Cost-Effective Global Surrogate Modeling of Planar Microwave Filters Using Multi-Fidelity Bayesian Support Vector Regression

J. Pieter Jacobs,[1] Slawomir Koziel[2]

[1] Department of Electrical, Electronic and Computer Engineering, University of Pretoria, Pretoria, South Africa
[2] Engineering Optimization and Modeling Center, School of Science and Engineering, Reykjavik University, Menntavegur 1, 101 Reykjavik, Iceland

Abstract: A computationally efficient method is presented for setting up accurate Bayesian support vector regression (BSVR) models of the highly nonlinear $|S_{21}|$ responses of planar microstrip filters using substantially reduced finely discretized training data (compared to traditional design of experiments techniques). Inexpensive coarse-discretiza-tion full-wave simulations are exploited in conjunction with the sparseness property of BSVR to identify the regions of the input space requiring denser sampling. The proposed technique allows for substantial reduction (by up to 51%) of the computational expense necessary to collect the finely discretized training data, with negligible loss in predictive accuracy. The accuracy of the reduced-data BSVR models is confirmed by their use within a space mapping optimization algorithm.

Introduction

High-fidelity full-wave electromagnetic simulations based on, e.g., the method of moments play a ubiquitous part in contemporary microwave engineering, as they permit stable and highly accurate evaluation of microwave structures such as planar filters. However, such simulations are computationally expensive, and their use for design tasks requiring numerous analyses (e.g., statistical analysis, 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 fast and accurate surrogate models of the structures under consideration are indispensable.

Unfortunately, the setting up of accurate models is often severely impeded by the high initial computational cost of gathering the training data from fine-discretization full-wave simulations. The standard approach is to use

Correspondence to: J. Pieter Jacobs; e-mail: jpjacobs@postino.up.ac.za

trial and error to determine the number of randomly distributed fine-discretization training data points required to train a model to exhibit a certain predictive accuracy. Generally, uniformly randomly distributed data is used because the variation of the response surface throughout the multi-dimensional input space typically is not well understood (it would be preferable to populate regions of rapid variation more densely with training points). The trial-and-error approach involves guessing a starting number of training data points, training the model, testing it, and then simulating more data for further training if the predictive accuracy needs improving. Numerous iterations might be necessary until the desired accuracy is met. If the data takes very long to simulate (days or even weeks are not unheard of in some cases) and/or many data points are required (*e.g.*, thousands as in the case of our filters), the trial and error approach might become extremely expensive. Furthermore, it is likely that some of the final training data will be redundant given their uniform distribution.

In this article, we introduce a method for efficient selection of training data using Bayesian support vector regression (BSVR) [1] as a modeling tool. Specifically,

we exploit the sparseness property of BSVR within a multi-fidelity modeling approach to significantly reduce the amount of expensive high-fidelity data required for training accurate surrogate models (compared with the amount of training data that would be required using the above experimental design approach). Our approach entails 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 reduced training data set for the actual (high-fidelity) BSVR model. The role of the auxiliary model can be seen as identifying regions of the input space where more samples are needed compared with other regions, for example because the response is more variable there. It should be noted that we focus on global, or "library" type, surrogate models that give accurate predictions over the entire input space—as opposed to local/ trust region models used within iterative optimization frameworks [2, 3].

The reason for our preference of Bayesian support vector regression (BSVR) over standard support vector regression (SVR) [4] —which so far has been the most widely used kernel-based learning method for microwave modeling [5–7]—is that BSVR provides a mechanism for optimizing the multiple hyperparameters of the powerful Gaussian kernel with automatic relevance determination (ARD), while standard SVR does not. Standard SVR can only handle the less expressive isotropic Gaussian kernel [5–7], as it has no mechanism by which more than two hyperparameters can feasibly be optimized (for a discussion see Ref. 8). In Ref. 8, BSVR employing a Gaussian kernel with ARD was shown to significantly outperform standard SVR with an isotropic Gaussian kernel on a highly nonlinear microwave modeling task, namely $|S_{11}|$-against-frequency responses of CPW-fed slot antennas with multiple adjustable geometry variables. (The form of the Gaussian kernel with ARD, its feature selection capabilities, and the optimization of its hyperparameters are discussed in Section II.A.)

The article is organized as follows. Section II briefly considers theoretical aspects of BSVR, and presents details of our multi-fidelity approach for training accurate BSVR models using greatly reduced fine-discretization training sets. Section III describes how we set up reduced-data BSVR models for two challenging problems, namely modeling of the highly nonlinear $|S_{21}|$ responses (as a function of their design parameters and frequency) of a capacitively coupled dual-behavior resonator (CCDBR) microstrip bandpass filter with three adjustable geometry variables, and an open-loop ring-resonator (OLRR) microstrip bandpass filter with seven adjustable geometry variables. It is shown that the accuracy of the reduced-data models is on par with that of BSVR surrogate models trained on the full fine-discretization data sets. In Section IV, the robustness in terms of predictive accuracy of the reduced-data models is further evaluated by using them in filter optimization by means of space mapping [9]. Conclusions are presented in Section V.

II. BAYESIAN SUPPORT VECTOR REGRESSION (BSVR) MODELING

A. Theoretical Aspects of BSVR

BSVR [1] is fundamentally a version of Gaussian process regression (GPR) [10], and is based on Bayesian principles. The crucial innovation in BSVR is the use of the so-called soft-insensitive loss function $L_{\varepsilon,\beta}$ [1, Eq. (11)], which combines advantageous properties of both the ε-insensitive loss function of standard SVR (sparseness of solutions) [4], and Huber's loss function (differentiability):

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

In the above, $0 < \beta \le 1$, and $\varepsilon > 0$; in the standard Bayesian regression framework for GPR, training targets y_i corresponding to input vectors x_i are expressed as $y_i = f(\mathrm{x}_i) + \delta_i$, where the δ_i are independent, identically distributed noise variables, and the underlying function f is a random field – δ in Eq. (1) is equivalent to δ_i.
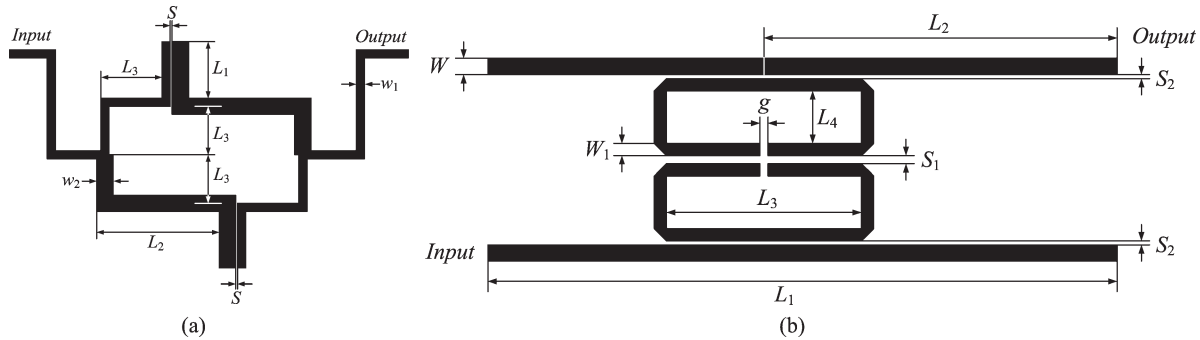
The result is that BSVR exhibits desirable SVR properties such as quadratic programming and sparseness of solutions, i.e., solutions that are fully characterized by the set of support vectors (a subset of the training set input vectors); while also exhibiting attractive features of GPR, such as likelihood-based optimization of hyperparameters. The crucial equations of the BSVR algorithm have already been summarized in a microwave modeling context in Ref. 8; a detailed treatment can be found in Ref. 1.

Training data sets in BSVR assume the form of n pairs of input vectors x_i and target (output) scalars y_i, $= \{(\mathrm{x}_i, y_i) \mid i = 1, ..., n\}$. As noted earlier, we use the Gaussian kernel with ARD, which takes the form

$$
k(\mathrm{x}_i, \mathrm{x}_j) = \sigma_f^2 \exp\left(-\frac{1}{2}\sum_{k=1}^{D} \frac{(x_{i_k} - x_{jk})^2}{\tau_k^2}\right) + \kappa
\tag{2}
$$

where x_{ik} and x_{jk} are the kth components of the D-dimensional input vectors x_i and x_j; $\tau_k > 0$ ($k = 1,..., D$) are tunable length-scale hyperparameters that determine the relevance for the regression of x_{ik} and x_{jk}, in effect presenting a built-in mechanism for feature selection; and σ_f and κ are additional tunable hyperparameters. The optimal hyperparameters (which include σ_f, τ_k, and κ in Eq. (2) as well as C and ε of the dual problem [1, Eq. (29)]) can be found as the set of hyperparameters that minimize the negative log probability of the data given the hyperparameters [1, Eq. (38)]. The dual problem is

$$
\min_{\alpha,\alpha^*} \frac{1}{2}(\alpha - \alpha^*)^\top \Sigma (\alpha - \alpha^*) - \sum_{i=1}^{n} y_i(\alpha_i - \alpha_i^*) \\
+ \sum_{i=1}^{n} (\alpha_i + \alpha_i^*)(1 - \beta)\varepsilon + \frac{\beta\varepsilon}{C} \sum_{i=1}^{n} y_i(\alpha_i^2 + \alpha_i^{*2})
\tag{3}
$$

Figure 1 Geometry of the (a) CCDBR bandpass filter [11], and (b) OLRR bandpass filter [13].

subject to $0 \leq \alpha_i, \alpha_i^* \leq C, \quad i = 1, ..., n$. In the above, Σ is a $n \times n$ matrix with $\Sigma_{ij} = k(\boldsymbol{x}_i, \boldsymbol{x}_j)$ and $k(\cdot)$ the kernel function.

The regression estimate at a test input $\boldsymbol{x}^*$ can be expressed as $f(\boldsymbol{x}^*) = \sum_{i=1}^{n} k(\boldsymbol{x}_i, \boldsymbol{x}^*)(\alpha_i - \alpha_i^*)$, with $0 \leq \alpha_i, \alpha_i^* \leq C$. The subset of training input vectors $\boldsymbol{x}_i$ corresponding to $|\alpha_i - \alpha_i^*| > 0$ are the support vectors. Training input vectors corresponding to $(\alpha_i - \alpha_i^*) = 0$ do not play a role in the regression estimate. In general, the lower the value of the parameter β in the loss function Eq. (1), the smaller the number of support vectors.

B. Multi-Fidelity Method for Training Accurate Models from Reduced High-Fidelity Data Sets

The model response of interest is $|S_{21}|$ over a specified frequency range for a particular filter geometry. A model (column) input vector $\boldsymbol{x}$ consists of the set of adjustable geometry parameters $\boldsymbol{x}_g$ and a frequency value f; thus we have $\boldsymbol{x} = [\boldsymbol{x}_g^T \ f]^T$. The (scalar) response of a model at a specific frequency, for example the model response $\boldsymbol{R}_f$ which is the fine-discretization full-wave simulated $|S_{21}|$ value, is denoted as $\boldsymbol{R}_f(\boldsymbol{x})$, or $\boldsymbol{R}_f(\boldsymbol{x}_g, f)$ in the space-mapping context of Section IV (other models are denoted similarly). Suppose now that a BSVR surrogate $\boldsymbol{R}_s$ of the CPU-intensive high-fidelity model $\boldsymbol{R}_f$ has to be constructed. As noted earlier, the computational cost of gathering sufficient data to train $\boldsymbol{R}_s$ typically is high. To address this, first, an auxiliary BSVR model $\boldsymbol{R}_{s.aux}$ of the 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 $\boldsymbol{R}_c$). The training set for $\boldsymbol{R}_{s.aux}$ consists of n input vectors $\boldsymbol{x}_i$, $i = 1,...,n$; and associated targets $y_i = \boldsymbol{R}_c(\boldsymbol{x}_i)$, where $\boldsymbol{x}_i$ contains geometry parameters and a frequency value as noted above, and y_i is the corresponding simulated $|S_{21}|$ value. The support vectors obtained from $\boldsymbol{R}_{s.aux}$ is then simulated at the (high) mesh density of $\boldsymbol{R}_f$, providing the reduced fine-discretization training set for $\boldsymbol{R}_s$.

As experience has shown that the coarsely simulated targets $\boldsymbol{R}_c$ and finely simulated targets $\boldsymbol{R}_f$ of microwave structures such as filters are usually reasonably well correlated (in Section III we give coarse and fine meshing densities for both filters as an indication of by how much these can differ), we assume in the above that the regions of the input space which support the crucial variations in the coarse response surface also will support the crucial variations in the fine response surface. Hence the support vectors 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., $\boldsymbol{R}_s$.

III. MODELING EXAMPLES

A. Capacitively Coupled Dual-Behavior Resonator (CCDBR) Microstrip Bandpass Filter

Consider the capacitively coupled dual-behavior resonator (CCDBR) bandpass filter [11] implemented in microstripline, shown in Figure 1a. The three design variables were $\boldsymbol{x}_g = [L_1 \ L_2 \ L_3]^T$. The design variable space for the BSVR models was defined by the center vector $\boldsymbol{x}_g^0 = [3 \ 5 \ 1.5]^T$ mm and size vector $\boldsymbol{\delta} = [1 \ 1 \ 0.5]^T$ mm such that the variable ranges were $\boldsymbol{x}_g^0 \pm \boldsymbol{\delta}$ mm ($\boldsymbol{x}_g^0$ 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 $\varepsilon_r = 9.9$; the value of S was 0.05 mm, while 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 $\boldsymbol{R}_f$ of the filter was simulated using FEKO [12] (total mesh number 715, simulation time about 15 seconds per frequency). The low-fidelity model $\boldsymbol{R}_c$ was also simulated in FEKO (total mesh number 136, simulation time 0.6 seconds 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). For each geometry, 12 simulation frequencies were selected by uniform random sampling from the above frequency range, yielding a total of $n = 400 \times 12 = 4800$ training input vectors of the form $\{\boldsymbol{x}_i = [L_{1i} \ L_{2i} \ L_{3i} \ f_i]^T | \ i = 1,..., n\}$, with L_{1i}, L_{2i}, and L_{3i} the design variables corresponding to the ith input vector, and f_i a frequency value within the range of interest. The corresponding output scalars, obtained from FEKO simulations, were $y_i = |S_{21i}|$. Test data were likewise obtained from 50 new geometries, also obtained

TABLE I Predictive Errors of Surrogate Filter Models

		RMSE (%)				
	β	$R_{s,aux}$	R_s	$R_{s,full}$	n_{SV}	n_{SV}/n (%)
CCDBR Filter	0.1	4.17	6.05	5.98	2488	52
($n = 4800$)	0.2	4.12	5.87	5.95	2563	53
	0.3	4.11	5.95	5.86	2886	60
OLRR Filter	0.1	4.12	4.26	4.21	2360	49
($n = 4800$)	0.2	3.77	4.47	4.32	2565	53
	0.3	3.46	3.90	3.64	2744	57

$R_{s,aux}$ 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 support vectors of $R_{s,aux}$.

via LHS, with 41 equally spaced frequencies per geometry. Training data were simulated at the R_c mesh density, and used to train the BSVR model $R_{s,aux}$ for $\beta = 0.1, 0.2$, and 0.3 (β is the loss function parameter in Eq. (1) discussed above). $R_{s,aux}$ was used to make predictions on the test data (also simulated at the 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} support vectors of $R_{s,aux}$ were simulated at the R_f mesh density. BSVR models fitted to these reduced training sets gave the desired surrogate models R_s. For comparison, surrogates $R_{s,full}$ trained on the full fine-discretization training data ($n = 4800$) were also set up. Table I gives, for the three β values, the %RMSEs obtained with $R_{s,aux}$ on the coarse test data, and with R_s and $R_{s,full}$ on the fine test data; as well as the number of support vectors obtained in each instance. The highly similar %RMSEs obtained with R_s and $R_{s,full}$ 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 2a shows typical predictive results for $|S_{21}|$ against frequency obtained for the test geometry $x_g = [2.794\ 4.407\ 1.491]^T$
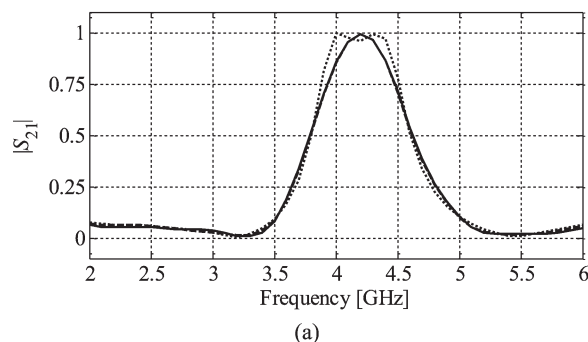
mm. Some discrepancy can be observed when comparing the RMSE values for $R_{s,aux}$ in Table I to those for R_s (and $R_{s,full}$). It follows from the fact that the coarse model responses are slightly smoother as functions of frequency (*i.e.*, they do not contain as much detail particularly in the 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 R_s and $R_{s,full}$ 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 Section IV. The total computational time necessary to gather the training data for setting up $R_{s,full}$ was 20 hours, whereas the corresponding time for setting R_s (including both low- and high-fidelity model evaluations) was 11.2, 11.4, and 12.8 hours (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 $\beta = 0.1$).

B. Open-Loop Ring-Resonator (OLRR) Bandpass Filter
The filter geometry [13] is shown in Figure 1b. The seven design variables were $x_g = [L_1\ L_2\ L_3\ L_4\ S_1\ S_2\ G]^T$. The design space was described by $x_g^0 = [40\ 8\ 6\ 4\ 0.2\ 0.1\ 1]^T$ mm and $\delta = [2\ 1\ 0.4\ 0.4\ 0.1\ 0.05\ 0.2]^T$ mm. Substrate parameters were $h = 0.635$ mm and $\varepsilon_r = 10.2$, while 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 [12] (total mesh number 1084 and simulation time 40 seconds per frequency for R_f; total mesh number 148 and simulation time 0.8 seconds per frequency for R_c). Training data were comprised of 400 geometries obtained by LHS, with 12 randomly selected frequencies per geometry ($n = 4800$), while test data were 50 new LHS geometries with 81 equally spaced frequencies per geometry. Setting up $R_{s,aux}$, R_s and $R_{s,full}$ proceeded in a manner similar to the earlier filter. Table I gives the %RMSE values obtained with $R_{s,aux}$ on the coarse test data and with R_s and $R_{s,full}$ on the fine test data; as well as support vector

(a)

(b)
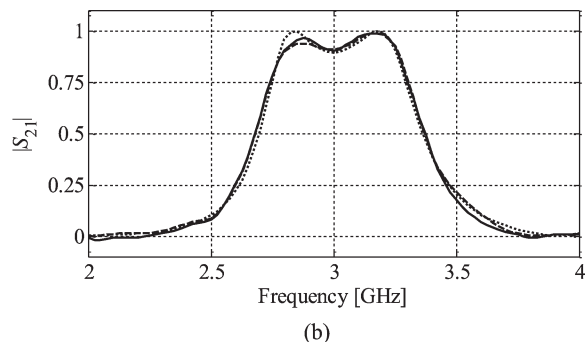
Figure 2 Predictive response of BSVR models R_s (—) and $R_{s,full}$ (- - -), and high-fidelity model R_f (⋯) for (a) the CCDBR bandpass filter test geometry $x_g = [2.794\ 4.407\ 1.491]^T$ mm, and (b) the OLRR bandpass filter test geometry $x_g = [38.088\ 8.306\ 5.882\ 4.029\ 0.193\ 0.061\ 0.985]^T$ mm.
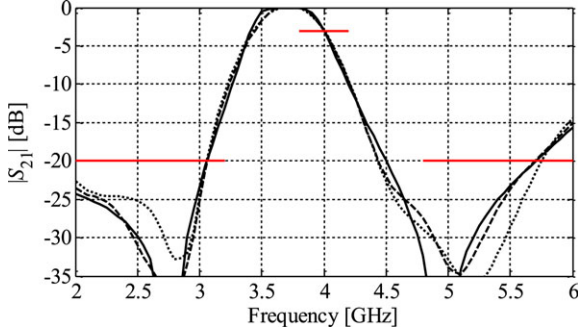
Figure 3 CCDBR filter: responses of the high-fidelity model (—), and BSVR models $R_{s,full}$ (- - -) and R_s (····) at the initial design $x_g^{(0)}$. Design specification as per horizontal solid (red) lines.

counts. In general %RMSE values of R_s were only marginally higher than those of $R_{s,full}$, suggesting as before that reducing the number of training points from n to n_{SV} by using the support vectors of $R_{s,aux}$ as fine-discretization training points for R_s has little effect on prediction accuracy. The greatest reduction in data (51%) was obtained for $\beta = 0.1$. Figure 2b shows representative predictive results for $|S_{21}|$ against frequency, in particular for the test geometry x_g = [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 $R_{s,full}$ was 53.3 hours. The corresponding time for setting up R_s (including both low- and high-fidelity model evaluations) was 27.2, 29.3, and 31.4 hours ($\beta = 0.1, 0.2,$ and 0.3, respectively). The computational savings due to the proposed technique range from 41% (for $\beta = 0.3$) to 49% (for $\beta = 0.1$).

IV. FILTER OPTIMIZATION USING BSVR SURROGATES

The BSVR models can be used to perform design optimization. Below, we denote the vector response of a model, defined as the set of scalar responses at selected frequency points $f = [f_1\ f_2\ ...\ f_m]$ over the range of interest, as $R_f(x_g,f)$ (models other than R_f are denoted similarly). Given that 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 standard space mapping surrogate-based optimization (SBO) [9] algorithm with frequency scaling [14]. In particular,

$$x_g^{(i+1)} = \arg\min_{x_g} U\left(R_{ss}^{(i)}(x_g,f)\right) \quad (4)$$

where U is the objective function implementing the design specifications, and $R_{ss}^{(i)}$ is a surrogate model, enhanced by output space mapping (OSM) of the form $R_{ss}^{(i)}(x_g,f) = R_{cc}(x_g,f) + [R_f(x_g^{(i)},f) - R_{cc}(x_g^{(i)},f)]$, which ensures that $R_{ss}^{(i)}(x_g^{(i)},f) = R_f(x_g^{(i)},f)$. In the above, $R_{cc} = R_s$ or $R_{s,full}$ (the BSVR models). The optimization problem (4) is considered in the regions of validity of the respective BSVR

models, which are defined at the beginning of Sections III.A and III.B for both filters, respectively.

The objective function U is a minimax objective function. More specifically, if $R_{ss}(x_g,f) = [R_{ss}(x_g,f_1)\ R_{ss}(x_g,f_2)...\ R_{ss}(x_g,f_k)...\ R_{ss}(x_g,f_m)]^T$, where $R_{ss}(x_g,f_k)$ is the scalar response of R_{ss} at frequency f_k, and the design specifications are $R_{ss}(x_g,f_j) \leq u_j, j = k_1,...,k_{Nu}$ (upper specifications), and $R_{ss}(x_g,f_j) \geq l_j, j = m_1,..., m_{Nl}$ (lower specifications), then the objective function is defined as

$$U(R_{ss}(x_g,f)) = \max(\max\{R_{ss}(x_g,f_j) - u_j; j = k_1, ..., k_{Nu}\}, \max\{l_j - R_{ss}(x_g,f_j); j = m_1, ..., m_{Nl}\}) \quad (5)$$

i.e., as the maximum violation of the design specifications for all frequencies where the specifications are set up. The design x_g satisfies the specifications if and only if $U(R_{ss}(x_g,f))$ is nonpositive, that is, if the upper/lower specifications are not violated for any frequency of interest.

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 [cf. Eq. (4)] 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 3 shows the responses of the high-fidelity model R_f as well as the responses of R_s and $R_{s,full}$ at the initial design $x_g^{(0)} = [3\ 6\ 1.5]^T$ mm. The high-fidelity model responses at the optimized designs found using both R_s and $R_{s,full}$ are shown in Figure 4 (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 OLRR 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 $3.5 \leq f \leq 4.0$ GHz. Figure 5 shows the responses of R_f, R_s, and $R_{s,full}$ at the initial design $x_g^{(0)} = [40.0\ 8.0\ 6.0\ 4.0\ 0.1\ 0.1\ 1.0]^T$ mm. The responses at the optimized designs obtained using R_s and $R_{s,full}$ are shown in Figure 6 (these designs were
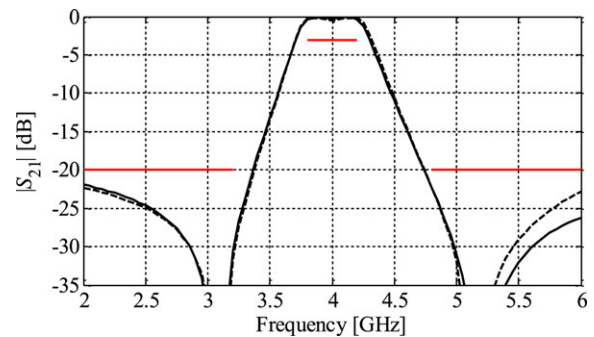
Figure 4 CCDBR filter: responses of the high-fidelity model at the optimized designs found using $R_{s,full}$ (—) and R_s (- - -). Design specification as per horizontal solid (red) lines.
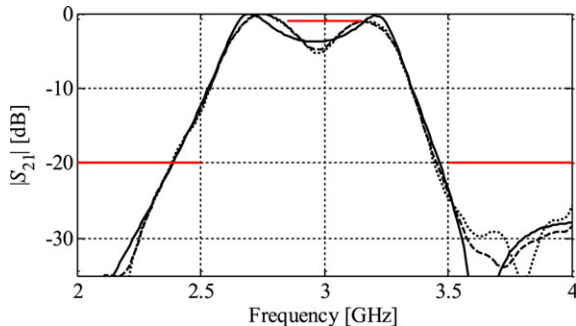
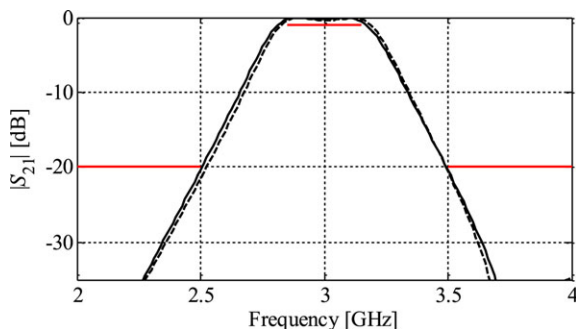**Figure 5** OLRR bandpass filter: responses of the high-fidelity model (—), and BSVR models $\boldsymbol{R}_{s.full}$ (- - -) and $\boldsymbol{R}_s$ (····) at the initial design $\boldsymbol{x}_g^{(0)}$. Design specification as per horizontal solid (red) lines.

$[39.605\ 8.619\ 6.369\ 3.718\ 0.300\ 0.069\ 0.986]^T$ 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.

V. CONCLUSIONS

A major impediment to the setting up of accurate surrogate models is the costly and lengthy process of collecting training data from fine-discretization full-wave simulations. We introduce accurate Bayesian support vector regression modeling of highly nonlinear $|S_{21}|$ responses of microstrip filters using substantially reduced fine-discretization training data sets (compared with data sets generated using standard experimental design methods). The reduction in the number of fine-discretization training points (by as much as 51% in the case of the open-loop ring resonator bandpass filter) is achieved by performing BSVR modeling on the coarse-discretization EM simulation data, and then obtaining high-fidelity simulations only for the points that contribute to the initial BSVR model in a non-trivial way (i.e., the support vectors). Computational savings achieved in this manner had negligible effect on predictive accuracy. Within a space mapping optimization context, the reduced-data BSVR

Figure 6 OLRR bandpass filter: responses of the high-fidelity model at the optimized designs found using $\boldsymbol{R}_{s.full}$ (—) and $\boldsymbol{R}_s$ (- - -). Design specification as per horizontal solid (red) lines.

models exhibited performance on a par with BSVR surrogates trained on the full fine-discretization data sets, validating our approach. A major advantage of BSVR is that only a single parameter has to be set by the user, namely β (this stands in contrast to for example neural network-based methodologies that might require the trial-and-error-based tuning of a variety of architectural/performance parameters). We note that an approach similar to ours was adopted for the modeling of microstrip transmission lines in [6]; however, these authors only used standard SVR with an isotropic Gaussian kernel to model functions that were relatively simple compared to our filter responses. Furthermore, the models in [6] had input vectors that were limited to three components/features; our BSVR filter models could successfully handle input vectors with up to eight features, confirming the efficacy of the Gaussian kernel with ARD (which cannot be used optimally within standard SVR).

REFERENCES

1. W. Chu, S.S. Keerthi, and C.J. Ong, Bayesian support vector regression using a unified loss function, IEEE Trans Neural Networks 15 (2004), 29–44.
2. 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 49 (2001), 2282–2291.
3. I. Couckuyt, F. Declercq, T. Dhaene, H. Rogier, and L. Knockaert, Surrogate-based infill optimization applied to electromagnetic problems, Int J RF Microwave CAE 20 (2010), 492–501.
4. B. Schölkopf and A.J. Smola, Learning with Kernels: Support Vector Machines, Regularization, Optimization, and Beyond, MIT Press, Cambridge, MA, 2002.
5. G. Angiulli, M. Cacciola, and M. Versaci, Microwave devices and antennas modelling by support vector regression machines, IEEE Trans Magnetics 43 (2007), 1589–1592.
6. N.T. Tokan, F. Gunes, Knowledge-based support vector synthesis of the microstrip lines, Prog Electromag Res 92 (2009), 65–77.
7. Y.Q. Wu, Z.X. Tang, Y.H. Xu, X. He, and B. Zhang, Permittivity measurement of ferroelectric thin film based on CPW transmission line, J Electromagn Waves Appl 22 (2008), 555–562.
8. J.P. Jacobs, Bayesian support vector regression with automatic relevance determination kernel for modeling of antenna input characteristics, IEEE Trans Antennas Propagat 60 (2012), 2114–2118.
9. S. Koziel, J.W. Bandler, and K. Madsen, A space mapping framework for engineering optimization: Theory and implementation, IEEE Trans Microwave Theory Tech 54 (2006), 3721–3730.
10. C.E. Rasmussen and C.K.I. Williams, Gaussian Processes for Machine Learning, MIT Press, Cambridge, MA, 2006.

11. A. Manchec, C. Quendo, J.-F. Favennec, E. Rius, and C. Person, Synthesis of capacitive-coupled dual-behavior resonator (CCDBR) filters, IEEE Trans Microwave Theory Tech 54 (2006), 2346–2355.
12. FEKO® User's Manual,Suite 6.0, EM Software & Systems-S.A. (Pty) Ltd, 32 Techno Lane, Technopark, Stellenbosch, 7600, South Africa, 2010.
13. C.Y. Chen, C.Y. Hsu, A simple and effective method for microstrip dual-band filters design, IEEE Microwave Wireless Comp Lett 16 (2006), 246–248.
14. S. Koziel, D. Echeverría-Ciaurri, and L. Leifsson, Surrogate-based methods, in S. Koziel and X.S. Yang, Eds., Computational Optimization, Methods and Algorithms, Series: Studies in Computational Intelligence, Springer-Verlag, pp.33—60, 2011.

BIOGRAPHIES

Pieter Jacobs holds B.Eng., M.Eng., and PhD degrees in electronic engineering from the University of Pretoria, South Africa, and a doctorate in music from Yale University. He is at present a Senior Lecturer in the Department of Electrical, Electronic and Computer Engineering at the University of Pretoria. His research interests include slot antennas and arrays, computational electromagnetics, and modeling and optimization of microwave antennas and devices. He is a registered Professional Engineer in South Africa.

Slawomir Koziel received the M.Sc. and Ph.D. degrees in electronic engineering from Gdansk University of Technology, Poland, in 1995 and 2000, respectively. He also received the M.Sc. degrees in theoretical physics and in mathematics, in 2000 and 2002, respectively, as well as the PhD in mathematics in 2003, from the University of Gdansk, Poland. He is currently a Professor with the School of Science and Engineering, Reykjavik University, Iceland. His research interests include CAD and modeling of microwave circuits, simulation-driven design, surrogate-based optimization, space mapping, circuit theory, analog signal processing, evolutionary computation and numerical analysis.