Bayesian Support Vector Regression With Automatic Relevance Determination Kernel for Modeling of Antenna Input Characteristics

J. P. Jacobs

***Abstract*—The modeling of microwave antennas and devices typically requires that non-linear input-output mappings be determined between a set of variable parameters (such as geometry dimensions and frequency), and the corresponding scattering parameter(s). Support vector regression (SVR) employing an isotropic Gaussian kernel has been widely used for such tasks; this kernel has one tunable hyperparameter that can be optimized (along with the penalty constant C) using a standard procedure that involves a parameter grid search combined with cross-validation. The isotropic kernel however suffers from limited expressiveness, and might provide inadequate predictive accuracy for nonlinear mappings that involve multiple tunable input variables. The present study shows that Bayesian support vector regression using the inherently more flexible Gaussian kernel with automatic relevance determination (ARD) is eminently suitable for highly non-linear modeling tasks, such as the input reflection coefficient magnitude $|S_{11}|$ of broadband and ultrawideband antennas. The Bayesian framework enables efficient training of the multiple kernel ARD hyperparameters—a task that would be computationally infeasible for the grid search/cross-validation approach of standard SVR.**

***Index Terms*—Gaussian processes, regression, slot antennas, support vector machines.**

1. Introduction

The modeling of microwave antennas and devices requires the accurate prediction of latent underlying non-linear input-output mappings, where the input might be tunable geometry parameters and frequency, and the output one or more scattering parameters. Optimization procedures such as genetic algorithms might require thousands of analyses of different geometries of the antenna/structure to be optimized [1]—use of a model, as opposed to direct full-wave simulations, can produce considerable computational savings. In recent years, regression with support vector machines (SVMs) has been frequently used for such modeling tasks—structures that have been characterized in this manner include printed microstrip and slot antennas [2], [3], printed transmission lines [4], [5], and vertical interconnects in microwave packaging structures [6]. Advantages of the SVM formulation include a convex quadratic optimization problem (unlike neural networks, SVMs do not suffer from local minima), and sparseness of solutions (i.e., regression estimates) that are fully characterized by the set of support vectors, a subset of the training set.

In the standard non-linear support vector regression (SVR) formulation, input vectors are projected into a high dimensional feature space by means of a set of basis functions; linear regression is then carried out in this space. A kernel function is used to compute inner products between projections of input vectors in feature space. Performance of the algorithm on a particular data set crucially depends on the choice of kernel; common examples are the linear, polynomial, sigmoid, and Gaussian or radial basis function kernels [7]. A kernel that has been widely used within standard SVR for microwave modeling applications (e.g., [2]–[6]) is the isotropic Gaussian kernel [7]

$$k(\mathbf{x}_i, \mathbf{x}_j) = \exp\left(-\gamma \|\mathbf{x}_i - \mathbf{x}_j\|_2^2\right) \tag{1}$$

where $\mathbf{x}_i$ and $\mathbf{x}_j$ are the ith and jth input vectors of dimension D, $\gamma = 1/(2\sigma^2)$ is an adjustable hyperparameter with σ the kernel width, and $\|\cdot\|_2$ denotes the Euclidean norm (this kernel has also been used in array [8] and direction of arrival [9] applications). Use of the isotropic Gaussian kernel requires that only two hyperparameters need to be optimized during the SVM "training" phase, namely the kernel hyperparameter γ and the penalty constant C. This is usually done by means of a standard procedure that involves a parameter grid search combined with cross-validation [7], [10].

A form of the Gaussian kernel that is inherently more flexible than the isotropic kernel is

$$k(\mathbf{x}_i, \mathbf{x}_j) \propto \exp\left(-\sum_{k=1}^{D} \tau_k (x_{ik} - x_{jk})^2\right) \tag{2}$$

where $\tau_k > 0$ is an automatic relevance determination (ARD) hyperparameter that determines the relevance of the kth component of the input vectors $\mathbf{x}_i$ and $\mathbf{x}_j$ (i.e., x_{ik} and x_{jk}) for the prediction of the output. ARD therefore can be viewed as an in-built mechanism for feature selection. However, optimizing more than two hyperparameters using the above grid search/cross-validation approach is usually not feasible because of exponentially increasing computational costs associated with the multi-dimensional parameter grid.

The present communication demonstrates the applicability to highly non-linear antenna modeling problems of an extended SVR formulation which allows for the hyperparameters of the ARD kernel (2) to be optimized in an efficient and principled manner, namely the Bayesian support vector regression algorithm of Chu *et al.* [11] (henceforth referred to as BSVR). Their approach combines the Bayesian framework for Gaussian process regression (GPR) [12] with standard SVR [7], retaining benefits of both methods such as the convex quadratic programming of SVR; and the GPR-like property that the ARD hyperparameters can be inferred by minimizing the negative log probability of the data given the hyperparameters.

The modeling capability of BSVR (which relies on the effectiveness of its kernel hyperparameter optimization) is demonstrated by means of two case studies involving CPW-fed slot antennas: a square slot with T-shaped tuning stub intended for broadband operation [13], and an ultrawideband (UWB) slot antenna with U-shaped tuning stub [14]. In both cases, highly non-linear underlying functions map several tunable antenna dimensions and frequency to S_{11}. Previously it was shown that GPR could model the UWB antenna with sufficient accuracy for a genetic algorithm based design procedure using this model to achieve a rigorous bandwidth specification [15]—hence an objective was to establish whether BSVR could achieve similar or better accuracy on this modeling task.

The layout of the article is as follows. Section II briefly provides theoretical background to standard SVR [7], BSVR [11], and GPR [12]; key steps and equations are given. Section III describes the implementation of BSVR (as well as GPR and standard SVR for comparison) towards modeling $|S_{11}|$ as a function of the tunable dimensions of each antenna and frequency, and results are discussed. Conclusions are presented in Section IV.

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

2. Theoretical Overview

A. Standard Support Vector Regression

Consider a training data set of n observations, $\mathcal{D} = \{(\mathbf{x}_i, y_i)|i = 1, \ldots, n\}$. The inputs $\mathbf{x}_i$ are column vectors of dimension D, while the corresponding output targets y_i are scalars. Assume that a linear underlying function $f(\mathbf{x}) = \mathbf{w}^\top \mathbf{x} + b$, needs to be estimated and that the loss due to predicting $f(\mathbf{x})$ instead of y is given by the ε-insensitive loss function [7]

$$L_\varepsilon(\mathbf{x}, y, f(\mathbf{x})) = \max\{0, |y - f(\mathbf{x})| - \varepsilon\} \tag{3}$$

where ε is the radius of the regression tube. Use of slack variables ξ_i and ξ_i^* [7] leads to the "soft margin" primal constrained optimization problem [7]

$$\min_{\mathbf{w}, \xi, \xi^*, b} \frac{1}{2}\mathbf{w}^\top \mathbf{w} + C \sum_{i=1}^{n} (\xi_i + \xi_i^*) \tag{4}$$

subject to $\mathbf{w}^\top \mathbf{x}_i + b - y_i \leq \varepsilon + \xi_i$, $y_i - \mathbf{w}^\top \mathbf{x}_i - b \leq \varepsilon + \xi_i^*$, and $\xi_i, \xi_i^* \geq 0$, $i = 1, \ldots, n$; in the above, "$\top$" denotes the transpose operator, and $C > 0$ is a constant determining the trade-off between the complexity of f and the extent to which deviations larger than ε can be allowed. Introducing Lagrange parameters $\boldsymbol{\alpha}, \boldsymbol{\alpha}^*$ leads to the dual problem for non-linear, kernel-based regression [7]:

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

subject to

$$\sum_{i=1}^{n} (\alpha_i - \alpha_i^*) = 0;\ 0 \leq \alpha_i, \alpha_i^* \leq C;\ i = 1, \ldots, n \tag{5}$$

where Σ is a $n \times n$ matrix with $\Sigma_{ij} = k(\mathbf{x}_i, \mathbf{x}_j)$ and $k(\cdot)$ is the kernel function. The regression estimate evaluated at a test data point $\mathbf{x}^*$ is given by [7]

$$f(\mathbf{x}^*) = \sum_{i=1}^{n} (-\alpha_i + \alpha_i^*)\, k(\mathbf{x}_i, \mathbf{x}^*) + b \tag{6}$$

and in effect only takes training points that are support vectors into account.

B. Bayesian Support Vector Machine Regression

The BSVR formulation [11] follows the standard Bayesian regression framework which assumes that targets y_i can be expressed as $y_i = f(\mathbf{x}_i) + \delta_i$, where the δ_i are independent, identically distributed noise variables, and the underlying function f is considered a random field. For $\mathbf{f} = [f(\mathbf{x}_1)\ f(\mathbf{x}_2) \ldots f(\mathbf{x}_n)]$, Bayes's theorem gives the posterior probability of $\mathbf{f}$ given the training data $\mathcal{D}$ as

$$p(\mathbf{f}|\mathcal{D}) = \frac{p(\mathcal{D}|\mathbf{f})p(\mathbf{f})}{p(\mathcal{D})} \tag{7}$$

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 can be expressed as

$$p(\mathcal{D}|\mathbf{f}) = \prod_{i=1}^{n} p(\delta_i) \tag{8}$$

Furthermore, $p(\delta_i) \propto \exp(-\zeta L(\delta_i))$ with $L(\delta_i)$ a loss function and ζ a constant; in standard GPR [12] the loss function is quadratic. The crucial point in the BSVR formulation is that a new loss function, the soft insensitive loss function, is introduced in the likelihood, combining advantageous properties from the ε-insensitive loss function (3) (sparseness of solutions) and Huber's loss function (differentiability). It is defined as [11]

$$L_{\varepsilon,\beta}(\delta) = \begin{cases} -\delta - \varepsilon; & \delta \in (-\infty, -(1+\beta)\varepsilon) \\ \frac{(\delta + (1-\beta)\varepsilon)^2}{4\beta\varepsilon}; & \delta \in [-(1+\beta)\varepsilon, -(1-\beta)\varepsilon] \\ 0; & \delta \in (-(1-\beta)\varepsilon, (1-\beta)\varepsilon) \\ \frac{(\delta - (1-\beta)\varepsilon)^2}{4\beta\varepsilon} & \delta \in [(1-\beta)\varepsilon, (1+\beta)\varepsilon] \\ \delta - \varepsilon; & \delta \in ((1+\beta)\varepsilon, +\infty) \end{cases} \tag{9}$$

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

Solving for the maximum a posteriori (MAP) estimate of the function values then becomes equivalent to solving a new primal problem [11, Eqs. (19)–(21)], with the corresponding dual problem given by

$$\min_{\boldsymbol{\alpha}, \boldsymbol{\alpha}^*} \frac{1}{2}(\boldsymbol{\alpha} - \boldsymbol{\alpha}^*)^\top \Sigma (\boldsymbol{\alpha} - \boldsymbol{\alpha}^*) - \sum_{i=1}^{n} y_i (\alpha_i - \alpha_i^*) + \sum_{i=1}^{n} (\alpha_i + \alpha_i^*)(1-\beta)\varepsilon + \frac{\beta\varepsilon}{C} \sum_{i=1}^{n} y_i \left(\alpha_i^2 + \alpha_i^{*2}\right) \tag{10}$$

subject to $0 \leq \alpha_i, \alpha_i^* \leq C$, $i = 1, \ldots, n$. The regression estimate at a test input $\mathbf{x}^*$ can be expressed as $f(\mathbf{x}^*) = \sum_{i=1}^{n} k(\mathbf{x}_i, \mathbf{x}^*)(\alpha_i - \alpha_i^*)$; (10) reduces to the ε-insensitive dual problem (5) when $\beta = 0$.

Assume the use of an ARD kernel given by

$$k(\mathbf{x}_i, \mathbf{x}_j) = \sigma_f^2 \exp\left(-\frac{1}{2}\sum_{k=1}^{D} \frac{(x_{ik} - x_{jk})^2}{\ell_k^2}\right) + \kappa \tag{11}$$

with hyperparameters $\sigma_f^2; \ell_k, k = 1, \ldots, D$ (the so-called length-scale parameters); and κ. (This kernel was used in all BSVR and GPR regressions reported below). The length-scale associated with a particular input dimension can be viewed as the distance that has to be traveled along that dimension before the output can change significantly [12]. Then the hyperparameter vector $\boldsymbol{\theta}$, which includes the kernel hyperparameters as well as C and ε, can be determined by minimizing the negative log probability of the data given the hyperparameters [11],

$$-\ln p(\mathcal{D}|\boldsymbol{\theta}) = \frac{1}{2}(\boldsymbol{\alpha} - \boldsymbol{\alpha}^*)^\top \Sigma (\boldsymbol{\alpha} - \boldsymbol{\alpha}^*) + C \sum_{i=1}^{n} L_{\varepsilon,\beta}(y_i - f_{MP}(x_i)) + \frac{1}{2}\ln\left|\mathbf{I} + \frac{C}{2\beta\varepsilon}\Sigma_M\right| + n \ln Z_S \tag{12}$$

with $\mathbf{f}_{MP} = \Sigma(\boldsymbol{\alpha} - \boldsymbol{\alpha}^*)$, Σ_M an $m \times m$ submatrix of Σ corresponding to the off-bound support vectors, $\mathbf{I}$ the $m \times m$ identity matrix, and Z_s defined as [11, Eq. (15)].

C. Gaussian Process Regression

In the function space approach to GPR [12], predictions are made by assuming a jointly Gaussian distribution over the n random variables which represent the training outputs, and the n^* random variables representing the test outputs. The distribution of the test outputs conditioned on the known training outputs $\mathbf{y}$ (the posterior distribution) is

again multivariate Gaussian, with mean vector $\mathbf{m}$ and covariance matrix Σ [12]

$$\mathbf{m} = K(X^*, X)K(X, X)^{-1}\mathbf{y} \tag{13}$$

$$\Sigma = K(X^*, X^*) - K(X^*, X)K(X, X)^{-1}K(X, X^*) \tag{14}$$

with $K(X, X^*)$ the $n \times n^*$ matrix of covariances, evaluated using (11), between all possible pairs of n training and n^* test outputs (the columns of the $D \times n$ matrix X are the training input vectors, and X^* contains the test input vectors; $K(X, X)$, $K(X^*, X)$, and $K(X^*, X^*)$ are similarly defined). $\mathbf{m}$ contains the regression estimate associated with the test inputs, and the diagonal of Σ gives the predictive variances. These equations show an exact correspondence with the weight-space approach to GPR involving Bayesian linear regression in feature space [12].

The hyperparameters of (11) may be determined by minimizing the negative log marginal likelihood [12]:

$$-\ln p(\mathbf{y}|X) = \frac{1}{2}\mathbf{y}^\top K^{-1}\mathbf{y} + \frac{1}{2}\ln|K| + \frac{n}{2}\ln 2\pi \tag{15}$$

where $|K|$ is the determinant of $K(X, X)$.

3. Application to Antenna Modeling

A. *Broadband CPW-Fed Square Slot Antenna With T-Shaped Tuning Stub*

Fig. 1 shows a CPW-fed square slot antenna with T-shaped tuning stub on a single dielectric layer substrate, intended for broadband operation [13]. BSVR, GPR and standard SVR were used to model the magnitude of the antenna input reflection coefficient $|S_{11}|$ over a four-dimensional input space spanned by the three tunable stub dimensions u_1, u_2, and u_3, and frequency f. Parameter ranges were $(21 \leq u_1 \leq 36)$ mm, $(16 \leq u_2 \leq 27)$ mm, $(0.3 \leq u_3 \leq 2)$ mm, and $(1.2 \leq f \leq 3)$ GHz. (Fixed dimensions and parameters were $h = 1.6$ mm, $\varepsilon_r = 3.38$, $A = 68$ mm, $B = 44$ mm, $w = 6.37$ mm, and $s = 0.5$ mm.) Visual inspection at randomly selected locations throughout the multi-dimensional geometry space revealed notable variability of $|S_{11}|$-against-frequency responses. For example, in addition to single-band responses exhibiting relatively wide -10 dB bandwidths, narrower responses located towards either end of the frequency range, dual-band responses, and responses that were larger than -10 dB throughout the frequency range were observed. $\mathrm{Re}[S_{11}]$ and $\mathrm{Im}[S_{11}]$, the real and imaginary parts of S_{11}, were modeled separately. This was preferred to setting up a single regression model for $|S_{11}|$, as GPR with a Gaussian kernel performs best on stationary smooth functions [12] (numerical experiments confirmed that this also holds for BSVR).

For training data, 400 geometries comprised of sets of tunable dimensions were uniformly randomly sampled from the input space, with two randomly selected frequencies per geometry (i.e., different frequencies were independently selected for each geometry). This resulted in a set of $n = 800$ training input vectors of the form $\mathbf{x}_i = [u_{1i}, u_{2i}, u_{3i}, f_i]^\top$, $i = 1, \ldots, n$ (where f_i is the ith frequency value). The corresponding output target scalars y_i (either $\mathrm{Re}[S_{11}]$ or $\mathrm{Im}[S_{11}]$) were determined using IE3D [16], a moment-method-based full-wave simulator. Test data were comprised of 150 new randomly selected geometries, with 37 equally spaced frequencies over the range 1.2 to 3 GHz per geometry, resulting in a total of 5550 test points.

During training, the hyperparameters of (11) were determined for each of the BSVR and GPR models by minimizing (12) and (15) respectively. A standard SVR model with isotropic kernel (1) was also trained; its hyperparameters $[C, \gamma]$ optimized using a 5-fold cross-validation procedure were $[2^{15}, 2^{-6.5}]$ for $\mathrm{Re}[S_{11}]$ and $[2^{14}, 2^{-6}]$ for $\mathrm{Im}[S_{11}]$ with $\varepsilon = 0.05$. Predictions were then made for the test data set. Table I gives for each method the percentage normalized root mean square error (%RMSE) for the test set (predictions were normalized with respect to the range of the target values), as well as the linear correlation coefficient R for predictions and targets. BSVR and GPR achieved highly accurate results throughout, with normalized $\mathrm{RMSEs} \leq 0.6\%$ in all cases, and correspondingly high correlation coefficients. In contrast, normalized RMSEs for SVR $\mathrm{Re}[S_{11}]$ and $\mathrm{Im}[S_{11}]$ test predictions were significantly worse (at 4.50% and 5.37% respectively).

Figs. 2 and 3 pertain to a test geometry $[u_1, u_2, u_3] = [29, 22.5, 0.8]$ mm that exhibited a near-optimal fractional -10 dB bandwidth of 67% with respect to the centre of the band at 2.1 GHz. Fig. 2 shows $\mathrm{Re}[S_{11}]$ against frequency as predicted by BSVR,

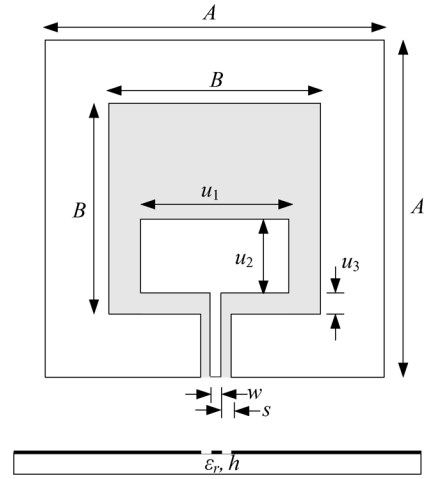

Fig. 1. CPW-fed square slot antenna with T-shaped tuning stub for broadband operation (top and side views). The slot aperture is shaded.

TABLE I
%RMSE AND CORRELATION COEFFICIENT R FOR TEST DATA PREDICTIONS

Antenna	Method	%RMSE		R	
		Re[S_{11}]	Im[S_{11}]	Re[S_{11}]	Im[S_{11}]
Broadband	SVR	4.50	5.37	0.9705	0.9650
	BSVR	0.48	0.33	0.9997	0.9999
	GPR	0.60	0.52	0.9995	0.9997
Ultrawide-band	SVR	8.81	9.60	0.8217	0.7659
	BSVR	1.60	1.51	0.9945	0.9949
	GPR	1.55	1.43	0.9949	0.9955

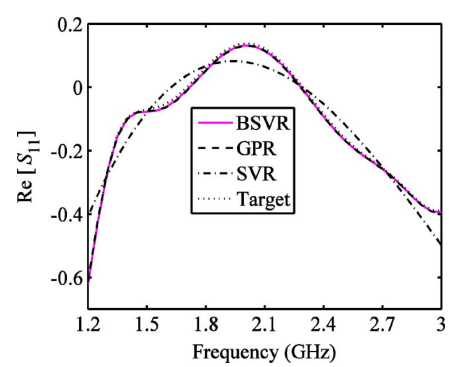

Fig. 2. $\mathrm{Re}[S_{11}]$ against frequency of broadband slot antenna corresponding to test geometry $[u_1, u_2, u_3] = [29, 22.5, 0.8]$ mm.

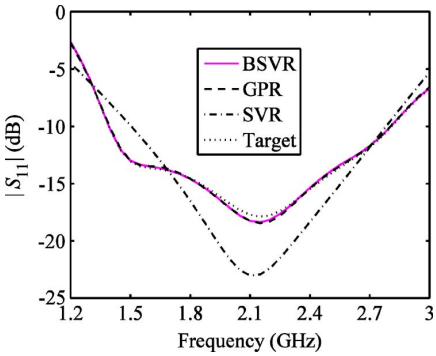

Fig. 3. Magnitude of reflection coefficient $|S_{11}|$ against frequency for broadband slot antenna.

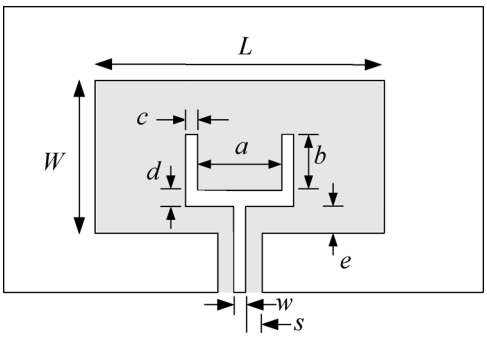

Fig. 4. Ultrawideband CPW-fed slot antenna with U-shaped tuning stub (top view only). The slot aperture is shaded.

GPR, and standard SVR [10]; and the full-wave simulated target values. Fig. 3 shows the corresponding $|\mathrm{S}_{11}|$ plots constructed from the $\mathrm{Re}[S_{11}]$ and $\mathrm{Im}[S_{11}]$ predictions (the normalized RMSEs for the 37 $\mathrm{Re}[S_{11}]$ test predictions were 0.51%, 0.60%, and 4.48% for BSVR, GPR, and SVR respectively; and 0.16%, 0.37%, and 3.75% for the corresponding $\mathrm{Im}[S_{11}]$ predictions). While BSVR and GPR predictive curves correspond closely to full-wave simulations, the limited expressiveness of the isotropic kernel can be seen in the failure of the SVR curves to replicate the local peaks and valleys of the target curves.

It may be noted that both the BSVR and GPR formulations provide the option of plotting confidence regions about the mean, which is a consequence of their probabilistic frameworks. For the sake of clarity these have been omitted from Fig. 2; given the high accuracy of the predictions the confidence regions were in any event very narrow about the respective means.

B. Ultrawide-Band CPW-Fed Slot Antenna With U-Shaped Tuning Stub

Fig. 4 shows a CPW-fed rectangular slot antenna with U-shaped tuning stub intended for ultrawideband operation [14]. The substrate was a single dielectric layer with height $h = 0.813$ mm and dielectric constant $\varepsilon_r = 3.38$. BSVR and GPR were used to model S_{11} over a six-dimensional input space spanned by the five tunable U-stub dimensions a, b, c, d, and e, and frequency f. Parameter ranges were $(6 \leq a \leq 14)$ mm, $(6 \leq b \leq 12)$ mm, $(0.5 \leq c \leq 2.5)$ mm, $(0.5 \leq d \leq 2.5)$ mm, $(0.5 \leq e \leq 2.5)$ mm, and $(2 \leq f \leq 10)$ GHz (fixed dimensions were $W = 32.2$ mm, $L = 21.1$ mm, $w = 1.88$ mm and $s = 0.125$ mm; infinite ground planes and dielectrics were assumed). Preliminary inspection revealed substantial variability of $|S_{11}|$-against-frequency responses throughout the geometry input space.

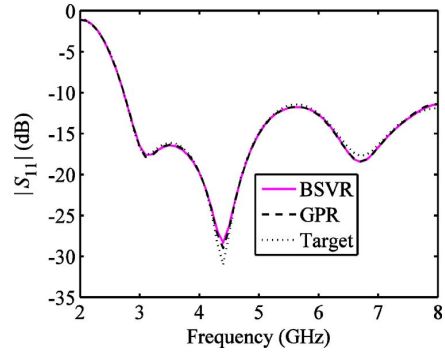

Fig. 5. $|S_{11}|$ against frequency for ultrawideband slot antenna constructed from the regression models for the real and imaginary parts of S_{11} for the test geometry $[a, b, c, d, e] = [12, 9, 2, 1, 1]$ mm.

Training data consisted of 625 geometries (randomly sampled sets of tunable dimensions), with seven frequencies per geometry selected randomly as before, resulting in 4375 training points (more frequencies per geometry were required, as $\mathrm{Re}[S_{11}]$ and $\mathrm{Im}[S_{11}]$ responses were less smooth compared to the broadband case). Test data consisted of 176 randomly selected geometries not used for training, with 81 equally spaced frequencies over the range 2 to 10 GHz per geometry, giving a total of 14256 test points.

Table I gives the %RMSE and R for each regression (including standard SVR with an isotropic kernel). For both BSVR and GPR the predictive accuracy was very good given the highly non-linear underlying function that had to be fitted, with normalized $\mathrm{RMSEs} \leq 1.6\%$ in all cases. Not unexpectedly—given the limited expressivity of its isotropic kernel with respect to the multi-dimensional input space—standard SVR fared substantially worse, with normalized RMSEs for $\mathrm{Re}[S_{11}]$ and $\mathrm{Im}[S_{11}]$ of respectively 8.81% and 9.60%. Fig. 5 shows $|S_{11}|$ against frequency, constructed from the $\mathrm{Re}[S_{11}]$ and $\mathrm{Im}[S_{11}]$ values predicted by BSVR and GPR, for a geometry that displays ultrawideband behavior, namely $[a, b, c, d, e] = [12, 9, 2, 1, 1]$ mm (the normalized RMSEs for the 81 $\mathrm{Re}[S_{11}]$ test predictions were 0.83% and 1.04% for BSVR and GPR; and 1.00% and 0.93% for the corresponding $\mathrm{Im}[S_{11}]$ predictions). Very good agreement with the full-wave-simulated results is exhibited by both methods. The BSVR kernel length-scale parameters for the $\mathrm{Re}[S_{11}]$ regression corresponding to input variables $[a, b, c, d, e, f]$ were $[\ell_1, \ell_2, \ell_3, \ell_4, \ell_5, \ell_6] = [1.287, 1.091, 4.137, 2.858, 3.297, 0.204]$, and [1.299, 1.095, 4.760, 2.918, 3.134, 0.204] for $\mathrm{Im}[S_{11}]$, indicating that the input dimensions most influential on output predictions were frequency and the stub vertical and horizontal dimensions b and a.

4. Conclusion

Results were presented confirming that Bayesian support vector regression (which is an extended formulation of the widely used standard SVR) with a Gaussian ARD kernel is eminently suitable for highly nonlinear regression problems such as modeling the input reflection coefficient of antennas with multiple tunable geometry variables. BSVR provides a systematic and efficient mechanism for inferring the multiple ARD kernel hyperparameters (i.e., minimization of the negative log probability of the data given the hyperparameters)—a task which is computationally infeasible for standard SVR with its grid search/cross-validation approach to hyperparameter optimization.

While BSVR achieved predictive accuracies similar to those obtained from GPR for these particular two antennas, the BSVR formulation has (because of its SVR-related properties) certain general advantages over GPR that are of direct interest for antenna modeling. The first pertains to efficiency in handling large training data sets: the

basic computational cost of GPR is $O(n^3)$ [12], due to the fact that an $n \times n$ matrix needs to be inverted for both the gradient calculations required for hyperparameter optimization [12, eq. (5.9)]; and inference (13). This computational expense may become prohibitive for large n. In contrast, even though hyperparameter optimization under BSVR likewise requires matrix inversion for computing the gradient of the evidence [11, eqs. (39)–(41)]), this involves the $m \times m$ matrix Σ_M (cf. (12)), where m usually is considerably smaller than n [11]. To make predictions, BSVR does not require matrix inversion, but rather solving of the convex quadratic optimization problem (10), for which multiple efficient techniques exist [7]. Second, the sparseness property of BSVR can be used towards adaptive data selection when training data is expensive to generate. The sparseness property alludes to the fact that the predictive function can be expressed as a weighted sum of kernel functions centered at the support vectors (SVs), where the SVs are a (usually significantly) reduced subset of the training input vectors. Suppose that a "coarse" regression model using many inexpensive coarsely simulated training data points is set up, and that SVs are identified. It has been shown that the SVs, re-simulated at a high meshing density, can form a sufficient training set for an accurate "fine" model [17]—resulting in substantial computational savings compared to when the full original training set is simulated at the high meshing density. Finally, BSVR's soft insensitive loss function (9) is not overly sensitive to outliers that might arise if training data were obtained from noisy measurements. In contrast, outliers contribute disproportionally to the quadratic loss function in GPR—to the extent that relatively few outliers could adversely affect the solution [11].

It is anticipated that BSVR might be applied with good effect to other antenna-related regression problems that involve non-linear input-output relationships and multi-dimensional inputs.

References

[1] Y. Kim, S. Keely, J. Ghosh, and H. Ling, "Application of artificial neural networks to broadband antenna design based on a parametric frequency model," *IEEE Trans. Antennas Propag.*, vol. 55, pp. 669–674, 2007.

[2] 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.

[3] N. Chauhan, A. Mittal, and M. V. Kartikeyan, "Support vector driven genetic algorithm for the design of circular polarized microstrip antenna," *Int J Infrared Millim. Waves*, vol. 29, pp. 558–569, 2008.

[4] F. Gunes, N. T. Tokan, and F. Gurgen, "A consensual modeling of the expert systems applied to microwave devices," *Int. J. RF Microw. Comput.-Aided Engrg.*, vol. 20, pp. 430–440, 2010.

[5] F. Gunes, N. T. Tokan, and F. Gurgen, "A knowledge-based support vector synthesis of the transmission lines for use in microwave integrated circuits," *Expert Syst. Appl.*, vol. 37, pp. 3302–3309, 2010.

[6] L. Xia, J. C. Meng, R. M. Xu, B. Yan, and Y. C. Guo, "Modeling of 3-D vertical interconnect using support vector machine regression," *IEEE Microw. Wireless Compon. Lett.*, vol. 16, pp. 639–641, 2006.

[7] B. Schölkopf and A. J. Smola, *Learning With Kernels: Support Vector Machines, Regularization, Optimization, and Beyond*. Cambridge, MA: MIT Press, 2002.

[8] M. Martinez-Ramon, J. L. Rojo-Alvarez, G. Camps-Valls, and C. G. Christodoulou, "Kernel antenna array processing," *IEEE Trans. Antennas Propag.*, vol. 55, pp. 642–650, 2007.

[9] A. Randazzo, M. A. Abou-Khousa, M. Pastorino, and R. Zoughi, "Direction of arrival estimation based on support vector regression: Experimental validation and comparison with MUSIC," *IEEE Antennas Wireless Propag. Lett.*, vol. 6, pp. 379–382, 2007.

[10] C. C. Chang and C. J. Lin, LIBSVM: A Library for Support Vector Machines 2001.

[11] W. Chu, S. S. Keerthi, and C. J. Ong, "Bayesian support vector regression using a unified loss function," *IEEE Trans. Neural Netw.*, vol. 15, pp. 29–44, 2004.

[12] C. E. Rasmussen and C. K. I. Williams, *Gaussian Processes for Machine Learning*. Cambridge, MA: MIT Press, 2006.

[13] H. D. Chen, "Broadband CPW-fed square slot antennas with a widened tuning stub," *IEEE Trans. Antennas Propag.*, vol. 51, pp. 1982–1986, 2003.

[14] R. Chair, A. A. Kishk, and K. F. Lee, "Ultrawide-band coplanar waveguide-fed rectangular slot antenna," *IEEE Antennas Wireless Propag. Lett.*, vol. 3, pp. 227–229, 2004.

[15] J. P. Jacobs and J. P. De Villiers, "Gaussian-process-regression-based design of ultrawide-band and dual-band CPW-fed slot antennas," *J. Electromagn. Waves Applicat.*, vol. 24, pp. 1763–1772, 2010.

[16] Zeland Software, IE3D User's Manual vol. 14, 2007.

[17] N. T. Tokan and F. Gunes, "Knowledge-based support vector synthesis of the microstrip lines," *Progr. Electromagn. Res.*, vol. 92, pp. 65–77, 2009.