

Computationally Efficient Multi-Fidelity Bayesian Support Vector Regression Modeling of Planar Antenna Input Characteristics

J. P. Jacobs, S. Koziel, and S. Ogurtsov

Abstract—Bayesian support vector regression (BSVR) modeling of planar antennas with reduced training sets for computational efficiency is presented. Coarse-discretization electromagnetic (EM) simulations are exploited in order to find a reduced number of fine-discretization training points for establishing a high-fidelity BSVR model of the antenna. As demonstrated using three planar antennas with different response types, the proposed technique allows substantial reduction (up to 48%) of the computational effort necessary to set up the fine-discretization training data sets for the high-fidelity models with negligible loss in predictive power. The accuracy of the reduced-data BSVR models is confirmed by their successful use within a space mapping optimization/design algorithm.

Index Terms—Gaussian processes, microwave antennas, optimization, predictive models, support vector machines.

I. Introduction

Full-wave electromagnetic simulations based on, e.g., the method of moments or finite elements play a pervasive part in microwave engineering, as they permit highly accurate evaluation of microwave structures such as planar antennas. However, such simulations are costly in computational terms and their use for tasks requiring numerous analyses (e.g., statistical analysis, parametric design optimization) might become infeasible under certain conditions (for instance, a genetic algorithm optimization might require 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 (e.g., adjustable antenna geometry parameters and frequency as input, and a performance characteristic such as 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, e.g., [1]. It has recently been demonstrated [2] that Bayesian support vector regression (BSVR) [3] using a Gaussian kernel with automatic relevance determination (ARD) significantly outperforms the above standard SVR with isotropic kernel in modeling $|S_{11}|$ against frequency of CPW-fed slot antennas with multiple tunable geometry variables. BSVR is based on Gaussian process regression (GPR) [4]; the Bayesian framework enables efficient training of the multiple hyperparameters of the above ARD kernel by minimizing the negative log probability of

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: jjacobs@postino.up.ac.za).

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

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 task. In addition to its advantageous Bayesian-based features, BSVR also exhibits certain desirable properties of standard SVR, such as quadratic programming and (important for our purposes) sparseness of solutions, i.e., solutions that are fully characterized by the set of support vectors (SVs), which is a subset of the training set.

As for many other approximation-based methods, the drawback of BSVR modeling is the high initial cost of gathering the fine-discretization full-wave simulation data necessary to train the model so that it has high predictive accuracy. In the present study, we address this problem by exploiting the sparseness property of BSVR to reduce—in an optimal manner—the amount of expensive high-fidelity data required for training. Previous 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 by iterative identification of the model and adding 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] are local/trust region models; in contrast, our focus is on global, or “library” type, surrogate models that give accurate predictions over the entire input space, and that can be used for a variety of applications (e.g., statistical analysis, optimization). 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—see [2] for a discussion of the problems of standard SVR with respect to modeling highly non-linear antenna $|S_{11}|$ responses due to multivariate inputs). 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 three examples of antennas with highly non-linear $|S_{11}|$ 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 [8]–[12].

The article is organized as follows. Section II provides a brief theoretical overview of BSVR along the lines of [3], and describes the multi-fidelity modeling approach. Section III describes how the BSVR models were set up for each of the above antennas, and compares predictive accuracies of models trained on the reduced fine-discretization data to those of models trained on the original full data sets (i.e., selected by means of standard experimental design procedures). In Section IV, the reduced-data models are further evaluated by using them as basis for antenna optimization using a space mapping algorithm [13]. Conclusions are presented in Section V.

II. Bayesian Support Vector Regression Modeling

A. Bayesian Support Vector Regression Overview

Consider a training data set of n observations, $\mathcal{D} = \{(\mathbf{u}_i, y_i) | i = 1, \dots, n\}$. The BSVR formulation, which

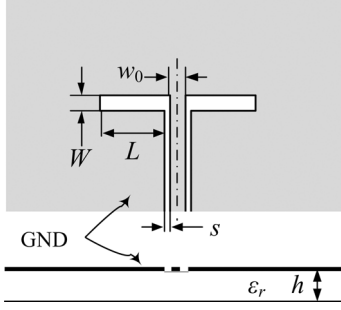


Fig. 1. Geometry of a CPW-fed slot dipole antenna (Antenna 1). The ground plane (GND) is of infinite lateral extent.

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 the 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 [14], 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

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

subject to $0 \leq \alpha_i, \alpha_i^* \leq C, i = 1, \dots, n$. In the above, Σ is a $n \times n$ matrix with $\Sigma_{ij} = k(\mathbf{u}_i, \mathbf{u}_j)$ and $k(\cdot)$ the kernel function. In particular, the Gaussian kernel with ARD (used throughout) 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 θ , which includes $\sigma_f^2, \tau_k, \kappa, C$, and

ε can be determined by minimizing the negative log probability of the data given the hyperparameters [3],

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

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, $\mathbf{f}_{MP} = \Sigma(\alpha - \alpha^*)$, 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.

B. BSVR Modeling With Reduced Data Sets

Suppose that a BSVR surrogate \mathbf{R}_s of the high-fidelity antenna model \mathbf{R}_f has to be constructed; \mathbf{R}_f is the response from CPU-intensive fine-discretization (EM) simulations. As noted earlier, the cost of gathering sufficient data to train \mathbf{R}_s typically is high. This problem is addressed as follows. First, an auxiliary BSVR model $\mathbf{R}_{s,aux}$ of the antenna is set up with training data obtained from coarse-discretization full-wave simulations (these simulations are referred to as the low-fidelity EM model \mathbf{R}_c). Specifically, the training set consists of n input vectors \mathbf{u}_i and associated targets $y_i = \mathbf{R}_c(\mathbf{u}_i)$, where \mathbf{u}_i contains geometry parameters and a frequency value within the range of interest, and y_i is the corresponding simulated $|S_{11}|$ value. The SVs obtained from $\mathbf{R}_{s,aux}$ (a subset of the training set) is then simulated at the \mathbf{R}_f (high) mesh density/fine discretization, providing the reduced training set for the high-fidelity BSVR surrogate model \mathbf{R}_s .

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 variables were $\mathbf{x} = [W \ L]^T$ mm, and the input space was defined by $5 \leq W \leq 10$ 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 $\varepsilon_r = 4.4$. We were interested in $|S_{11}|$ over the frequency band 2.0-to-2.7 GHz (visual inspection revealed that $|S_{11}|$ -against-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 (~ 130 000 mesh cells, simulation time 12 min), and a low-fidelity model \mathbf{R}_c (~ 5000 mesh cells, simulation time 30 s).

For training input data, 99 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, 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 $\{\mathbf{u}_i = [\mathbf{x}_i^T f_i]^T =$

Table 1: Predictive Errors of surrogate Antenna Models

Antenna No.	β	RMSE (%)			n_{SV}	n_{SV}/n (%)	Model Cost ¹	
		$\mathbf{R}_{s,aux}$	\mathbf{R}	$\mathbf{R}_{s,full}$			\mathbf{R}_s^2	$\mathbf{R}_{s,full}$
1	0.05	1.14	1.77	1.68	168	56	168	
	0.15	1.13	1.71	1.72	176	59	176	297
	0.25	1.15	1.68	1.68	206	69	206	
2	0.05	3.37	4.14	3.83	1819	52	1819	
	0.15	3.14	4.11	3.52	2010	57	2010	3528
	0.25	3.13	3.89	3.50	2225	63	2225	
3	0.15	6.59	5.77	5.53	3658	57	3658	
	0.20	6.54	5.87	5.41	3937	62	3937	6400
	0.25	6.59	5.71	5.27	3916	61	3916	

¹In terms of number of \mathbf{R}_f evaluations, i.e., fine-discretization full-wave simulations, required to generate the training data.

²The actual cost of setting up \mathbf{R}_s is slightly higher (by 1 percent for Antenna 2, 2 percent for Antenna 3, and 4 percent for Antenna 1) due to the coarse-discretization EM simulations used in the model construction process (cf. Section II.B).

$[W_i L_i f_i]^T | i = 1, \dots, n$, 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 were comprised 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 $\mathbf{R}_{s,aux}$ for 3 different values of β at the low end of its possible range ($\beta \in \{0.05, 0.15, 0.25\}$; as noted earlier, it is usually the case that the smaller the value of β , the smaller the number of SVs). Each $\mathbf{R}_{s,aux}$ was used to make predictions on 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 $\mathbf{R}_{s,aux}$ 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 purposes of comparison, surrogate models $\mathbf{R}_{s,full}$ trained on the full fine-discretization training data set ($n = 297$) were also set up. Table I gives, for each of the β values, the %RMSE values obtained with $\mathbf{R}_{s,aux}$ on the coarse-discretization test data, and \mathbf{R}_s , and $\mathbf{R}_{s,full}$ on the fine-discretization test data. Also shown in the table is n_{SV} , the number of SVs associated with $\mathbf{R}_{s,aux}$ (and therefore the number of training points for \mathbf{R}_s); and n_{SV}/n , the proportion of the full training data that were support vectors of $\mathbf{R}_{s,aux}$. The %RMSE values obtained for \mathbf{R}_s and $\mathbf{R}_{s,full}$ were either the same, or only marginally higher in the case of \mathbf{R}_s , indicating that reducing the number of training points from n to n_{SV} by using the SVs of $\mathbf{R}_{s,aux}$ as training points for \mathbf{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 ease of comparison, Table I 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 approximately 12 h for \mathbf{R}_s ($\beta = 0.15$), and 20 h for $\mathbf{R}_{s,full}$.

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

Fig. 2 shows the antenna layout [16]. The design variables were $\mathbf{x} = [a_x \ a_y \ ab]^T$ 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

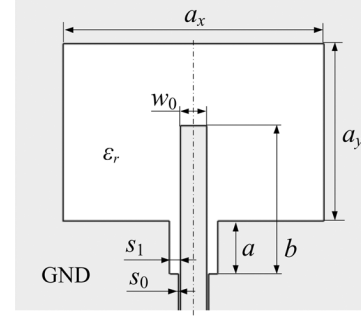


Fig. 2. Geometry of an UWB CPW-fed T-shaped slot antenna (Antenna 2; top view). The ground plane (GND) is of infinite lateral extent.

had height $h = 0.813$ mm and dielectric constant $\epsilon_r = 3.38$). The frequency band of interest was 2-to-8 GHz (as before, visual inspection confirmed that $|S_{11}|$ -against-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).

Training data were comprised of 294 geometries obtained by LHS, with 12 frequencies per geometry, randomly selected as before ($n = 3528$). Test data were comprised of 49 new LHS geometries, with 121 equally-spaced frequencies per geometry (as before, the value of n was determined by the performance of $\mathbf{R}_{s,aux}$ on the test data simulated at the coarse mesh density).

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

In general %RMSE values of \mathbf{R}_s were only somewhat higher than those of $\mathbf{R}_{s,full}$, suggesting as before that reducing the number of training points from n to n_{SV} by using the SVs of $\mathbf{R}_{s,aux}$ 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 $\mathbf{R}_{s,full}$ was 103 h.

C. Microstrip Antenna With Parasitic Elements (Antenna 3)

The antenna geometry is shown in Fig. 3 [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 dimensions were $a_0 = 5.8$ 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. The pin was 0.5 mm in diameter. The frequency band of interest was 4-to-7 GHz.

Training data were 400 geometries obtained by LHS, with 16 randomly selected frequencies per geometry ($n = 6400$). Test data were comprised of 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 fairly 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

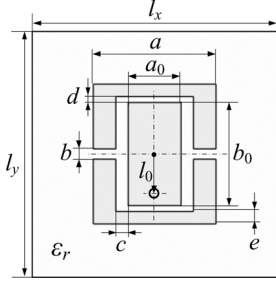


Fig. 3. Geometry of a broadband probe-fed microstrip patch antenna with two U-shaped parasitic elements (Antenna 3; top view). The dielectric substrate and ground plane share the lateral dimensions l_x and l_y . The empty circle below the centre of the patch indicates the position of the feed pin.

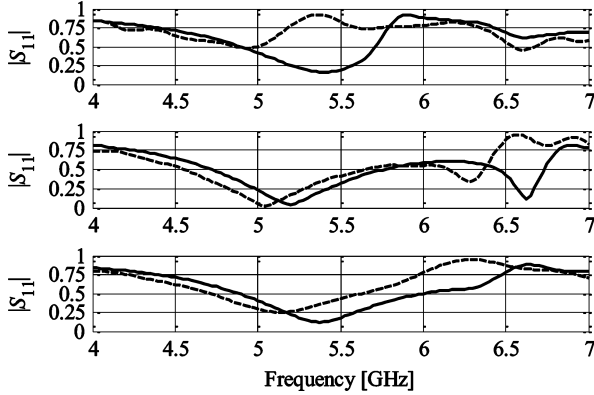


Fig. 4. Full-wave-simulated responses \mathbf{R}_c (---) and \mathbf{R}_f (—) for the Antenna 3 training geometries $\mathbf{x} = [16.2267, 0.7904, 1.8943, 0.4476, 1.1148]^T$ mm (top), $[14.8396, 0.6347, 0.9242, 0.9358, 0.8196]^T$ mm (center), and $[15.2672, 1.3194, 1.2827, 0.5165, 1.0004]^T$ mm (bottom).

coarse and fine responses agreed to some extent for some regions of the frequency band, there were considerable differences for others.

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

In order to explore further the influence of mesh density on our method, a second coarse model \mathbf{R}_{cc} (~ 8000 mesh cells, simulation time 8 s), i.e., coarser than \mathbf{R}_c , was generated, and corresponding surrogate models constructed. 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 3992 (cf. Table I). In order to assess 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 that there is some robustness to our procedure.

IV. Application Examples: Antenna Optimization

The BSVR models (both full and reduced ones) were used to perform design optimization of the antenna structures considered in Section III. As noted earlier, our models are intended to be multi-purpose global models that give accurate predictions for the whole of the input space; 4

multiple optimization runs corresponding to any number of sets of design specifications constitute one kind of repeated-use application. In each case, the initial design is the center of the region of interest $\mathbf{x}^{(0)}$. The design process starts from 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 dB) are optimized), the design is further refined using the space mapping iterative process [13]

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

where $\mathbf{R}_{\text{su}}^{(i)}$ is a surrogate model, enhanced by frequency and output space mapping [13]. The surrogate model setup is performed using an evaluation of \mathbf{R}_f at $\mathbf{x}^{(i)}$. U implements design specifications. For the sake of simplicity, we simply 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 space mapping context. Let $\mathbf{R}_{co}(\mathbf{x}, F)$ denotes 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

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

with

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

and

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

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

$$\begin{aligned} [\alpha_0^{(i)} \alpha_1^{(i)}] = \arg \min_{\mathbf{x}} & \left\| \mathbf{R}_f(\mathbf{x}^{(i)}) \right. \\ & \left. - \mathbf{R}_{co}(\mathbf{x}^{(i)}, \alpha_0^{(i)} + \alpha_1^{(i)} F) \right\| \quad (12) \end{aligned}$$

i.e., to minimize the misalignment between the high- and the scaled low-fidelity model response 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) that ensured 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 3 to 4 iterations to yield an optimized design, with the cost of each iteration effectively equal to a single evaluation of the high-fidelity model.

Fig. 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 corresponded to β values in Table I of 0.15 (Antenna 1), 0.05 (Antenna 2), and 0.15 (Antenna 3). Table II summarizes the results. It can be observed 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 3 \mathbf{R}_f evaluations (Antennas 1 and 2) and 4 \mathbf{R}_f evaluations (Antenna 3) were 36 min, 63 min, and 48 min, respectively).

For purposes of 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 maximum $|S_{11}|$ values at the final designs 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 space-mapping procedure, the computational expense

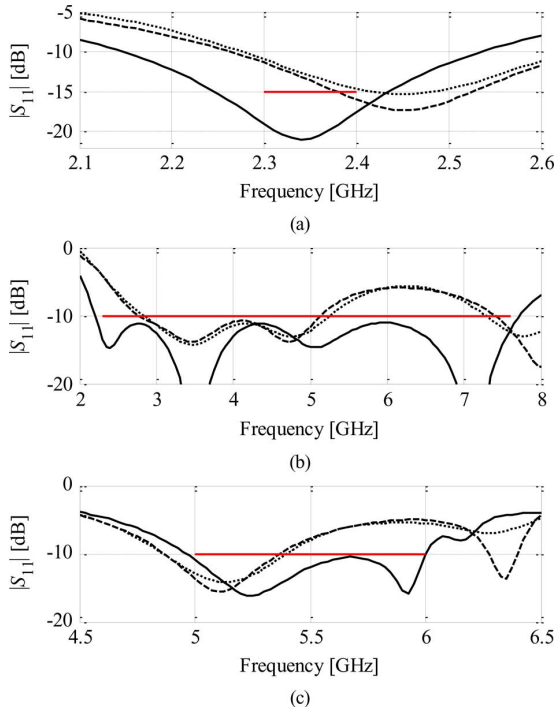


Fig. 5. Optimization results: responses of the BSVR model with reduced data set (\cdots) and the fine model at the initial design ($---$), and the fine model at the optimized design ($—$) for (a) Antenna 1, (b) Antenna 2, and (c) Antenna 3. Design specifications marked with horizontal solid line.

TABLE II
ANTENNAS OPTIMIZATION RESULTS

Antenna	Model	max $ S_{11} $ at Final Design ¹	Optimization Cost ²
1	$R_{s,full}$	-18.3 dB	3
	R_s	-17.6 dB	3
2	$R_{s,full}$	-11.0 dB	3
	R_s	-11.0 dB	3
3	$R_{s,full}$	-10.2 dB	4
	R_s	-10.2 dB	4

¹max $|S_{11}|$ 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). ²Number of R_f evaluations including evaluation at the initial design.

for the conventional optimization was at least an order of magnitude larger (i.e., 40, 148, and 201 R_f evaluations for Antennas 1, 2, and 3 respectively, compared to the 3, 3, and 4 R_f evaluations reported in Table II). This emphasizes how much faster optimization can be realized when accurate models such as BSVR models are available. Our primary contribution is that we reduce by up to 48% the high initial cost of setting up these global multi-use models (by comparison with which the cost of the optimization using space mapping is insignificant).

V. CONCLUSION

Accurate Bayesian support vector regression modeling of the highly non-linear input characteristics of planar antennas using reduced high-fidelity training data sets is presented. The reduction of the number of fine-discretization training points is realized by performing BSVR modeling on the coarse-discretization EM simulation data (selected by means of standard experimental design) and then obtaining high-fidelity simulations only for the points that contribute to the initial BSVR model in a non-trivial way. Computational savings thus ob-

tained had little effect on modeling accuracy. We also demonstrate 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 has to be set by the user, namely β (hyperparameters are initialized randomly during training). This stands in contrast to for instance neural network-based methodologies for regression that might require the tuning of a variety of architectural/performance parameters (e.g., number of hidden units, learning rate, momentum).

References

- [1] G. Angiulli, M. Cacciola, and M. Versaci, "Microwave devices and antennas modelling by support vector regression machines," *IEEE Trans. Magn.*, vol. 43, pp. 1589–1592, 2007.
- [2] 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, 2012.
- [3] W. Chu, S. S. Keerthi, and C. J. Ong, "Bayesian support vector regression using a unified loss function," *IEEE Trans. Neural Netw.*, vol. 15, no. 1, pp. 29–44, 2004.
- [4] C. E. Rasmussen and C. K. I. Williams, *Gaussian Processes for Machine Learning*. Cambridge, MA: MIT Press, 2006.
- [5] 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. Microwave Theory Tech.*, vol. 49, pp. 2282–2291, 2001.
- [6] 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.
- [7] N. T. Tokan and F. Gunes, "Knowledge-based support vector synthesis of the microstrip lines," *Prog. Electromagn. Res.*, vol. 92, pp. 65–77, 2009.
- [8] J. W. Bandler, N. Georgieva, M. A. Ismail, J. E. Rayas-Sánchez, and Q. J. Zhang, "A generalized space mapping tableau approach to device modeling," *IEEE Trans. Microwave Theory Tech.*, vol. 49, no. 1, pp. 67–79, 2001.
- [9] E. Rayas-Sánchez and V. Gutierrez-Ayala, "EM-based Monte Carlo analysis and yield prediction of microwave circuits using linear-input neural-output space mapping," *IEEE Trans. Microwave Theory Tech.*, vol. 54, no. 12, pp. 4528–4537, 2006.
- [10] S. Koziel and J. W. Bandler, "Recent advances in space-mapping-based modeling of microwave devices," *Int. J. Numerical Model.*, vol. 23, no. 6, pp. 425–446, 2010.
- [11] L. Zhang, Q. J. Zhang, and J. Wood, "Statistical neuro-space mapping technique for large-signal modeling of nonlinear devices," *IEEE Trans. Microwave Theory Tech.*, vol. 56, no. 1, pp. 2453–2467, 2011.
- [12] J. W. Bandler, Q. S. Cheng, and S. Koziel, "Simplified space mapping approach to enhancement of microwave device models," *Int. J. RF Microw. Comput.-Aided Eng.*, vol. 16, no. 5, pp. 518–535, 2006.
- [13] S. Koziel, J. W. Bandler, and K. Madsen, "A space mapping framework for engineering optimization: Theory and implementation," *IEEE Trans. Microwave Theory Tech.*, vol. 54, no. 10, pp. 3721–3730, 2006.
- [14] B. Schölkopf and A. J. Smola, *Learning With Kernels: Support Vector Machines, Regularization, Optimization, and Beyond*. Cambridge, MA: MIT Press, 2002.
- [15] CST Microwave Studio. Darmstadt, Germany, 2011, ver. 2011, CST AG, Bad Nauheimer Str. 19, D-64289.
- [16] J.-J. Jiao, G. Zhao, F.-S. Zhang, H.-W. Yuan, and Y.-C. Jiao, "A broadband CPW-fed T-shape slot antenna," *Prog. Electromagn. Res.*, vol. 76, pp. 237–242, 2007.
- [17] S.-H. Wi, Y.-S. Lee, and J.-G. Yook, "Wideband microstrip patch antenna with U-shaped parasitic elements," *IEEE Trans. Antennas Propag.*, vol. 55, no. 4, pp. 1196–1199, 2007.
- [18] N. M. Alexandrov and R. M. Lewis, "An overview of first-order model management for engineering optimization," *Opt. Eng.*, vol. 2, no. 4, pp. 413–430, 2001.
- [19] 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.
- [20] S. Koziel, "Multi-fidelity multi-grid design optimization of planar microwave structures with sonnet," in *Proc. 26th Review of Progress in Applied Computational Electromagnetics*, Tampere, Finland, Apr. 26–29, 2010, pp. 719–724.