coarse-discretization EM simulations used in the model construction process

the β values, the %RMSE values obtained with $R_{s,aux}$ on the coarse-discretization test data, and R_s and $R_{s,full}$ on the fine-discretization test data.

Also given in the table is n_{SV} , the number of SVs associated with $R_{s,aux}$ (and therefore the number of training points for R_s), and n_{SV}/n , the proportion of the full training data that were SVs of $R_{s,aux}$. The %RMSE values obtained for R_s and $R_{s,full}$ were either the same, or only marginally higher in the case of R_s , indicating that reducing the number of training points from n to n_{SV} by using the SVs of $R_{s,aux}$ as training points for R_s incurred insignificant accuracy loss. In all cases, the reduction in training data was considerable: for example, for $\beta = 0.15$ the number of SVs was 176, which is 59 % of the original training data set. For ready comparison, Table 1 also explicitly lists the computational cost of generating the training data

for the models, expressed in terms of the number of fine-discretization simulations \mathbf{R}_f (for each model it simply equals the number of training points). In terms of total CPU time (which was proportional to the costs in the table), these numbers translate to about 12 h for \mathbf{R}_s ($\beta = 0.15$), and 20 h for $\mathbf{R}_{s,\text{full}}$.

4.2 UWB T-Shaped Slot Antenna (Antenna 2)

Figure 2 shows the antenna layout [16]. The design variables were $\mathbf{x} = [a_x \ a_y \ a \ b]^T$ mm, with design space $35 \leq a_x \leq 45$ mm, $20 \leq a_y \leq 35$ mm, $2 \leq a \leq 12$ mm, and $10 \leq b \leq 30$ mm ($w_0 = 4.0$ mm, $s_0 = 0.3$ mm, $s_1 = 1.7$ 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 $|S_{11}|$ -versus-frequency responses varied substantially throughout the geometry input space). Using CST Microwave Studio [15], we considered a high-fidelity model \mathbf{R}_f ($\sim 2,962,000$ mesh cells, simulation time 21 min) and a low-fidelity model \mathbf{R}_c ($\sim 44,500$ mesh cells, simulation time 20 s).

The training data consisted of 294 geometries obtained by LHS, with 12 frequencies per geometry, randomly selected as before ($n = 3,528$). The test data comprised 49 new LHS geometries, with 121 equally spaced frequencies per geometry (as before, the value of n was determined by the performance of $\mathbf{R}_{s,\text{aux}}$ on the test data simulated at the coarse mesh density).

The surrogate models $\mathbf{R}_{s,\text{aux}}$, \mathbf{R}_s , and $\mathbf{R}_{s,\text{full}}$ were set up similarly to those for Antenna 1. Table 1 gives, for three values of β , the %RMSEs obtained with $\mathbf{R}_{s,\text{aux}}$ on the coarse test data and with \mathbf{R}_s , and $\mathbf{R}_{s,\text{full}}$ on the fine test data; as well as counts of SVs.

In general, %RMSE values of \mathbf{R}_s were only somewhat higher than those of $\mathbf{R}_{s,\text{full}}$, suggesting as before that reducing the number of training points from n to n_{SV} by using the SVs of $\mathbf{R}_{s,\text{aux}}$ as fine-discretization training points for \mathbf{R}_s has little effect on prediction accuracy. The CPU time required to generate fine-discretization training data for \mathbf{R}_s in the case $\beta = 0.05$ (i.e., the model used in the optimization below) was approximately 56 h; the CPU time for $\mathbf{R}_{s,\text{full}}$ was 103 h.

4.3 Microstrip Antenna with Parasitic Elements (Antenna 3)

Figure 3 shows the antenna geometry [17]. The design variables were $\mathbf{x} = [a \ b \ c \ d \ e]^T$ mm, with design space $14 \leq a \leq 22$ mm, $0.4 \leq b \leq 2$ mm, $0.4 \leq c \leq 2$ mm, $0.4 \leq d \leq 2$ mm, and $0.4 \leq e \leq 2$ mm. The main patch had dimensions $a_0 = 5.8$ mm and $b_0 = 13.1$ mm. The lateral dimensions of the dielectric material and the metal ground were $l_x = 20$ mm and $l_y = 25$ mm. The dielectric substrate height, h , was 0.4 mm, and its relative permittivity, ε_r , was 4.3. The feed pin offset from the main patch center, l_0 , was 5.05 mm, and the pin was 0.5 mm in diameter. The frequency band of interest was 4–7 GHz.

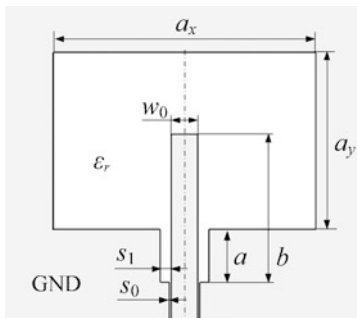


Fig. 2 Geometry of a UWB CPW-fed T-shaped slot antenna (Antenna 2, top view). The ground plane (GND) has infinite lateral extent

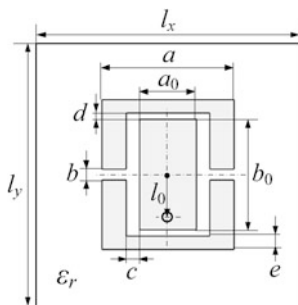


Fig. 3 Geometry of a broadband probe-fed microstrip patch antenna with two U-shaped parasitic elements (Antenna 3, top view). The dielectric substrate and ground plane both have lateral dimensions l_x and l_y . The empty circle below the center of the patch indicates the position of the feed pin

The training data were 400 geometries obtained by LHS, with 16 randomly selected frequencies per geometry ($n = 6,400$). The test data comprised 50 new LHS geometries with 121 equally spaced frequencies per geometry. We considered a high-fidelity model \mathbf{R}_f ($\sim 440,500$ mesh cells, simulation time 12 min) and a low-fidelity model \mathbf{R}_c ($\sim 25,700$ mesh cells, simulation time 15 s). It is instructive to consider three randomly picked responses from the training data, shown in Fig. 4. In spite of what appears to be a narrowly circumscribed input space (cf. the boundaries on the b , c , d , and e dimensions), the responses show considerable variety from one training point to the next. Furthermore, while within-training point coarse and fine responses agreed to some extent for some regions of the frequency band, there were considerable differences for others.

Surrogate models were constructed as before. The %RMSE values obtained with $\mathbf{R}_{s,aux}$ on the coarse test data, and \mathbf{R}_s and $\mathbf{R}_{s,full}$ on the fine test data for $\beta \in \{0.15, 0.20, 0.25\}$ are shown in Table 1, as well as n_{SV} . The greatest data reduction, by 43 %, occurred for $\beta = 0.15$, while the %RMSE only increased from 5.53 % (full model) to 5.77 % (reduced model). The CPU time necessary to simu-

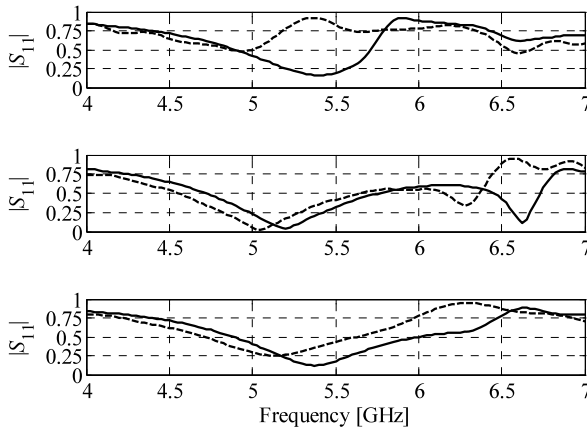


Fig. 4 Full-wave simulation responses \mathbf{R}_c (- -) and \mathbf{R}_f (—) for the Antenna 3 training geometries $\mathbf{x} = [16.2267 \ 0.7904 \ 1.8943 \ 0.4476 \ 1.1148]^T$ mm (*top*), $[14.8396 \ 0.6347 \ 0.9242 \ 0.9358 \ 0.8196]^T$ mm (*center*), and $[15.2672 \ 1.3194 \ 1.2827 \ 0.5165 \ 1.0004]^T$ mm (*bottom*)

late fine-discretization training data for \mathbf{R}_s for $\beta = 0.05$ (i.e., the model used for the optimization) was approximately 49 h; for $\mathbf{R}_{s,\text{full}}$ it was 80 h.

To further explore the influence of mesh density on our method, a second coarse model \mathbf{R}_{cc} ($\sim 8,000$ mesh cells, simulation time 8 s), i.e., coarser than \mathbf{R}_c , was generated, and corresponding surrogate models constructed. The predictive results for the new \mathbf{R}_s were similar to previous results; e.g., for $\beta = 0.2$ the predictive %RMSE was 5.82 %, although the number of SVs increased somewhat to 3,992 (see Table 1). In order to evaluate the general similarity between the coarsely and finely simulated data, Pearson product-moment correlation coefficients were computed for the respective $|S_{11}|$ values, i.e., for $|S_{11}|$ of \mathbf{R}_c and \mathbf{R}_f ; and also for \mathbf{R}_{cc} and \mathbf{R}_f (using all training geometries with 121 equally spaced frequency points per geometry). The correlation coefficients were 0.74 and 0.51, respectively, suggesting some robustness to our procedure.

4.4 Application Examples: Antenna Optimization

The full and reduced BSVR models were used to perform design optimization of the antenna structures considered in Sects. 4.1 through 4.3. We again note that our models are intended as multipurpose global models that give accurate predictions for the whole of the input space; multiple optimization runs corresponding to any number of sets of design specifications constitute one kind of repeated-use application. The initial design in each case is the center of the region of interest $\mathbf{x}^{(0)}$. The design process starts by directly optimizing the BSVR model. Because of some limitations in the accuracy of the models given the design context (linear responses

were modeled—the preferred choice given the Gaussian kernel—but logarithmic responses (in decibels) are optimized), the design is further refined by means of the SM iterative process [14]

$$\mathbf{x}^{(i+1)} = \arg \min_{\mathbf{x}} U(\mathbf{R}_{\text{su}}^{(i)}(\mathbf{x})), \quad (8)$$

where $\mathbf{R}_{\text{su}}^{(i)}$ is a surrogate model, enhanced by frequency and output SM [14]. The surrogate model setup is performed using an evaluation of \mathbf{R}_f at $\mathbf{x}^{(i)}$. U implements design specifications. For simplicity, we use the symbol \mathbf{R}_{co} to denote either of $\mathbf{R}_{s,\text{full}}$ or \mathbf{R}_s , which can be considered the “coarse” models in the SM context. Let $\mathbf{R}_{\text{co}}(\mathbf{x}, F)$ denote the explicit dependency of the model on the frequency (F is the set of frequencies of interest at which the model is evaluated). The surrogate model is defined as

$$\mathbf{R}_{\text{su}}^{(i)}(\mathbf{x}) = \mathbf{R}_{\text{co}}(\mathbf{x}, \boldsymbol{\alpha}^{(i)} F) + \mathbf{d}^{(i)} \quad (9)$$

with

$$\mathbf{d}^{(i)} = \mathbf{R}_f(\mathbf{x}^{(i)}) - \mathbf{R}_{\text{co}}(\mathbf{x}^{(i)}, \boldsymbol{\alpha}^{(i)} F) \quad (10)$$

and

$$\boldsymbol{\alpha}^{(i)} F = \alpha_0^{(i)} + \alpha_1^{(i)} F \quad (11)$$

the affine frequency scaling (shift and scaling). The frequency scaling parameters are calculated as

$$[\alpha_0^{(i)} \alpha_1^{(i)}] = \arg \min_{\mathbf{x}} \|\mathbf{R}_f(\mathbf{x}^{(i)}) - \mathbf{R}_{\text{co}}(\mathbf{x}^{(i)}, \alpha_0^{(i)} + \alpha_1^{(i)} F)\| \quad (12)$$

i.e., to minimize the misalignment between the high-fidelity and the scaled low-fidelity model responses at $\mathbf{x}^{(i)}$. Although the models are evaluated at a discrete set of frequencies, the information at other frequencies can be obtained through interpolation. The misalignment is further reduced by the output SM (10); this ensures zero-order consistency (i.e., $\mathbf{R}_{\text{su}}^{(i)}(\mathbf{x}^{(i)}) = \mathbf{R}_f(\mathbf{x}^{(i)})$) between the surrogate and \mathbf{R}_f [18]. The algorithm (8) working with the SM surrogate model (9)–(12) typically requires only three to four iterations to yield an optimized design, with the cost of each iteration effectively equal to a single evaluation of the high-fidelity model.

Figure 5 shows the responses of the reduced BSVR and fine models at the initial designs as well as the responses of the fine models at the final designs obtained for both antenna structures. The reduced BSVR models correspond to β values in Table 1 of 0.15 (Antenna 1), 0.05 (Antenna 2), and 0.15 (Antenna 3). Table 2 summarizes the results. One can see that the design quality and cost (expressed in terms of number of \mathbf{R}_f evaluations) are very similar for the BSVR models obtained using full and reduced data sets (the CPU times associated with three \mathbf{R}_f evaluations (Antennas 1 and 2) and four \mathbf{R}_f evaluations (Antenna 3) were 36 min, 63 min, and 48 min, respectively).

For comparison, we also optimized the three antennas using a conventional (not surrogate-based) method, namely a state-of-the-art pattern search algorithm [19, 20] that directly relied on fine-discretization full-wave simulations (i.e., \mathbf{R}_f) for its objective function evaluations. While the maximum $|S_{11}|$ values at the final designs

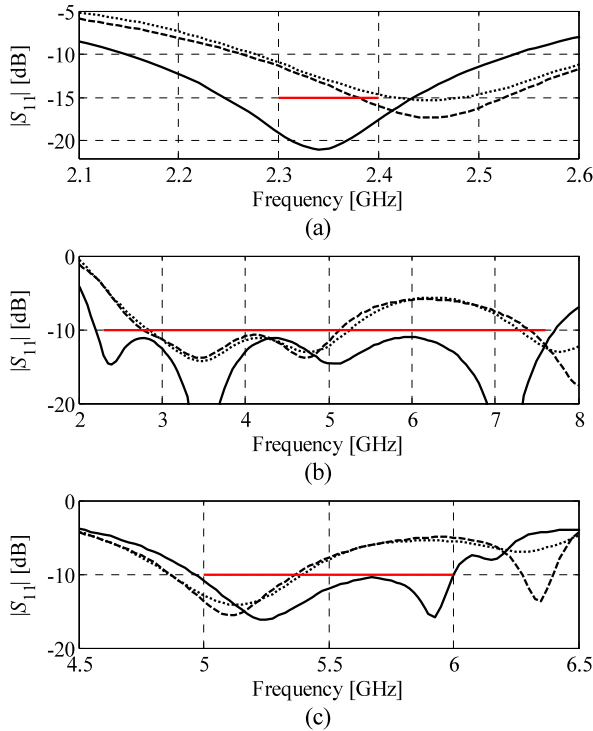


Fig. 5 Optimization results: responses of the BSVR model with reduced data set (\cdots), the fine model at the initial design ($---$), and the fine model at the optimized design ($—$) for (a) Antenna 1, (b) Antenna 2, and (c) Antenna 3. Design specifications are marked with *horizontal solid line*

Table 2 Results of antenna optimizations

Antenna	Model	max $ S_{11} $ at final design ^a	Optimization cost ^b
1	$R_{s,\text{full}}$	−18.3 dB	3
	R_s	−17.6 dB	3
2	$R_{s,\text{full}}$	−11.0 dB	3
	R_s	−11.0 dB	3
3	$R_{s,\text{full}}$	−10.2 dB	4
	R_s	−10.2 dB	4

^amax $|S_{11}|$ at the frequency band of interest: 2.3 to 2.4 GHz (Antenna 1), 2.3 to 7.7 GHz (Antenna 2), and 5 to 6 GHz (Antenna 3)

^bNumber of R_f evaluations including evaluation at the initial design

obtained for Antennas 1, 2, and 3 (−21.6 dB, −11.6 dB, and −10.7 dB, respectively) were similar to those obtained using our BSVR models and the above SM procedure, the computational expense for the conventional optimization was at least

an order of magnitude larger (i.e., 40, 148, and 201 \mathbf{R}_f evaluations for Antennas 1, 2, and 3, respectively, compared to the 3, 3, and 4 \mathbf{R}_f evaluations reported in Table 2). This accentuates how much faster optimization can be realized when accurate models such as BSVR models are available: our approach reduces by up to 43 % the high initial cost of setting up these multipurpose global models (in comparison the cost of the optimization using SM is insignificant).

5 Modeling and Optimization of Filters Using BSVR

Here we discuss global BSVR models based on reduced finely discretized data sets for the $|S_{21}|$ responses of two microstrip filters: a capacitively coupled dual-behavior resonator microstrip bandpass filter, and an open-loop ring resonator (OLRR) bandpass filter. As before, we use these models for design optimization.

5.1 *Capacitively Coupled Dual-Behavior Resonator (CCDBR) Microstrip Bandpass Filter*

Consider the capacitively coupled dual-behavior resonator (CCDBR) bandpass filter [21] implemented in microstrip lines, shown in Fig. 6(a). The three design variables were $\mathbf{x} = [L_1 \ L_2 \ L_3]^T$. The design variable space for the BSVR models was defined by the center vector $\mathbf{x}^0 = [3 \ 5 \ 1.5]^T$ mm and size vector $\boldsymbol{\delta} = [1 \ 1 \ 0.5]^T$ mm such that the variable ranges were $\mathbf{x}^0 \pm \boldsymbol{\delta}$ mm (\mathbf{x}^0 and $\boldsymbol{\delta}$ were guesses, guided to some extent by expert knowledge of the filters and a very rudimentary exploration of the design space). The substrate height was $h = 0.254$ mm and the relative permittivity was $\epsilon_r = 9.9$; the value of S was 0.05 mm, while the microstrip line widths w_1 and w_2 were 0.25 mm and 0.5 mm, respectively. We were interested in the filter response over the frequency range 2 to 6 GHz. The high-fidelity model \mathbf{R}_f of the filter was simulated using FEKO [22] (total mesh number 715, simulation time about 15 s per frequency). The low-fidelity model \mathbf{R}_c was also simulated in FEKO (total mesh number 136, simulation time 0.6 s per frequency).

In order to set up the training data input vectors, 400 geometries were randomly selected from the design space using Latin hypercube sampling (LHS) [23]. For each geometry, 12 simulation frequencies were selected by uniform random sampling from the above frequency range, yielding a total of $n = 400 \times 12 = 4,800$ training input vectors of the form $\{\mathbf{x}_i = [L_{1i} \ L_{2i} \ L_{3i} \ f_i]^T | i = 1, \dots, n\}$, with L_{1i} , L_{2i} , and L_{3i} the design variables corresponding to the i th input vector, and f_i a frequency value within the range of interest. The corresponding output scalars, obtained from FEKO simulations, were $y_i = |S_{21i}|$. The test data were likewise obtained from 50 new geometries, also obtained via LHS, with 41 equally spaced frequencies per geometry. The training data were simulated at the \mathbf{R}_c mesh density and used to train the BSVR model $\mathbf{R}_{s,\text{aux}}$ for $\beta = 0.1, 0.2, \text{ and } 0.3$ (β is the loss function parameter

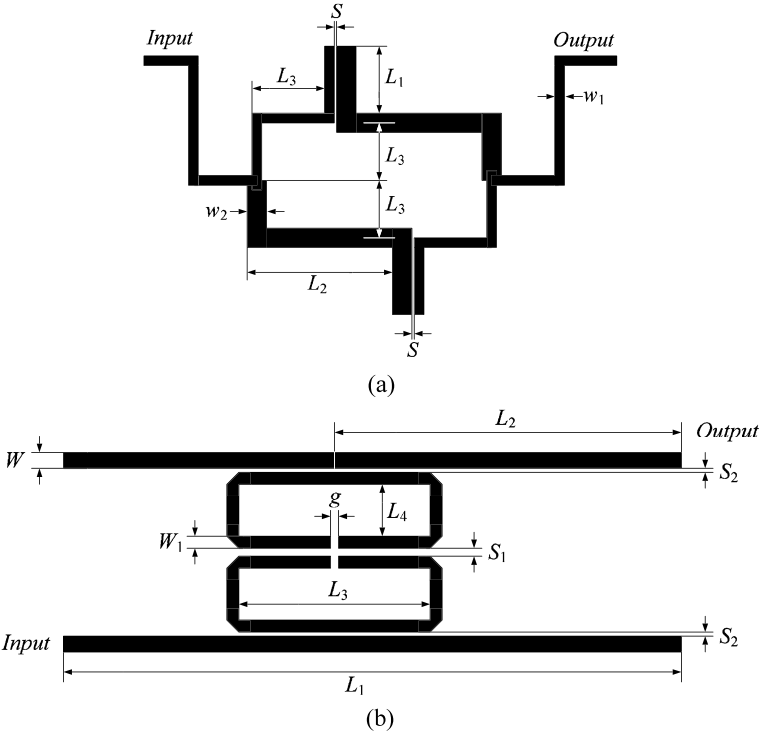


Fig. 6 Geometry of the (a) CCDBR bandpass filter [21] and (b) OLRR bandpass filter [24]

in Eq. (3) discussed above). $\mathbf{R}_{s,\text{aux}}$ was used to make predictions on the test data (also simulated at the \mathbf{R}_c mesh density).

The %RMSE (root mean square error normalized to the target range expressed as a percentage) was around 4.13 % for the three β values, which was acceptable for this highly nonlinear problem, and indicated that the training set was sufficiently large. Next, for each value of β the n_{SV} SVs of $\mathbf{R}_{s,\text{aux}}$ were simulated at the \mathbf{R}_f mesh density. BSVR models fitted to these reduced training sets gave the desired surrogate models \mathbf{R}_s .

For comparison, surrogates $\mathbf{R}_{s,\text{full}}$ trained on the full fine-discretization training data ($n = 4,800$) were also set up. Table 3 gives, for the three β values, the %RMSEs obtained with $\mathbf{R}_{s,\text{aux}}$ on the coarse test data, and with \mathbf{R}_s and $\mathbf{R}_{s,\text{full}}$ on the fine test data; as well as the number of SVs obtained in each instance. The highly similar %RMSEs obtained with \mathbf{R}_s and $\mathbf{R}_{s,\text{full}}$ indicate that reducing the number of expensive fine-discretization training points from n to n_{SV} incurred negligible accuracy loss, even for a reduction in data as large as 48 % ($\beta = 0.1$). Figure 7(a) shows typical predictive results for $|S_{21}|$ versus frequency obtained for the test geometry $\mathbf{x} = [2.794 \ 4.407 \ 1.491]^T$ mm. Some discrepancy can be observed when comparing the RMSE values for $\mathbf{R}_{s,\text{aux}}$ in Table 3 to those for \mathbf{R}_s (and $\mathbf{R}_{s,\text{full}}$). This occurs because the coarse model responses are slightly smoother as functions of fre-

Table 3 Predictive errors of surrogate filter models. $R_{s,aux}$ are BSVR models trained on all n data points obtained from coarsely discretized simulations, $R_{s,full}$ are BSVR models trained on all n data points obtained from finely discretized simulations, and R_s are BSVR models trained on the n_{SV} finely discretized data points corresponding to SVs of $R_{s,aux}$

Filter	β	RMSE (%)			n_{SV}	n_{SV}/n (%)
		$R_{s,aux}$	R_s	$R_{s,full}$		
CCDBR ($n = 4,800$)	0.1	4.17	6.05	5.98	2,488	52
	0.2	4.12	5.87	5.95	2,563	53
	0.3	4.11	5.95	5.86	2,886	60
OLRR ($n = 4,800$)	0.1	4.12	4.26	4.21	2,360	49
	0.2	3.77	4.47	4.32	2,565	53
	0.3	3.46	3.90	3.64	2,744	57

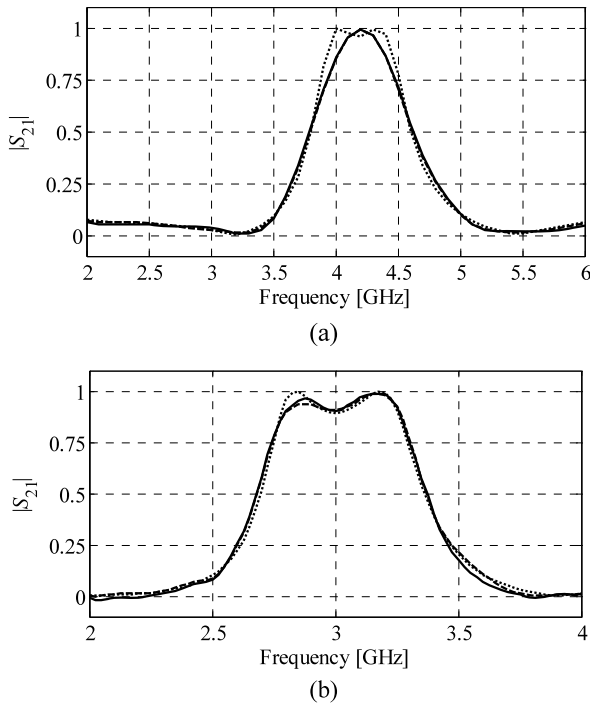


Fig. 7 Predictive response of BSVR models R_s (—) and $R_{s,full}$ (---), and high-fidelity model R_f (····) for (a) the CCDBR bandpass filter test geometry $x = [2.794 \ 4.407 \ 1.491]^T$ mm, and (b) the OLRR bandpass filter test geometry $x = [38.088 \ 8.306 \ 5.882 \ 4.029 \ 0.193 \ 0.061 \ 0.985]^T$ mm

quency (i.e., they do not contain as much detail particularly in the passband) than the fine model ones, which make them easier to model; this results in a lower value of %RMSE. Using a finer mesh for the R_c model would reduce this discrepancy. The

RMSE values for \mathbf{R}_s and $\mathbf{R}_{s,\text{full}}$ nevertheless were good given the highly nonlinear nature of the modeling problem, and of sufficient accuracy to yield good optimization results, as we show in Sect. 4. The total computational time necessary to gather the training data for setting up $\mathbf{R}_{s,\text{full}}$ was 20 h, whereas the corresponding time for setting \mathbf{R}_s (including both low- and high-fidelity model evaluations) was 11.2, 11.4, and 12.8 h (for $\beta = 0.1, 0.2,$ and $0.3,$ respectively) on a quad core PC with a 2.66 GHz Intel processor and 4 GB RAM. Thus the computational savings due to the proposed technique vary between 36 % (for $\beta = 0.3$) to 44 % (for $\beta = 0.1$).

5.2 Open-Loop Ring Resonator (OLRR) Bandpass Filter

The filter geometry [24] is shown in Fig. 6(b). The seven design variables were $\mathbf{x} = [L_1 \ L_2 \ L_3 \ L_4 \ S_1 \ S_2 \ G]^T$. The design space was described by $\mathbf{x}^0 = [40 \ 8 \ 6 \ 4 \ 0.2 \ 0.1 \ 1]^T$ mm and $\boldsymbol{\delta} = [2 \ 1 \ 0.4 \ 0.4 \ 0.1 \ 0.05 \ 0.2]^T$ mm. The substrate parameters were $h = 0.635$ mm and $\varepsilon_r = 10.2$, while the microstrip line widths were $W_1 = 0.4$ mm and $W = 0.6$ mm. The frequency range of interest was 2 to 4 GHz. High- and low-fidelity models were simulated in FEKO [22] (total mesh number 1,084 and simulation time 40 s per frequency for \mathbf{R}_f ; total mesh number 148 and simulation time 0.8 s per frequency for \mathbf{R}_c). The training data comprised 400 geometries obtained by LHS [23], with 12 randomly selected frequencies per geometry ($n = 4,800$), while the test data were 50 new LHS geometries with 81 equally spaced frequencies per geometry. Setting up $\mathbf{R}_{s,\text{aux}}, \mathbf{R}_s,$ and $\mathbf{R}_{s,\text{full}}$ proceeded in a manner similar to the earlier filter. Table 3 gives the %RMSE values obtained with $\mathbf{R}_{s,\text{aux}}$ on the coarse test data and with \mathbf{R}_s and $\mathbf{R}_{s,\text{full}}$ on the fine test data; as well as the SV counts. In general, %RMSE values of \mathbf{R}_s were only marginally higher than those of $\mathbf{R}_{s,\text{full}}$, suggesting as before that reducing the number of training points from n to n_{SV} by using the SVs of $\mathbf{R}_{s,\text{aux}}$ as fine-discretization training points for \mathbf{R}_s has little effect on the prediction accuracy. The greatest reduction in data (51 %) was obtained for $\beta = 0.1$. Figure 7(b) shows representative predictive results for $|S_{21}|$ versus frequency, in particular for the test geometry $\mathbf{x} = [38.088 \ 8.306 \ 5.882 \ 4.029 \ 0.193 \ 0.061 \ 0.985]^T$ mm. The computational time necessary to gather the training data for setting up $\mathbf{R}_{s,\text{full}}$ was 53.3 h. The corresponding time for setting up \mathbf{R}_s (including both low- and high-fidelity model evaluations) was 27.2, 29.3, and 31.4 h ($\beta = 0.1, 0.2,$ and $0.3,$ respectively). The computational savings due to the proposed technique range from 41 % (for $\beta = 0.3$) to 49 % (for $\beta = 0.1$).

5.3 Filter Optimization Using BSVR Surrogates

The BSVR developed in Sects. 5.1 and 5.2 is used to perform design optimization of the filters. The optimization methodology is essentially the same as described in

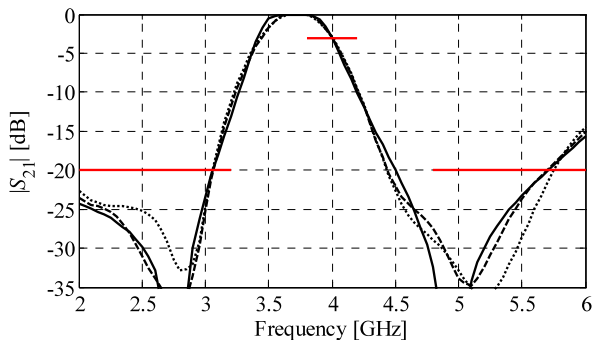


Fig. 8 CCDBR filter: responses of the high-fidelity model (—) and BSVR models $\mathbf{R}_{s,\text{full}}$ (- - -) and \mathbf{R}_s (· · ·) at the initial design $\mathbf{x}^{(0)}$. Design specification as per *horizontal solid (red) lines*

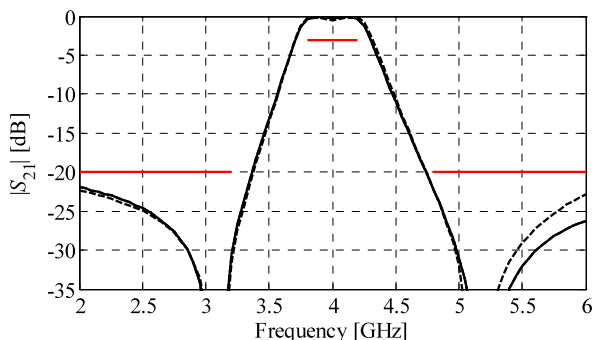


Fig. 9 CCDBR filter: responses of the high-fidelity model at the optimized designs found using $\mathbf{R}_{s,\text{full}}$ (—) and \mathbf{R}_s (- - -). Design specification as per *horizontal solid (red) lines*

Sect. 4.4 and involves iterative correction and optimization of the surrogates in order to (locally) improve their accuracy [25].

The design process starts by directly optimizing the BSVR model (for each filter, we used the BSVR surrogates corresponding to $\beta = 0.1$). Each iteration (see Eq. (8)) requires only one evaluation of the high-fidelity model. The CCDBR bandpass filter had design specifications $|S_{21}| \geq -3$ dB for $3.8 \leq f \leq 4.2$ GHz; and $|S_{21}| \leq -20$ dB for $2.0 \leq f \leq 3.2$ GHz, and $4.8 \leq f \leq 6.0$ GHz (f denotes frequency). Figure 8 shows the responses of the high-fidelity model \mathbf{R}_f as well as the responses of \mathbf{R}_s and $\mathbf{R}_{s,\text{full}}$ at the initial design $\mathbf{x}^{(0)} = [3 \ 6 \ 1.5]^T$ mm. The high-fidelity model responses at the optimized designs found using both \mathbf{R}_s and $\mathbf{R}_{s,\text{full}}$ are shown in Fig. 9 (these designs were $[3.47 \ 4.79 \ 1.01]^T$ mm and $[3.21 \ 4.87 \ 1.22]^T$ mm, respectively). In both cases, the design process is accomplished in three iterations, which correspond to a design cost of four high-fidelity model evaluations.

For the OLRB bandpass filter, the design specifications were $|S_{21}| \geq -1$ dB for $2.85 \leq f \leq 3.15$ GHz, and $|S_{21}| \leq -20$ dB for $2.0 \leq f \leq 2.5$ GHz and

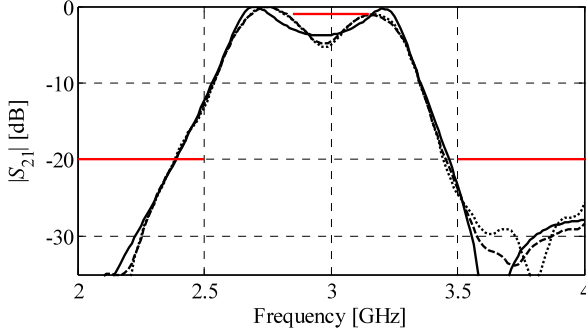


Fig. 10 OLRR bandpass filter: responses of the high-fidelity model (—) and BSVR models $\mathbf{R}_{s,\text{full}}$ (- - -) and \mathbf{R}_s (· · ·) at the initial design $\mathbf{x}^{(0)}$. Design specification as per *horizontal solid (red) lines*

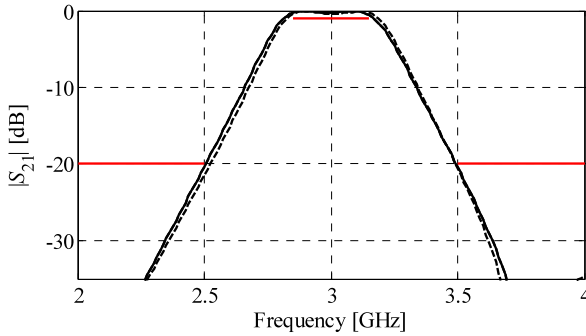


Fig. 11 OLRR bandpass filter: responses of the high-fidelity model at the optimized designs found using $\mathbf{R}_{s,\text{full}}$ (—) and \mathbf{R}_s (- - -). Design specification as per *horizontal solid (red) lines*

$3.5 \leq f \leq 4.0$ GHz. Figure 10 shows the responses of \mathbf{R}_f , \mathbf{R}_s , and $\mathbf{R}_{s,\text{full}}$ at the initial design $\mathbf{x}^{(0)} = [40.0 \ 8.0 \ 6.0 \ 4.0 \ 0.1 \ 0.1 \ 1.0]^T$ mm. The responses at the optimized designs obtained using \mathbf{R}_s and $\mathbf{R}_{s,\text{full}}$ are shown in Fig. 11 (these designs were $[39.605 \ 8.619 \ 6.369 \ 3.718 \ 0.300 \ 0.069 \ 0.986]^T$ mm and $[39.010 \ 8.219 \ 5.786 \ 4.260 \ 0.268 \ 0.050 \ 1.068]^T$ mm, respectively). In both cases, the design process is accomplished in two iterations, which correspond to the design cost of three high-fidelity model evaluations.

6 Variable-Fidelity Optimization Using Local BSVR Surrogates

In this section, we discuss the application of BSVR surrogates defined locally (i.e., in a specific region of the input space) towards low-cost design optimization of antennas [26].

As we have shown, BSVR models can be accurate; however, similarly to other approximation-based modeling methods, considerable computational overhead is necessary to acquire the training data. This is not convenient when using approximation surrogates for *ad hoc* optimization of a specific structure.

Here, we describe a computationally efficient antenna design methodology that combines space mapping (as the optimization engine), and coarse model response surface approximation implemented through BSVR. BSVR serves to create a fast coarse model of the antenna structure. In order to reduce the computational cost of creating the latter, it is obtained from coarse discretization EM simulation data.

6.1 Optimization Algorithm

As mentioned before, the main optimization engine is space mapping (SM) [26]. The generic SM optimization algorithm produces a sequence of approximate solutions to the problem (1), $\mathbf{x}^{(0)}, \mathbf{x}^{(1)}, \dots$, as follows:

$$\mathbf{x}^{(i+1)} = \arg \min_{\mathbf{x}} U(\mathbf{R}_s^{(i)}(\mathbf{x})), \quad (13)$$

where $\mathbf{R}_s^{(i)}$ is the SM surrogate model at iteration i . $\mathbf{x}^{(0)}$ is the initial design. The surrogate model is constructed using the underlying coarse model \mathbf{R}_c and suitable auxiliary transformations [27]. The elementary SM transformations include input SM [28] with the surrogate defined as $\mathbf{R}_s^{(i)}(\mathbf{x}) = \mathbf{R}_c(\mathbf{x} + \mathbf{c}^{(i)})$, multiplicative output SM [27], $\mathbf{R}_s^{(i)}(\mathbf{x}) = \mathbf{A}^{(i)} \cdot \mathbf{R}_c(\mathbf{x})$, additive output SM [28], $\mathbf{R}_s^{(i)}(\mathbf{x}) = \mathbf{R}_c(\mathbf{x}) + \mathbf{d}^{(i)}$, and frequency scaling [27], $\mathbf{R}_s^{(i)}(\mathbf{x}) = \mathbf{R}_{c,f}(\mathbf{x}; \mathbf{F}^{(i)})$. In frequency SM, it is assumed that the coarse model is an evaluation of a given performance parameter over a certain frequency range, i.e., $\mathbf{R}_c(\mathbf{x}) = [R_c(\mathbf{x}; \omega_1) \dots R_c(\mathbf{x}; \omega_m)]^T$, and the frequency-scaled model is then given as $\mathbf{R}_{c,f}(\mathbf{x}; \mathbf{F}^{(i)}) = [R_c(\mathbf{x}; s(\omega_1)) \dots R_c(\mathbf{x}; s(\omega_m))]^T$, where s is a scaling function dependent on the set of parameters $\mathbf{F}^{(i)}$. Typically, a linear scaling function $s(\omega) = f_0^{(i)} + f_1^{(i)}\omega$ is used.

Parameters of SM transformations are obtained using the parameter extraction (PE) process, which, in the case of input SM, takes the form

$$\mathbf{c}^{(i)} = \arg \min_{\mathbf{c}} \sum_{k=0}^i \|\mathbf{R}_f(\mathbf{x}^{(k)}) - \mathbf{R}_c(\mathbf{x}^{(k)} + \mathbf{c})\|. \quad (14)$$

Formulation of PE for the other transformations is similar [28].

Because PE and surrogate model optimization may require a large number of coarse model evaluations, it is beneficial that \mathbf{R}_c is fast, which is usually not possible for antenna structures, where the only universally available (and yet accurate) type of coarse model is the output of coarse-discretization EM simulations. To alleviate this problem, we construct the coarse model by using a fixed number of such low-fidelity simulations as training data for the coarse model, so that, once set up, the coarse model can be used by the SM algorithm without further reference to the EM solver.

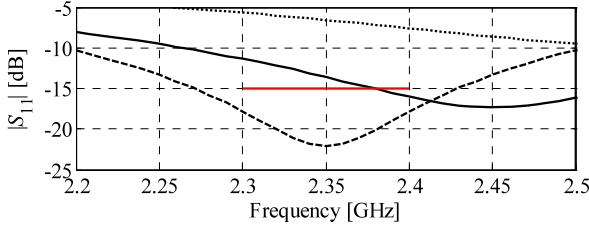


Fig. 12 CPW-fed slot dipole antenna: responses of the low- (· · ·) and high- (—) fidelity models at the initial design \mathbf{x}^{init} , and the response of the low-fidelity model at its approximate optimum $\mathbf{x}^{(0)}$ (- -)

In order to improve the convergence properties of the algorithm, it is embedded in the trust region framework [29], so that the new design $\mathbf{x}^{(i+1)}$ is found only in the vicinity of the current one, $\mathbf{x}^{(i)}$, as follows:

$$\mathbf{x}^{(i+1)} = \arg \min_{\mathbf{x}, \|\mathbf{x} - \mathbf{x}^{(i)}\| \leq \delta^{(i)}} U(\mathbf{R}_s^{(i)}(\mathbf{x})), \quad (15)$$

where $\delta^{(i)}$ is the trust region radius updated in each iteration according to the standard rules [29]. Within this framework, the designs that do not improve the specification error are rejected, and the search is repeated with the reduced value of $\delta^{(i)}$.

The steps in the modeling procedure are as follows. First, we find an approximate optimum of the coarse model \mathbf{R}_{cd} (i.e., the low-fidelity full-wave EM simulations). Then we construct a BSVR surrogate \mathbf{R}_c , only using relatively densely spaced training data in the vicinity of this optimum, thus further enhancing the computational efficiency. Once constructed, \mathbf{R}_c is used as the basis for the iterative optimization (13)–(15).

6.2 Antenna Optimization Examples

As examples, we consider again the CPW-fed slot dipole antenna of Sect. 4.1 (Antenna 1, Fig. 1) and the CPW-fed T-shaped slot antenna of Sect. 4.2 (Antenna 2, Fig. 2).

For Antenna 1, we have two design variables, $\mathbf{x} = [W \ L]^T$ mm. Using CST Microwave Studio [15], we consider a high-fidelity model \mathbf{R}_f ($\sim 130,000$ mesh cells, simulation time 12 min), and a low-fidelity model \mathbf{R}_{cd} ($\sim 5,000$ mesh cells, simulation time 30 s). The initial design is $\mathbf{x}^{\text{init}} = [7.5 \ 39.0]^T$ mm. The starting point of the SM optimization is the approximate optimum of the low-fidelity model, $\mathbf{x}^{(0)} = [5.0 \ 43.36]^T$ mm, found using a pattern search algorithm [19]. The computational cost of this step was 26 evaluations of \mathbf{R}_{cd} . The BSVR coarse model has been created using 100 low-fidelity model samples allocated using LHS [23] in the vicinity of $\mathbf{x}^{(0)}$ defined by deviation $\mathbf{d} = [1 \ 3]^T$ mm. The size of this vicinity must be sufficiently large to allow the coarse model to “absorb” the misalignment between the low- and high-fidelity models at $\mathbf{x}^{(0)}$ through appropriate SM transformations.

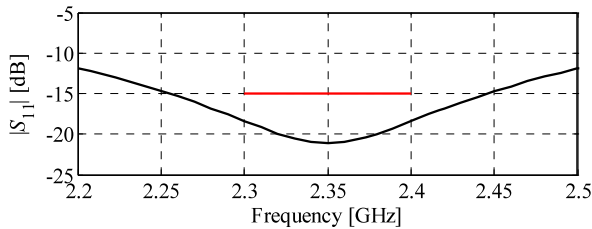


Fig. 13 CPW-fed slot dipole antenna: response of the high-fidelity model at the final design $\mathbf{x}^{(4)}$

Table 4 CPW-fed slot dipole antenna: design cost

Algorithm component	Number of model evaluations	Evaluation time	
		Absolute [min]	Relative to \mathbf{R}_f
Optimization of \mathbf{R}_{cd}	$26 \times \mathbf{R}_{cd}$	13	1.1
Creating BSVR coarse model	$100 \times \mathbf{R}_{cd}$	50	4.2
Evaluation of \mathbf{R}_f^a	$5 \times \mathbf{R}_f$	60	5.0
Total optimization time	N/A	123	10.3

^aIncludes evaluation of \mathbf{R}_f at $\mathbf{x}^{(0)}$

The low-fidelity model at the initial design, as well as the low- and high-fidelity model responses at $\mathbf{x}^{(0)}$, are shown in Fig. 12. In this case, the major discrepancy between the models is a frequency shift. Therefore, the primary SM transformation used for this example is frequency scaling applied to all designs, $\mathbf{x}^{(0)}, \mathbf{x}^{(1)}, \dots, \mathbf{x}^{(i)}$, considered during the optimization run. The SM surrogate is then enhanced using a local additive output SM [27] so that the entire SM model has the form $\mathbf{R}_s^{(i)}(\mathbf{x}) = \mathbf{R}_{c,f}(\mathbf{x}; \mathbf{F}^{(i)}) + [\mathbf{R}_f(\mathbf{x}^{(i)}) - \mathbf{R}_c(\mathbf{x}^{(i)})]$.

The final design, $\mathbf{x}^{(4)} = [5.00 \ 41.56]^T$ mm, is obtained in four SM iterations. The high-fidelity model response at $\mathbf{x}^{(4)}$ is shown in Fig. 13. At this design, we have $|S_{11}| \leq -18.3$ dB over the entire frequency band of interest. The design cost is summarized in Table 4 and corresponds to about ten evaluations of the high-fidelity model. Figure 14 shows the convergence plot and the evolution of the specification error versus iteration index.

For Antenna 2, the design variables are $\mathbf{x} = [a_x \ a_y \ a \ b]^T$ mm. The design specifications are $|S_{11}| \leq -12$ dB for 2.3 to 7.6 GHz. The high-fidelity model \mathbf{R}_f is evaluated with the CST MWS transient solver [15] (3,556,224 mesh cells, simulated in 60 min). The low-fidelity model \mathbf{R}_{cd} is also evaluated in CST MWS but with a coarser mesh: 110,208 mesh cells, simulation time 1.5 min. The initial design is $\mathbf{x}^{\text{init}} = [40 \ 30 \ 10 \ 20]^T$ mm. The approximate low-fidelity model optimum, $\mathbf{x}^{(0)} = [40.33 \ 25.6 \ 8.4 \ 20.8]^T$ mm, has been found using a pattern search algorithm, at the cost of 85 evaluations of \mathbf{R}_{cd} . The BSVR coarse model has been created using 100 low-fidelity model samples allocated using LHS in the vicinity of $\mathbf{x}^{(0)}$ defined by deviation $\mathbf{d} = [2 \ 2 \ 1 \ 1]^T$ mm. The BSVR model \mathbf{R}_c was subsequently used as a

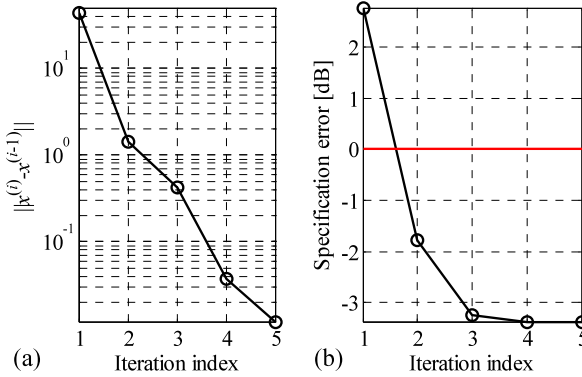


Fig. 14 CPW-fed slot dipole antenna: (a) convergence plot, (b) specification error versus iteration index

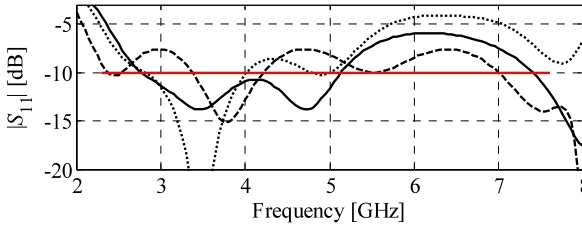


Fig. 15 Broadband CPW-fed T-shaped slot antenna: responses of the low- (\cdots) and high- (—) fidelity models at the initial design \mathbf{x}^{init} and the response of the low-fidelity model at its approximate optimum $\mathbf{x}^{(0)}$ (- - -)

coarse model for the SM algorithm. Figure 15 shows the low-fidelity model at the initial design, as well as the low- and high-fidelity model responses at $\mathbf{x}^{(0)}$. Because the major discrepancy between the models is a vertical shift, the primary SM transformation used for this example is the multiplicative response correction applied to all designs, $\mathbf{x}^{(0)}, \mathbf{x}^{(1)}, \dots, \mathbf{x}^{(i)}$, considered during the optimization run. The SM surrogate is then enhanced using an additive output SM with the SM model having the form $\mathbf{R}_s^{(i)}(\mathbf{x}) = \mathbf{A}^{(i)} \cdot \mathbf{R}_c(\mathbf{x}) + [\mathbf{R}_f(\mathbf{x}^{(i)}) - \mathbf{R}_c(\mathbf{x}^{(i)})]$.

The final design, $\mathbf{x}^{(7)} = [39.84 \ 24.52 \ 8.84 \ 21.40]^T$ mm, is obtained in seven SM iterations. The high-fidelity model response at $\mathbf{x}^{(7)}$ is shown in Fig. 16. At that design, we have $|S_{11}| \leq -10.9$ dB for 2.3 GHz to 7.6 GHz. The overall design cost is summarized in Table 5 and corresponds to about 15 evaluations of the high-fidelity model. The convergence of the algorithm (Fig. 17) is consistent with that for the previous example.

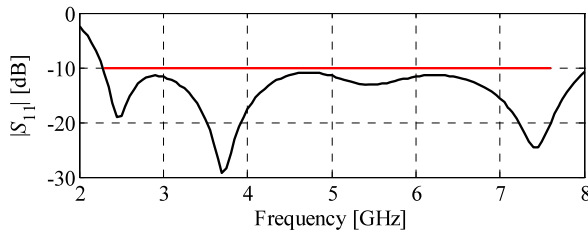


Fig. 16 Broadband CPW-fed T-shaped slot antenna: response of the high-fidelity model at the final design $\mathbf{x}^{(7)}$

Table 5 Broadband CPW-fed T-shaped slot antenna: design cost

Algorithm component	Number of model evaluations	Evaluation time	
		Absolute [min]	Relative to \mathbf{R}_f
Optimization of \mathbf{R}_{cd}	$85 \times \mathbf{R}_{cd}$	128	2.1
Creating BSVR coarse model	$100 \times \mathbf{R}_{cd}$	150	2.5
Evaluation of \mathbf{R}_f^a	$10 \times \mathbf{R}_f$	600	10.0
Total optimization time	N/A	878	14.6

^aIncludes evaluation of \mathbf{R}_f at $\mathbf{x}^{(0)}$

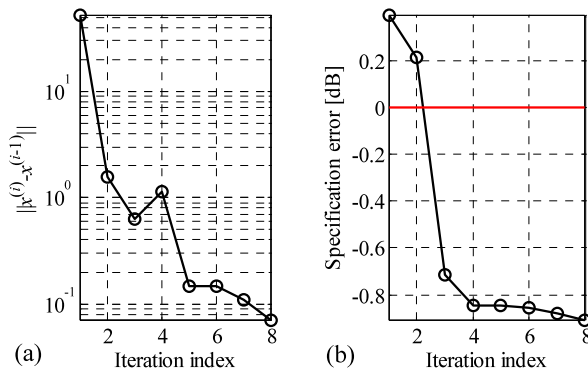


Fig. 17 Broadband CPW-fed T-shaped slot antenna: (a) convergence plot, (b) specification error versus iteration index

7 Conclusion

In this chapter, we presented a Bayesian support vector regression methodology for accurate modeling of microwave components and structures. We demonstrated the possibility of reduction of the number of fine-discretization training points by performing BSVR modeling on coarse-discretization EM simulation data (selected by standard experimental design) and then obtaining high-fidelity simulations only for

the points that contribute to the initial BSVR model in a nontrivial way. The computational savings thus obtained had little effect on the modeling accuracy. We have also demonstrated that the reduced-training-set BSVR models perform as well as the full-training-set models in parametric optimization of antenna structures. A notable advantage of BSVR is that only a single parameter must be set by the user, namely β (hyperparameters are initialized randomly during training). This is in contrast to, for instance, neural network-based methodologies for regression, which might require the tuning of a variety of architectural/performance parameters (e.g., number of hidden units, learning rate, momentum). We also discussed the use of BSVR surrogates for variable-fidelity design optimization of antennas, where the main optimization engine is space mapping, whereas the underlying coarse model is obtained by approximating low-fidelity EM simulation data. As a result, the optimization can be accomplished at a low computational cost corresponding to a few evaluations of the high-fidelity EM simulations of the structure under design.

References

1. Angiulli, G., Cacciola, M., Versaci, M.: Microwave devices and antennas modelling by support vector regression machines. *IEEE Trans. Magn.* **43**, 1589–1592 (2007)
2. Jacobs, J.P.: Bayesian support vector regression with automatic relevance determination kernel for modeling of antenna input characteristics. *IEEE Trans. Antennas Propag.* **60**, 2114–2118 (2012)
3. Chu, W., Keerthi, S.S., Ong, C.J.: Bayesian support vector regression using a unified loss function. *IEEE Trans. Neural Netw.* **15**, 29–44 (2004)
4. Rasmussen, C.E., Williams, C.K.I.: *Gaussian Processes for Machine Learning*. MIT Press, Cambridge (2006)
5. Devabhaktuni, V.K., Yagoub, M.C.E., Zhang, Q.J.: A robust algorithm for automatic development of neural network models for microwave applications. *IEEE Trans. Microw. Theory Tech.* **49**, 2282–2291 (2001)
6. Couckuyt, I., Declercq, F., Dhaene, T., Rogier, H., Knockaert, L.: Surrogate-based infill optimization applied to electromagnetic problems. *Int. J. RF Microw. Comput.-Aided Eng.* **20**, 492–501 (2010)
7. Tokan, N.T., Gunes, F.: Knowledge-based support vector synthesis of the microstrip lines. *Prog. Electromagn. Res.* **92**, 65–77 (2009)
8. Schölkopf, B., Smola, A.J.: *Learning with Kernels: Support Vector Machines, Regularization, Optimization, and Beyond*. MIT Press, Cambridge (2002)
9. Bandler, J.W., Georgieva, N., Ismail, M.A., Rayas-Sánchez, J.E., Zhang, Q.J.: A generalized space mapping tableau approach to device modeling. *IEEE Trans. Microw. Theory Tech.* **49**, 67–79 (2001)
10. Rayas-Sánchez, E., Gutierrez-Ayala, V.: EM-based Monte Carlo analysis and yield prediction of microwave circuits using linear-input neural-output space mapping. *IEEE Trans. Microw. Theory Tech.* **54**, 4528–4537 (2006)
11. Koziel, S., Bandler, J.W.: Recent advances in space-mapping-based modeling of microwave devices. *Int. J. Numer. Model.* **23**, 425–446 (2010)
12. Zhang, L., Zhang, Q.J., Wood, J.: Statistical neuro-space mapping technique for large-signal modeling of nonlinear devices. *IEEE Trans. Microw. Theory Tech.* **56**, 2453–2467 (2011)
13. Bandler, J.W., Cheng, Q.S., Koziel, S.: Simplified space mapping approach to enhancement of microwave device models. *Int. J. RF Microw. Comput.-Aided Eng.* **16**, 518–535 (2006)

14. Koziel, S., Bandler, S.W., Madsen, K.: A space mapping framework for engineering optimization: theory and implementation. *IEEE Trans. Microw. Theory Tech.* **54**, 3721–3730 (2006)
15. CST Microwave Studio, ver. 2011. CST AG, Bad Nauheimer Str. 19, D-64289 Darmstadt, Germany (2012)
16. Jiao, J.-J., Zhao, G., Zhang, F.-S., Yuan, H.-W., Jiao, Y.-C.: A broadband CPW-fed T-shape slot antenna. *Prog. Electromagn. Res.* **76**, 237–242 (2007)
17. Wi, S.-H., Lee, Y.-S., Yook, J.-G.: Wideband microstrip patch antenna with U-shaped parasitic elements. *IEEE Trans. Antennas Propag.* **55**, 1196–1199 (2007)
18. Alexandrov, N.M., Lewis, R.M.: An overview of first-order model management for engineering optimization. *Optim. Eng.* **2**, 413–430 (2001)
19. Kolda, T.G., Lewis, R.M., Torczon, V.: Optimization by direct search: new perspectives on some classical and modern methods. *SIAM Rev.* **45**, 385–482 (2003)
20. Koziel, S.: Multi-fidelity multi-grid design optimization of planar microwave structures with Sonnet. In: *International Review of Progress in Applied Computational Electromagnetics*, April 26–29, Tampere, Finland, pp. 719–724 (2010)
21. Manchec, A., Quendo, C., Favennec, J.-F., Rius, E., Person, C.: Synthesis of capacitive-coupled dual-behavior resonator (CCDBR) filters. *IEEE Trans. Microw. Theory Tech.* **54**, 2346–2355 (2006)
22. FEKO[®] User's Manual, Suite 6.0. EM Software & Systems-S.A. (Pty) Ltd, 32 Techno Lane, Technopark, Stellenbosch, 7600, South Africa (2010)
23. Beachkofski, B., Grandhi, R.: Improved distributed hypercube sampling. American Institute of Aeronautics and Astronautics. Paper AIAA 2002-1274 (2002)
24. Chen, C.Y., Hsu, C.Y.: A simple and effective method for microstrip dual-band filters design. *IEEE Microw. Wirel. Compon. Lett.* **16**, 246–248 (2006)
25. Koziel, S., Echeverría-Ciaurri, D., Leifsson, L.: Surrogate-based methods. In: Koziel, S., Yang, X.S. (eds.) *Computational Optimization, Methods and Algorithms*. Studies in Computational Intelligence, pp. 33–60. Springer, Berlin (2011)
26. Koziel, S., Ogurtsov, S., Jacobs, J.P.: Low-cost design optimization of slot antennas using Bayesian support vector regression and space mapping. In: *Loughborough Antennas and Propagation Conf.* (2012). doi:[10.1109/LAPC.2012.6402988](https://doi.org/10.1109/LAPC.2012.6402988)
27. Koziel, S., Cheng, Q.S., Bandler, J.W.: Space mapping. *IEEE Microw. Mag.* **9**, 105–122 (2008)
28. Bandler, J.W., Cheng, Q.S., Dakrouy, S.A., Mohamed, A.S., Bakr, M.H., Madsen, K., Sondergaard, J.: Space mapping: the state of the art. *IEEE Trans. Microw. Theory Tech.* **52**, 337–361 (2004)
29. Conn, A.R., Gould, N.I.M., Toint, P.L.: *Trust Region Methods*. MPS-SIAM Series on Optimization. Springer, Berlin (2000)