

Received April 24, 2021, accepted May 6, 2021, date of publication May 10, 2021, date of current version May 19, 2021.

Digital Object Identifier 10.1109/ACCESS.2021.3078432

Improved Modeling of Microwave Structures Using Performance-Driven Fully-Connected Regression Surrogate

SLAWOMIR KOZIEL^{1,2}, (Senior Member, IEEE), PEYMAN MAHOUTI³, NURULLAH CALIK⁴, MEHMET ALI BELEN⁵, AND STANISLAW SZCZEPANSKI²

¹Engineering Optimization and Modeling Center, Department of Technology, Reykjavik University, Reykjavik 102, Iceland

²Faculty of Electronics, Telecommunications and Informatics, Gdansk University of Technology, 80-233 Gdansk, Poland

³Department of Electronic and Communication, Vocational School of Technical Sciences, Istanbul University–Cerrahpasa, 34500 Istanbul, Turkey

⁴Department of Biomedical Engineering, Istanbul Medeniyet University, 34700 Istanbul, Turkey

⁵Department of Electrical and Electronic Engineering, Iskenderun Technical University, 31200 Iskenderun, Turkey

Corresponding author: Slawomir Koziel (koziel@ru.is)

This work was supported in part by the Icelandic Centre for Research (RANNIS) under Grant 206606051, and in part by the National Science Centre of Poland under Grant 2020/37/B/ST7/01448.

ABSTRACT Fast replacement models (or surrogates) have been widely applied in the recent years to accelerate simulation-driven design procedures in microwave engineering. The fundamental reason is a considerable—and often prohibitive—CPU cost of massive full-wave electromagnetic (EM) analyses related to solving common tasks such as parametric optimization or uncertainty quantification. The most popular class of surrogates are data-driven models, which are fast to evaluate, versatile, and easy to handle. Notwithstanding, the curse of dimensionality as well as the utility demands (e.g., so that the model covers sufficiently broad ranges of the system operating conditions), limit the applicability of conventional methods. A performance-driven modeling paradigm allows for mitigating these issue by focusing the surrogate setup process in a constrained domain encapsulating designs being of high quality w.r.t. the assumed figures of interest. The nested kriging framework capitalizing on this idea, renders the constrained surrogate using kriging interpolation, and has been shown to surpass traditional approaches. In pursuit of further accuracy improvements, this work incorporates the performance-driven concept into the fully-connected regression model (FCRM). The latter has been recently introduced in the context of frequency selective surfaces, and combined deep neural networks with Bayesian optimization, the latter employed to determine the network architecture and hyper-parameters. Using two examples of miniaturized microstrip couplers, our methodology is demonstrated to outperform both conventional modeling techniques and nested kriging, with reliable models constructed over multi-dimensional parameters spaces using just a few hundreds of samples.

INDEX TERMS Data-driven modeling, surrogate modeling, performance-driven surrogates, nested kriging, deep regression model, Bayesian optimization.

I. INTRODUCTION

Contemporary microwave design is a challenging task, where the strive to meet stringent performance specifications is compromised by the necessity of satisfying various constraints related to device cost or its physical dimensions [1], [2]. Due to the complexity of modern microwave structures, one of the fundamental design tools is electromagnetic (EM)

simulation. While ensuring evaluation reliability, it may incur considerable computational expenditures. These are particularly pronounced when repetitive EM analyses are required, e.g., in the case of geometry parameter tuning [3], estimation of fabrication yield [4], or seeking for trade-off designs at the presence of multiple objectives [5]. The methods for accelerating EM-driven design procedures have been researched for over two decades [6], [7]. Available approaches include incorporation of adjoint sensitivities [8], [9], or sparse Jacobian updates [10], [11] into gradient-based

algorithms, utilization of custom EM solvers [12], as well as mesh deformation techniques [13]. Other methods rely on variable-fidelity or variable-resolution techniques [14], [15], often combined with physics-based surrogate modeling procedures (e.g., space mapping [16], response correction [17], Bayesian model fusion [18], co-kriging [19]). The latter are mostly applicable to local optimization purposes [20]. Surrogate-assisted methodologies involving data-driven (or approximation) models constitute a separate class of growing popularity. Widely used techniques include polynomial regression [21], kriging [22], radial basis functions [23], neural networks [24], or polynomial chaos expansion [25], often combined with sequential sampling schemes [26] or machine learning frameworks [27]. In practical scenarios, the meta-model is sequentially refined using the EM simulation data accumulated during the optimization process [28], with the infill criteria oriented towards improving the predictive power [29], or optimum identification [30].

From the perspective of simulation-driven design, replacing EM simulations by a fast surrogate is an ideal solution as it enables a rapid execution of a variety of tasks, including parametric optimization of microwave components. Unfortunately, constructing globally accurate surrogates is seriously hindered by the curse of dimensionality. Also, practically useful model has to cover a sufficiently broad range of the system parameters and operating frequencies, which is even more challenging given a typically high nonlinearity of microwave circuit characteristics. Possible workarounds include reduction of the problem dimensionality (model order reduction techniques [31], principal component analysis [32]), high-dimensional model representation (HDMR) [33], as well as variable-resolution approaches [34], [35]. On the other hand, efficient handling of nonlinear responses can be realized using deep learning (DL) procedures [36] with the representative example being Deep Neural Networks (DNN) [36], [38]. Nevertheless, DNN-based modeling is challenging due to the necessity of appropriate adjustment of the network architecture and its hyper-parameters, as well as to address potential issues such as overtraining [39], [40]. The latter can be avoided by automated architecture determination involving numerical optimization methods [41]. A recent example is a fully-connected regression model (FCRM) [42], where all components of the DNN, including its architecture, are adjusted through Bayesian optimization (BO) [43]. Yet another possibility is ensemble learning (EL) [44], where individual models (referred to as learners) are combined as building blocks of a more involved surrogate [45]. Unfortunately, neither a selection nor integration of the building blocks is a trivial task [46]. Performance-driven modeling methods [47] foster an entirely different way of handling the dimensionality problems by confining the surrogate model domain to a selected region of the parameter space, which contains high-quality designs with respect to the performance figures pertinent to the structure of interest. The constrained domain of the surrogate model is determined using pre-optimized reference designs and an inverse regression model,

whereas the final surrogate is constructed using kriging interpolation [48]. The domain confinement concept has been shown to significantly improve the modeling accuracy, and enabled reliable surrogate rendition over wide ranges of geometry parameters [49].

This paper proposes a novel approach to modeling of miniaturized microwave components. Our methodology combines the fully-connected regression model [42] involving Bayesian optimization for automated determination of the underlying DNN architecture, and the concept of domain confinement as formulated in the nested kriging framework [49]. The employment of FCRM allows us to more efficiently handle nonlinear frequency characteristics of the microwave components as well as account for specific allocation of the training data (particularly crucial for smaller data sets). On the other hand, setting up the model in a constrained domain considerably extends the range of applicability of FCRM in terms of the parameter space dimensionality of parameter ranges. Numerical verification of the presented technique is carried out using two miniaturized microstrip couplers, a rat-race, and a branch-line ones. Comprehensive benchmarking involves several conventional and performance-driven modeling techniques, and reveals superiority of the constrained FCRM in terms of the modeling accuracy as well as computational efficiency. On the one hand, it takes a full advantage of the benefits offered by domain confinement. On the other hand, it allows for exploitation of the inherent suitability of the DNN-based surrogates for representing nonlinear responses of microwave systems. In particular, the predictive power of the proposed models is better than that of the nested kriging framework, by a factor of 2.5 on the average. The modeling methodology presented in this paper can be viewed as a low-cost complement of conventional methods, especially for handling difficult cases featuring higher-dimensional parameter spaces and limited computational budget for training data acquisition.

II. PERFORMANCE-DRIVEN FULLY-CONNECTED REGRESSION MODEL

This section describes the proposed modeling methodology. We start by recalling the formulation of fully-connected regression model (FCRM) with automated architecture determination (Sections II. A and II. B), as well as the performance-driven modeling concept (Section II. C). The complete modeling procedure is summarized in Section II. D.

A. FULLY-CONNECTED REGRESSION MODEL

Here, we briefly formulate the fully-connected regression model (FCRM) [42], which is one of the two major components of the modeling methodology presented in this work. FCRM combines deep neural networks (DNN), and fully automated adjustment procedure that involved Bayesian optimization (BO) [43] (cf. Section II. B). DNNs feature an increased number of hidden layers as compared to conventional (shallow) neural networks, which offers certain advantages, in particular, improved flexibility in modeling

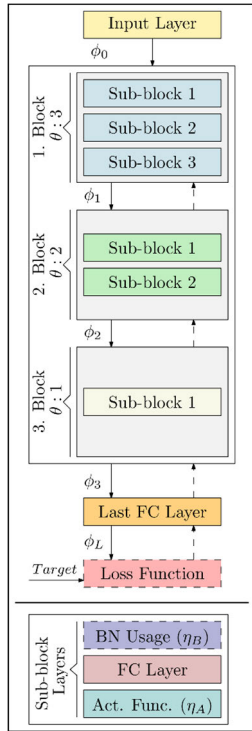


FIGURE 1. The generic architecture of FCRM, here, shown with $m = 3$ blocks [42]. The specific model architecture is obtained through Bayesian Optimization (BO) (Section II. B). In particular, BO provides the number θ of sub-blocks and the number ϕ of neurons for each block, decides about using BN ($\eta_B \in \{0, 1\}$), and selects the activation function ($\eta_A \in \{0, 1, 2\}$). The training phase is based on the Mean Absolute Error (MAE), employed as the loss function. Dashed lines mark the data flow during the training process (gradient values ϕ_m). The bottom panel shows the internal block structure.

of nonlinear system outputs [50]. On the other hand, their training process, referred to as Deep Learning (DL), is considerably more involved [51].

Figure 1 shows a block diagram of FCRM. Therein, one can distinguish several types of functional units, which include:

- *Blocks (BC)*. These are the highest-level units determining the overall architecture of the surrogate. Flexibility of the model in terms of its data processing capabilities is ensured by varying the number of neurons between the blocks. The number m of BCs is the only parameter not adjusted automatically in the training process (Section II. B).
- *Sub-blocks (SB)*. SBs establish the internal structure of the BCs. Their number is automatically determined as a part of the training phase.
- *Layers*. These are the lowest-level components of FCRM, by themselves consisting of three parts: batch normalization (BN), fully-connected layer (FC), and an activation function. The latter can be one of the three types: ReLU, Leak ReLU, or Tanh (see a description below).

The activation function plays an important role in handling the data processed by the network. In particular, in FCRM, the layers act as multi-layer perceptrons, with each

neuron receiving information from all neurons of the preceding layer [42]. A particular way of providing this information is decided upon by the activation function, which should be a nonlinear one to ensure sufficient flexibility of the model in representing the input data. While sigmoid and Tanh functions are preferred in conventional ANNs [52], the vanishing gradient phenomenon [53] makes them insufficient in the case of DNN training. The Rectified Linear Unit (ReLU) activation functions [54], exhibit much better properties with this respect, which are further improved in a *leaky* ReLU [55], defined as

$$\text{ReLU}_\alpha(x) = \begin{cases} x, & x \geq 0 \\ \alpha x, & \text{otherwise} \end{cases} \quad (1)$$

Both ReLU, leaky ReLU, and Tanh are used as potential activation functions in FCRM [42].

Batch normalization (BN) [56] is another component employed in FCRM in order to alleviate certain difficulties pertinent to DNN training, which is normally realized using gradient-based procedures. In the training process, the weights within the internal layers are adjusted based on their corresponding input data, which is obtained from the preceding layers, and, consequently, altered at the end of a particular training stage. As a result, there is a need to accommodate these changes and re-adjust the layers, which is known as the *internal covariate shift* [57]. Batch normalization allows for improving the stability of the training process through a normalization of the layer means and variances, performed on the subsets of the input dataset (referred to as mini-batches) rather than on its entirety. The normalization is realized as an affine transformation with the multiplication factor γ and the shift β , which are adjusted in the training process (cf. Section II. B).

All components of FCRM, including the number of layers, the number of layer neurons, the activation type, and optional incorporation of BN are determined in the learning process, and by accounting for the available training data, as elaborated on in Section II. B. It should be emphasized that the FCRM architecture can be automatically adjusted to the specific input data so that no user interaction/expert knowledge needs to be engaged [42].

B. ARCHITECTURE AND HYPER-PARAMETER SETUP BY BAYESIAN OPTIMIZATION

In FCRM, the adjustment of internal model parameters discussed in Section II. A is realized using Bayesian Optimization (BO) [43]. BO is a population-based procedure incorporating an underlying surrogate model [58]. The surrogate, generated using the initially acquired points, allows for rendering a prior distribution of the probabilistic model. The specific model associated with BO is the Gaussian Process (GP) [59]. The GP model defined to predict the objective function is minimized according to the assumed acquisition function, and it is further refined using the data accumulated during the optimization process.

The infill points are allocated using the acquisition function that helps to achieve a balance between exploration and exploitation of the parameter space [60]. Exploration aims at identifying new and promising regions not visited before. Whereas, exploitation generates the samples based on the posterior distribution within already explored regions, which are likely to contain the global optimum [60].

Given the objective function $f: X \rightarrow R$, BO attempts to determine the global minimum $x^* \in \operatorname{argmin}_{x \in X} f(x)$. BO requires an initial knowledge of a prior distribution $p(f)$ over the objective function f and an acquisition function $a_{p(f)}: X \rightarrow R$. The former encodes information about the promising input space locations; the latter guides the search towards the optimum. The major steps of BO include: (i) finding the most suitable $x_i \in \operatorname{argmax} a_{p(f)}(x)$ through direct optimization; (ii) evaluating $y_i = f(x_i)$, and adding the pair (x_i, y_i) to the observation set $D_i = \{x_i, f(x_i)\}_{i=1, \dots, n}$; (iii) updating the posterior distribution $p(f|D_i)$ and $a_{p(f|D_i)}$. As mentioned before, the preferred statistical model for BO is Gaussian Process [59]. The details of a particular GP implementation utilized in FCRM can be found in [42] and are omitted here for the sake of brevity.

The application of BO to FCRM surrogate training is described in the remaining part of this section. The number m of block is assumed to be fixed (set by the user). BO determines the number θ of sub-blocks and the number ϕ of neurons for each block (the same for each sub-block). Furthermore, it decides about the usage of batch normalization within the sub-blocks (parameter η_B), and selects the activation function (parameter η_A). Consequently, the FCRM control parameters are $\mathcal{S} = \{\Theta \in \mathbb{Z}^{(m \times 1)}, \Phi \in \mathbb{Z}^{(m \times 1)}, \eta_A, \eta_B\}$. As the sub-block and neuron sizes are block-dependent, the corresponding parameters are grouped in Θ and Φ , respectively. The activation function and the usage of BN are the same for the entire network, therefore, parameters η_A and η_B are scalars.

Figure 2 summarizes the FCRM architecture adjustment using BO (see [42] for more details). At the beginning of the process, BO randomly generates an initial set of control parameters S^t ($t = 0$). The Mean Absolute Error (MAE) is employed as the error metric for k -fold average score (L_{avg}) which also is the cost function of the BO process. The estimated set of improved hyper-parameters is obtained based on the average score of previously observed parameter set, with the goal being a reduction of L_{avg} . The termination condition is the maximum number of iterations, here, set to 30.

C. INCORPORATION DOMAIN CONFINEMENT

In this work, the inherent capability of the FCRM surrogate to represent nonlinear responses of high-frequency components is enhanced by combining it with the performance-driven modeling concept [47], which is briefly outlined in this section, including the analytical definition of the confined domain.

The two fundamental components to consider are the objective space F and design optimality. Given the figures

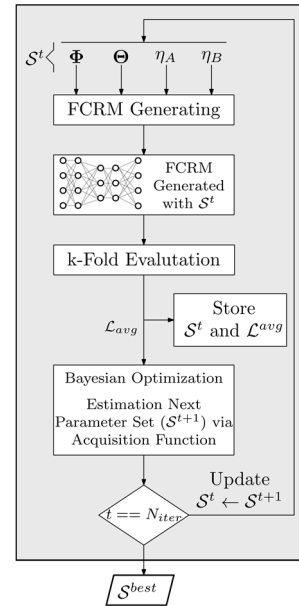


FIGURE 2. Block diagram of neural architecture search using Bayesian optimization [42].

of interest f_k , $k = 1, \dots, N$, e.g., target operating frequency or power split ratio, permittivity of the substrate the circuit is implemented on, the space F contains aggregated objectives $\mathbf{f} = [f_1 \dots f_N]^T$. The region of validity of the surrogate model is determined by the lower and upper bounds $f_{k,\min}$ and $f_{k,\max}$, $k = 1, \dots, N$, for f_k . The design optimality is defined in terms of the objective function $U(\mathbf{x}, \mathbf{f})$, assessing the design \mathbf{x} for the objective vector \mathbf{f} .

More specifically, the optimum design \mathbf{x}^* is understood as

$$\mathbf{x}^* = U_F(\mathbf{f}) = \operatorname{argmin}_{\mathbf{x}} U(\mathbf{x}, \mathbf{f}) \quad (2)$$

For example, suppose that the circuit of interest is a microwave coupler, which is to operate at frequency f_0 and to provide a power split ratio K . Furthermore, let us assume that the surrogate model of the coupler should be valid with in the following ranges: $f_{0,\min} \leq f_0 \leq f_{0,\max}$, and $K_{\min} \leq K \leq K_{\max}$. The objective space F is then $[f_{0,\min}, f_{0,\max}] \times [K_{\min}, K_{\max}]$, whereas the objective function may be defined as $U(\mathbf{x}, \mathbf{f}) = \max\{|S_{11}(\mathbf{x}, f_0)|, |S_{41}(\mathbf{x}, f_0)|\} + \beta(d_S(\mathbf{x}, f_0) - K)^2$, where $d_S(\mathbf{x}, f_0) = |S_{21}(\mathbf{x}, f_0)| - |S_{31}(\mathbf{x}, f_0)|$ (in dB) is the power split ratio at the design \mathbf{x} and frequency f . This objective function corresponds to the design goals being minimization of the coupler matching and isolation characteristics at f_0 while maintaining the power split ratio of K .

The set $U_F(F) = \{U_F(\mathbf{f}) : \mathbf{f} \in F\}$ consists of designs that are optimal according to (2) for all $\mathbf{f} \in F$. Note that as long as the aim is to construct the surrogate that adequately represents designs being of high quality with respect to vectors \mathbf{f} within F , it is sufficient to restrict the modeling process to the vicinity of $U_F(F)$ [47]. In the nested kriging framework [48], $U_F(F)$ is approximated by a set of reference designs $\mathbf{x}^{(j)} = [x_1^{(j)} \dots x_n^{(j)}]^T$, $j = 1, \dots, p$, optimized w.r.t. the $\mathbf{f}^{(j)} = [f_1^{(j)} \dots f_N^{(j)}]$, uniformly distributed in F .

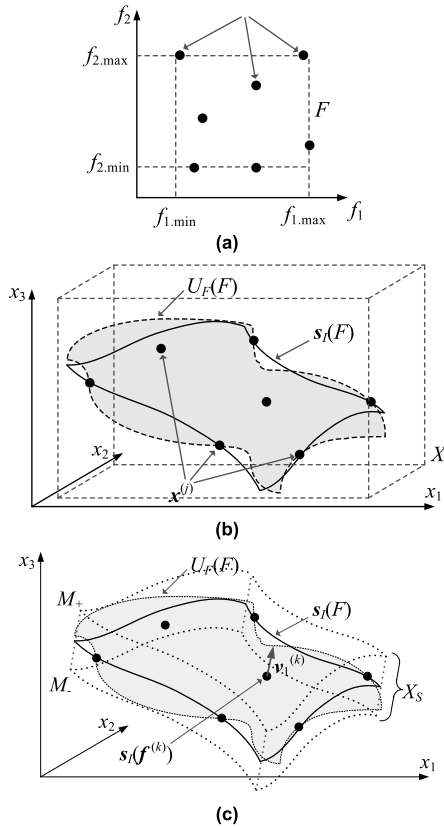


FIGURE 3. Conceptual illustration of performance-driven modeling: (a) objective space; (b) parameter space X , reference designs, and the optimum design manifold $U_F(F)$, as well as the image $s_I(F)$ of the first-level surrogate; (c) definition of the domain X_S through orthogonal extension of $s_I(F)$. The manifolds M_- and M_+ are marked using the dotted lines.

Using the pairs $\{f^{(j)}, \mathbf{x}^{(j)}\}$, $j = 1, \dots, p$, the first-level model $s_I(f): F \rightarrow X$ is established using kriging interpolation. Because s_I coincides with $U_F(F)$ only at the reference designs, $s_I(F)$ is extended to encapsulate the optimum design manifold. The extension is carried out using the vectors $\{v_n^{(k)}(f)\}$, $k = 1, \dots, n - N$, normal to $s_I(F)$ at f . Let $\mathbf{x}_d = \mathbf{x}_{\max} - \mathbf{x}_{\min}$, with $\mathbf{x}_{\max} = \max\{\mathbf{x}^{(k)}, k = 1, \dots, p\}$, $\mathbf{x}_{\min} = \min\{\mathbf{x}^{(k)}, k = 1, \dots, p\}$, and α be the extension factors $\alpha(f) = [\alpha_1(f) \dots \alpha_{n-N}(f)]^T = 0.5T [|\mathbf{x}_d v_n^{(1)}(f)| \dots |\mathbf{x}_d v_n^{(n-N)}(f)|]^T$, where T is a thickness coefficient. The surrogate model domain X_S is then defined as

$$X_S = \left\{ \mathbf{x} = s_I(f) + \sum_{k=1}^{n-N} \lambda_k \alpha_k(f) v_n^{(k)}(f) : f \in F, -1 \leq \lambda_k \leq 1, k = 1, \dots, n - N \right\} \quad (3)$$

Note that X_S is delimited by $M_{\pm} = \{\mathbf{x} \in X : \mathbf{x} = s_I(f) \pm \sum_{k=1}^{n-N} \alpha_k(f) v_n^{(k)}(f)\}$. A graphical illustration of the discussed concepts can be found in Fig. 3. The actual (final) surrogate is constructed in X_S using the data set $\{\mathbf{x}_B^{(k)}, \mathbf{R}(\mathbf{x}_B^{(k)})\}_{k=1, \dots, N_B}$, uniformly distributed within the domain [48]. In nested kriging, the surrogate is rendered using kriging interpolation. Here, the FCRM

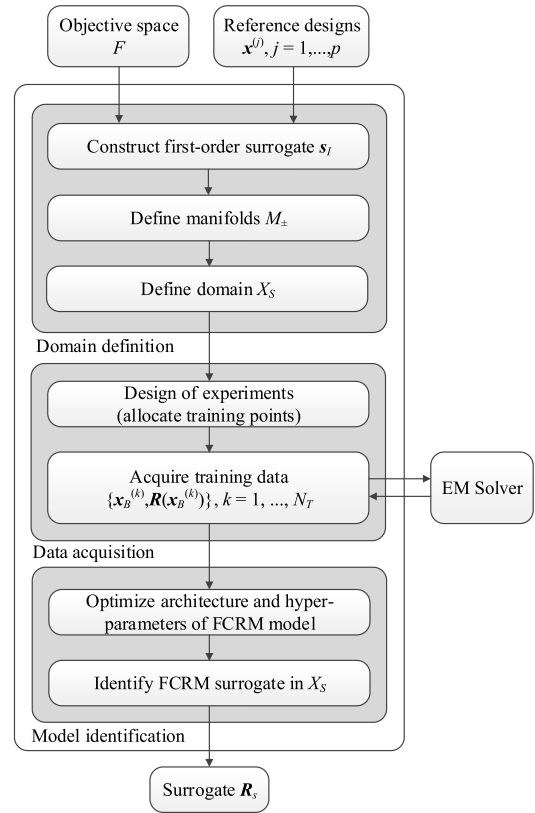


FIGURE 4. Flowchart of the proposed modeling framework involving PDRN with automated architecture determination and domain confinement.

model is used instead to capitalize on improved flexibility of DNNs.

D. MODELING PROCEDURE

The overall modeling procedure consists of several stages, cf. Fig. 4. Having the objective space and the objective function definition decided upon by the user, the reference designs are acquired (some of which may be available from, e.g., the previous work with the same circuit). These are used to define the model domain, the sampling process (cf. [48]), and acquisition of the training data. Subsequently, the FCRM surrogate is rendered using Bayesian optimization as discussed in Section II. B.

III. DEMONSTRATION EXAMPLES

This section discusses numerical verification of the modeling procedure introduced in the work. It involves two miniaturized microstrip couplers modeled over broad ranges of operating frequencies, power split ratio, and substrate permittivity. The constrained FCRM surrogate is established in both a conventional, interval-like, domain (to validate the advantages of FCRM in modeling scattering parameters of the considered microwave components), and the constrained domain defined as in Section II. C (to verify the benefits of incorporating performance-driven concepts into

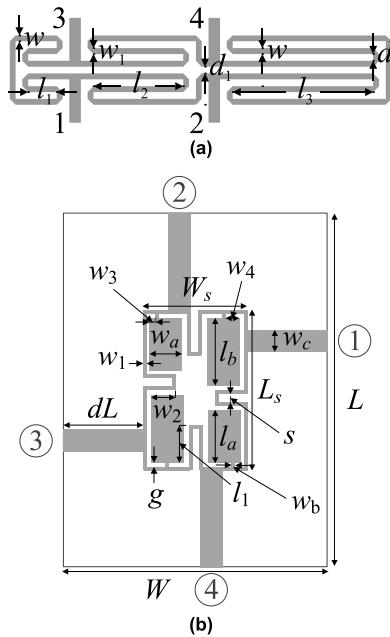


FIGURE 5. Compact microstrip circuit used as verification case studies: (a) rat-race coupler (RRC) [61], (b) branch-line coupler (BLC) [62].

the DNN-based metamodels). This is complemented by a comprehensive comparison with several benchmark methods, including conventional techniques, as well as the nested kriging framework.

A. VERIFICATION CASE STUDIES

Consider two miniaturized microstrip couplers shown in Fig. 5, a rat-race coupler (RRC) [61], and a branch-line coupler (BLC) [62]. The details concerning the substrates, design variables, parameter and objective spaces, as well as reference designs can be found in Table 1. The computational models of both circuits are implemented in CST Microwave Studio. Note that for the BLC, relative permittivity ϵ_r of the dielectric substrate is one of the components of the objective space. The surrogate models are supposed to be valid within the specified ranges of the figures of interest, defining the space F . It should be emphasized that both modeling tasks are very challenging, not only due to the dimensionality of the parameter space (six and ten, respectively) but mainly due to broad ranges of the geometry parameters (the average ratio of the upper-to-lower bounds of three) and the operating conditions. These are 1 GHz to 2 GHz (operating frequency), -6 dB to 0 dB (power split ratio), and 2.0 to 5.0 (substrate permittivity)

B. EXPERIMENTAL SETUP

The surrogate models are constructed within the objective spaces described in Table 1 (for constrained, or performance-driven models), and within the parameter space X (also provided in Table 1) for conventional surrogates. The conventional space corresponds to the smallest interval

containing all reference designs selected for a particular test case. In order to investigate the scalability properties of the models as a function of training data set size, the surrogates are constructed using $N_B = 50, 100, 200, 400,$ and 800 samples. The constrained domain X_S is established using the thickness parameter $T = 0.05$.

The modeling error is computed using a split-sample approach with 100 testing samples. The assumed error metric is the relative RMS error defined as $\|R(x) - R_s(x)\|/\|R(x)\|$ (here, R and R_s stand for the EM simulated coupler responses, and the characteristics predicted by the surrogate).

The comparison is executed at the level of complex responses, and averaged over all relevant characteristics ($|S_{11}|, |S_{21}|, |S_{31}|,$ and $|S_{41}|$) as well as all testing points. The performance of the proposed modeling approach, i.e., constrained FCRM surrogate, is compared to the following benchmark methods:

- Kriging interpolation [22];
- Radial basis functions (RBF) [63];
- Convolutional neural network (CNN) [54], [64];
- Ensemble learning [65];

The kriging surrogate uses second-order polynomial as a trend function, and Gaussian correlation function [22]. The RBF surrogate employs Gaussian basis functions. CNN [54], [64], with hyper-parameter setting of Depth size of 4 and first filter amount of 64 had been used another benchmark method. The final technique in our comparative study is the ensemble learning using LSBoost [65], with the learning rate of the model obtained using Bayesian Optimization.

C. RESULTS

Tables 2 through 5 provide the numerical results. These include the estimated modeling errors for the proposed and the benchmark techniques under two scenarios: the surrogate constructed in the conventional parameter space X , and within the constrained domain X_S . The conventional parameter space was considered to emphasize the advantages of the FCRM in modeling of nonlinear coupler characteristics even without domain confinement. The surrogate-predicted RRC and BLC characteristics versus their EM-simulated responses for selected testing locations have been shown in Figs. 6 and 7, respectively.

Based on the data included in the tables, one can formulate a number of observations concerning the performance of the presented modeling methodology, pertinent to both the original and the constrained domain. These are summarized below:

- The framework discussed in this paper ensures a modeling accuracy that by far surpasses the capability of all considered benchmark methods. This applies to both the original (unconstrained) space X , and the confined domain X_S .

TABLE 1. Verification case studies: rat-race coupler (RRC), and branch-line coupler (BLC).

	Coupler circuit	
	Rat-Race	Branch-Line
Substrate	RO4350 ($\epsilon_r = 3.38, h = 0.76$ mm)	$h = 0.76$ mm ϵ_r – one of objective space components
Independent parameters	$\mathbf{x} = [l_1 \ l_2 \ l_3 \ d \ w \ w_1]^T$	$\mathbf{x} = [g \ l_{1r} \ l_a \ l_b \ w_1 \ w_{2r} \ w_{3r} \ w_{4r} \ w_a \ w_b]^T$
Other parameters	$d_1 = d + w - w_1 , d = 1.0, w_0 = 1.7,$ and $l_0 = 15$	$L = 2dL + L_s, L_s = 4w_1 + 4g + s + l_a + l_b, W = 2dL + W_s, W_s = 4w_1 + 4g + s + 2w_a, l_1 = l_b l_{1r}, w_2 = w_a w_{2r}, w_3 = w_{3r} w_a$
Parameter space X (lower and upper bounds \mathbf{l} and \mathbf{u})	$\mathbf{l} = [2.0 \ 7.0 \ 12.5 \ 0.2 \ 0.7 \ 0.2]^T$ $\mathbf{u} = [4.5 \ 12.5 \ 22.0 \ 0.65 \ 1.5 \ 0.9]^T$	$\mathbf{l} = [0.4 \ 0.43 \ 5.9 \ 7.7 \ 0.68 \ 0.28 \ 0.1 \ 0.1 \ 2.0 \ 0.2]^T$ $\mathbf{u} = [1.0 \ 0.86 \ 14.0 \ 16.5 \ 1.5 \ 0.99 \ 0.65 \ 0.25 \ 5.5 \ 0.8]^T$
Objective space F	Operating frequency: $1.0 \text{ GHz} \leq f_0 \leq 2.0 \text{ GHz};$ Power split ratio $-6 \text{ dB} \leq K_P \leq 0 \text{ GHz}$ $\Rightarrow \mathbf{f} = [f_0 \ K_P]^T$	Operating frequency: $1.0 \text{ GHz} \leq f_0 \leq 2.0 \text{ GHz};$ Substrate permittivity $2.0 \leq \epsilon_r \leq 5.0$ $\Rightarrow \mathbf{f} = [f_0 \ \epsilon_r]^T$
Reference designs ^s	$\{f_0, K_P\} = \{1.0, -6\}, \{1.0, -2\}, \{1.0, 0\}, \{1.2, -4\}, \{1.3, 0\}, \{1.5, -5\}, \{1.5, -2\}, \{1.7, -6\}, \{1.7, 0\}, \{1.8, -3\}, \{2.0, -6\},$ and $\{2.0, 0\}$.	$\{f_0, \epsilon_r\} = \{1.0, 2.0\}, \{1.0, 3.5\}, \{1.0, 5.0\}, \{1.5, 2.0\}, \{1.5, 3.5\}, \{1.5, 5.0\}, \{2.0, 2.0\}, \{2.0, 3.5\},$ and $\{2.0, 5.0\}$.
Design optimality [#]	$U(\mathbf{x}, \mathbf{f}) = \max\{ S_{11}(\mathbf{x}, f_0) , S_{41}(\mathbf{x}, f_0) \} + \beta [K_P - [S_{21}(\mathbf{x}, f_0) - S_{31}(\mathbf{x}, f_0)]]^2$	$U(\mathbf{x}, \mathbf{f}) = \max\{ S_{11}(\mathbf{x}, f_0) , S_{41}(\mathbf{x}, f_0) \} + \beta [[S_{21}(\mathbf{x}, f_0) - S_{31}(\mathbf{x}, f_0)]]^2$

TABLE 2. Modeling results and benchmarking for the RRC of Fig. 5(a).

Number of training samples	Relative RMS error									
	Modeling in parameter space X					Modeling in confined domain X_S				
	Kriging	RBF	CNN	Ensemble	FCRM (this work)	Kriging	RBF	CNN	Ensemble	Constrained FCRM (this work)
50	25.7 %	28.3 %	22.9 %	32.7 %	17.7 %	6.9 %	7.8 %	11.4 %	15.8 %	5.8 %
100	17.9 %	19.1 %	12.7 %	28.1 %	12.8 %	5.7 %	6.3 %	5.7 %	11.5 %	4.3 %
200	13.5 %	13.9 %	8.0 %	25.0 %	5.1 %	3.8 %	4.9 %	3.7 %	12.3 %	2.9 %
400	9.9 %	10.3 %	5.5 %	22.8 %	4.1 %	3.5 %	4.6 %	3.3 %	8.7 %	2.2 %
800	8.0 %	8.9 %	4.5 %	19.1 %	3.6 %	3.1 %	4.1 %	3.1 %	8.3 %	1.9 %

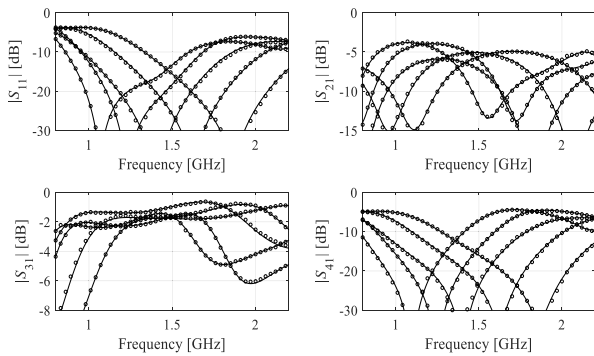


FIGURE 6. Miniaturized RRC of Fig. 5(a): S-parameters at the selected testing locations: EM simulations (—), and predictions by the proposed constrained FCRM surrogate (o). The surrogate set up using $N_B = 400$ training samples.

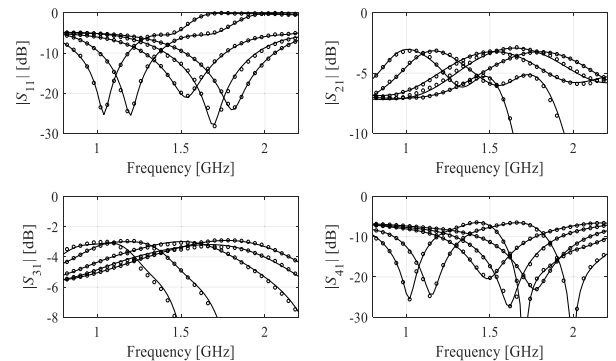


FIGURE 7. Miniaturized BLC of Fig. 5(b): S-parameters at the selected testing locations: EM simulations (—), and predictions by the proposed constrained FCRM surrogate (o). The surrogate set up using $N_B = 400$ training samples.

- The advantages of FCRM are particularly pronounced for larger data sets (200 samples and more); for example,

the accuracy of the FCRM surrogate in the original space is as good as that of kriging or RBF in the

TABLE 3. Constrained FCRM architectures for the RRC of Fig. 5(a).

Number of training samples	Domain	S-parameter	Φ	Θ	η_A	η_B
50	Original X	S_{11}	[4 1 1]	[256 8 32]	3	1
		S_{21}	[3 2 1]	[128 512 8]	3	1
		S_{31}	[3 4 4]	[1024 1024 128]	2	2
		S_{41}	[1 2 2]	[256 512 16]	3	1
	Confined X_S	S_{11}	[2 4 4]	[1024 64 64]	2	1
		S_{21}	[1 2 4]	[512 1024 256]	3	1
		S_{31}	[3 4 1]	[32 256 32]	2	1
		S_{41}	[1 2 3]	[256 64 32]	3	1
		S_{41}	[4 2 4]	[128 64 64]	2	3
100	Original X	S_{21}	[1 2 1]	[512 64 1024]	3	1
		S_{31}	[2 1 2]	[128 512 512]	3	1
		S_{41}	[3 1 1]	[256 256 128]	3	2
		S_{11}	[2 4 3]	[64 256 32]	2	1
	Confined X_S	S_{21}	[1 1 2]	[1024 16 128]	3	1
		S_{31}	[2 1 1]	[1024 32 128]	3	2
		S_{41}	[4 3 1]	[128 64 16]	3	1
		S_{11}	[3 2 2]	[128 512 128]	2	2
		S_{21}	[2 3 3]	[64 1024 8]	2	2
200	Original X	S_{31}	[2 2 2]	[256 256 128]	3	1
		S_{41}	[1 4 3]	[1024 1024 128]	3	2
		S_{11}	[2 3 2]	[32 128 8]	2	2
		S_{21}	[2 1 2]	[64 1024 64]	2	1
	Confined X_S	S_{31}	[2 2 1]	[32 512 8]	3	3
		S_{41}	[2 2 2]	[128 128 32]	2	1
		S_{11}	[3 2 2]	[1024 256 8]	3	2
		S_{21}	[4 2 1]	[256 8 128]	3	2
		S_{31}	[4 4 1]	[256 1024 64]	3	3
400	Original X	S_{41}	[3 2 1]	[256 512 64]	2	1
		S_{11}	[1 1 4]	[32 64 256]	2	2
		S_{21}	[1 1 2]	[128 1024 16]	2	2
		S_{31}	[3 1 2]	[256 512 16]	3	2
	Confined X_S	S_{41}	[1 3 3]	[64 64 256]	3	1
		S_{11}	[1 4 2]	[16 128 32]	2	2
		S_{21}	[2 4 2]	[128 256 128]	2	2
		S_{31}	[2 4 2]	[64 1024 8]	3	3
		S_{41}	[3 2 2]	[256 32 32]	3	2
800	Original X	S_{11}	[3 3 1]	[64 512 16]	3	3
		S_{21}	[2 2 4]	[256 64 128]	2	2
		S_{31}	[2 2 4]	[64 128 256]	2	2
		S_{41}	[3 1 3]	[128 256 32]	3	2
	Confined X_S	S_{11}	[1 4 2]	[16 128 32]	2	2
		S_{21}	[2 4 2]	[128 256 128]	2	2
		S_{31}	[2 4 2]	[64 1024 8]	3	3
		S_{41}	[3 2 2]	[256 32 32]	3	2
		S_{11}	[3 3 1]	[64 512 16]	3	3

TABLE 4. Modeling results and benchmarking for the BLC of Fig. 5(b).

Number of training samples	Relative RMS error									
	Modeling in parameter space X					Modeling in confined domain X_S				
	Kriging	RBF	CNN	Ensemble	FCRM (this work)	Kriging	RBF	CNN	Ensemble	Constrained FCRM (this work)
50	52.3 %	51.8 %	51.9 %	53.1 %	30.5 %	10.0 %	12.5 %	12.1 %	16.7 %	5.2 %
100	38.3 %	40.5 %	39.9 %	44.4 %	19.0 %	7.4 %	8.3 %	9.2 %	15.9 %	3.7 %
200	31.0 %	37.4 %	30.7 %	41.6 %	9.5 %	6.8 %	7.5 %	6.4 %	14.6 %	2.5 %
400	27.3 %	32.8 %	19.7 %	38.7 %	5.2 %	5.1 %	6.2 %	5.8 %	13.7 %	2.1 %
800	23.3 %	27.2 %	11.5 %	33.3 %	4.1 %	4.9 %	5.8 %	4.1 %	10.2 %	1.9 %

confined domain. This indicates a particular suitability of DNN-type of models for representing nonlinear characteristic of the coupling structures.

- FCRM is capable of exploring a particular arrangement of the training data which is evident from Tables 3 and 5. It can be observed that the model

TABLE 5. Constrained FCRM architectures for the BLC of Fig. 5(b).

Number of training samples	Domain	S-parameter	Φ	Θ	η_A	η_B
50	Original X	S_{11}	[4 1 1]	[256 32 128]	3	2
		S_{21}	[4 1 2]	[1024 1024 16]	2	2
		S_{31}	[2 4 3]	[1024 128 16]	3	1
		S_{41}	[2 1 2]	[1289 128 128]	3	1
	Confined X_S	S_{11}	[1 4 1]	[32 64 32]	3	1
		S_{21}	[1 2 4]	[64 1024 32]	3	1
		S_{31}	[2 2 2]	[256 256 64]	3	1
		S_{41}	[3 3 4]	[16 32 32]	2	1
100	Original X	S_{11}	[4 2 2]	[512 64 64]	3	1
		S_{21}	[4 2 2]	[1024 128 32]	3	1
		S_{31}	[3 3 1]	[512 128 1024]	3	2
		S_{41}	[4 4 1]	[1024 32 1024]	3	2
	Confined X_S	S_{11}	[3 1 3]	[64 16 64]	3	1
		S_{21}	[1 2 1]	[32 128 1024]	3	1
		S_{31}	[1 2 3]	[128 8 64]	3	1
		S_{41}	[4 3 1]	[256 32 8]	3	1
200	Original X	S_{11}	[4 3 1]	[64 1024 64]	3	3
		S_{21}	[2 2 3]	[1024 64 256]	3	2
		S_{31}	[2 2 4]	[1024 8 64]	3	1
		S_{41}	[3 4 1]	[512 512 8]	3	3
	Confined X_S	S_{11}	[1 1 2]	[128 256 32]	2	1
		S_{21}	[3 1 1]	[128 128 1024]	3	1
		S_{31}	[1 1 3]	[512 1024 512]	2	1
		S_{41}	[2 3 1]	[128 64 32]	3	1
400	Original X	S_{11}	[4 1 1]	[256 64 256]	3	2
		S_{21}	[1 1 3]	[512 32 32]	3	1
		S_{31}	[3 1 2]	[1024 128 16]	3	2
		S_{41}	[4 4 2]	[1024 256 64]	3	2
	Confined X_S	S_{11}	[1 2 3]	[512 128 32]	3	1
		S_{21}	[3 4 3]	[64 32 32]	3	1
		S_{31}	[2 1 1]	[256 256 1024]	3	1
		S_{41}	[3 1 2]	[256 1024 128]	3	1
800	Original X	S_{11}	[2 1 3]	[512 256 32]	3	2
		S_{21}	[3 1 1]	[256 128 128]	3	1
		S_{31}	[3 2 3]	[256 64 32]	3	1
		S_{41}	[3 4 2]	[512 128 8]	2	2
	Confined X_S	S_{11}	[3 4 1]	[64 512 128]	3	3
		S_{21}	[1 2 1]	[64 64 64]	3	1
		S_{31}	[1 3 2]	[256 1024 16]	3	2
		S_{41}	[2 3 1]	[256 128 128]	3	1

architecture, as adjusted through Bayesian optimization, depends on both the training data set size and the type of response to be modeled.

- FCRM takes a full advantage of domain confinement. It can be noted that the predictive power of constrained FCRM is better than nested kriging by a factor of two for the RRC and three for the BLC. In terms of specific figures, the average relative RMS error is as low as about two percent for 400 and 800 training samples. This is remarkable considering that both test cases are very challenging, especially in terms of the ranges of geometry and operating parameters covered by the region of validity of the surrogates.
- The results are consistent for both the RRC and BLC, which demonstrates that the considered method can

handle various modeling problems without any tuning of its control parameters.

IV. CONCLUSION

In this work, a novel approach to cost-efficient and reliable modeling of miniaturized microwave components has been proposed. The two major components of our technique are the fully-connected regression model, and the performance-driven concept. The former combines a deep neural network surrogate with automated architecture and hyper-parameter determination using Bayesian optimization. It allows us to take the advantage of the neural network flexibility in modeling of highly-nonlinear responses while avoiding the risk of overtraining, and to adjust the structure of the model to a specific input data set. The latter allows to overcoming

the issues related to parameter space dimensionality, and enables achieving good predictive power over broad ranges of geometry parameters and operating conditions of the device of interest. The numerical experiments based on two compact microstrip couplers, and conducted for the proposed method and a number of benchmark procedures (kriging, radial-basis functions, CNN, Ensemble Learning) conclusively demonstrate the advantages of the presented framework in terms of its ability to precisely model the circuit characteristics using small numbers of training data points, as well as capitalizing on the domain confinement concept. In particular, the accuracy improvement over nested kriging is by a factor of two and three for the first and the second test case, respectively. The modeling methodology discussed in this paper can be viewed as a viable alternative to existing procedures, especially under challenging scenarios that involve multi-dimensional spaces and wide ranges of both geometry and operating parameters that need to be covered by the surrogate.

ACKNOWLEDGMENT

The authors would like to thank Dassault Systemes, France, for making CST Microwave Studio available and Signal Processing for Computational Intelligence Group in Informatics Institute of Istanbul Technical University for providing computational resources.

REFERENCES

- [1] V. Palazzi, A. Cicioni, F. Alimenti, P. Mezzanotte, M. M. Tentzeris, and L. Roselli, "Compact 3-D-printed 4×4 butler matrix based on low-cost and curing-free additive manufacturing," *IEEE Microw. Wireless Compon. Lett.*, vol. 31, no. 2, pp. 125–128, Feb. 2021.
- [2] M. Fan, K. Song, Y. Zhu, and Y. Fan, "Compact bandpass-to-bandstop reconfigurable filter with wide tuning range," *IEEE Microw. Wireless Compon. Lett.*, vol. 29, no. 3, pp. 198–200, Mar. 2019.
- [3] C. Bao, X. Wang, Z. Ma, C.-P. Chen, and G. Lu, "An optimization algorithm in ultrawideband bandpass wilkinson power divider for controllable equal-ripple level," *IEEE Microw. Wireless Compon. Lett.*, vol. 30, no. 9, pp. 861–864, Sep. 2020.
- [4] Z. Zhang, H. Chen, Y. Yu, F. Jiang, and Q. S. Cheng, "Yield-constrained optimization design using polynomial chaos for microwave filters," *IEEE Access*, vol. 9, pp. 22408–22416, 2021.
- [5] F. Gunes, A. Uluclu, and P. Mahouti, "Pareto optimal characterization of a microwave transistor," *IEEE Access*, vol. 8, pp. 47900–47913, 2020.
- [6] J. W. Bandler, "Space mapping: The state of the art," *IEEE Trans. Microw. Theory Techn.*, vol. 52, no. 1, pp. 337–361, Jan. 2004.
- [7] J. E. Rayas-Sanchez, S. Koziel, and J. W. Bandler, "Advanced RF and microwave design optimization: A journey and a vision of future trends," *IEEE J. Microw.*, vol. 1, no. 1, pp. 481–493, 2021.
- [8] N. K. Nikolova, J. W. Bandler, and M. H. Bakr, "Adjoint techniques for sensitivity analysis in high-frequency structure CAD," *IEEE Trans. Microw. Theory Techn.*, vol. 52, no. 1, pp. 403–419, Jan. 2004.
- [9] S. Koziel, F. Mosler, S. Reitzinger, and P. Thoma, "Robust microwave design optimization using adjoint sensitivity and trust regions," *Int. J. RF Microw. Comput.-Aided Eng.*, vol. 22, no. 1, pp. 10–19, Jan. 2012.
- [10] S. Koziel and A. Pietrenko-Dabrowska, "Efficient gradient-based algorithm with numerical derivatives for expedited optimization of multi-parameter miniaturized impedance matching transformers," *Radioengineering*, vol. 27, no. 3, pp. 572–578, Sep. 2019.
- [11] A. Pietrenko-Dabrowska and S. Koziel, "Numerically efficient algorithm for compact microwave device optimization with flexible sensitivity updating scheme," *Int. J. RF Microw. Comput.-Aided Eng.*, vol. 29, no. 7, Jul. 2019, Art. no. e21714.
- [12] J. Wang, X. Yang, and B. Wang, "Efficient gradient-based optimisation of pixel antenna with large-scale connections," *IET Microw., Antennas Propag.*, vol. 12, no. 3, pp. 385–389, Feb. 2018.
- [13] F. Feng, J. Zhang, W. Zhang, Z. Zhao, J. Jin, and Q.-J. Zhang, "Coarse-and fine-mesh space mapping for EM optimization incorporating mesh deformation," *IEEE Microw. Wireless Compon. Lett.*, vol. 29, no. 8, pp. 510–512, Aug. 2019.
- [14] J. Ossorio, J. C. Melgarejo, V. E. Boria, M. Guglielmi, and J. W. Bandler, "On the alignment of low-fidelity and High-fidelity simulation spaces for the design of microwave waveguide filters," *IEEE Trans. Microw. Theory Techn.*, vol. 66, no. 12, pp. 5183–5196, Dec. 2018.
- [15] S. Koziel, "Shape-preserving response prediction for microwave design optimization," *IEEE Trans. Microw. Theory Techn.*, vol. 58, no. 11, pp. 2829–2837, Nov. 2010.
- [16] J. E. Rayas-Sanchez, "Power in simplicity with ASM: Tracing the aggressive space mapping algorithm over two decades of development and engineering applications," *IEEE Microw. Mag.*, vol. 17, no. 4, pp. 64–76, Apr. 2016.
- [17] S. Koziel and S. D. Unnsteinsson, "Expedited design closure of antennas by means of trust-region-based adaptive response scaling," *IEEE Antennas Wireless Propag. Lett.*, vol. 17, no. 6, pp. 1099–1103, Jun. 2018.
- [18] F. Wang, P. Cachecho, W. Zhang, S. Sun, X. Li, R. Kanj, and C. Gu, "Bayesian model fusion: Large-scale performance modeling of analog and mixed-signal circuits by reusing early-stage data," *IEEE Trans. Comput.-Aided Design Integr. Circuits Syst.*, vol. 35, no. 8, pp. 1255–1268, Aug. 2016.
- [19] M. Kennedy, "Predicting the output from a complex computer code when fast approximations are available," *Biometrika*, vol. 87, no. 1, pp. 1–13, Mar. 2000.
- [20] J. C. Melgarejo, J. Ossorio, S. Cogollos, M. Guglielmi, V. E. Boria, and J. W. Bandler, "On space mapping techniques for microwave filter tuning," *IEEE Trans. Microw. Theory Techn.*, vol. 67, no. 12, pp. 4860–4870, Dec. 2019.
- [21] J. L. Chavez-Hurtado and J. E. Rayas-Sanchez, "Polynomial-based surrogate modeling of RF and microwave circuits in frequency domain exploiting the multinomial theorem," *IEEE Trans. Microw. Theory Techn.*, vol. 64, no. 12, pp. 4371–4381, Dec. 2016.
- [22] S. Koziel and A. Pietrenko-Dabrowska, "Rapid optimization of compact microwave passives using kriging surrogates and iterative correction," *IEEE Access*, vol. 8, pp. 53587–53594, 2020.
- [23] P. Barmuta, F. Ferranti, G. P. Gibiino, A. Lewandowski, and D. M. M.-P. Schreurs, "Compact behavioral models of nonlinear active devices using response surface methodology," *IEEE Trans. Microw. Theory Techn.*, vol. 63, no. 1, pp. 56–64, Jan. 2015.
- [24] Z. Zhang, Q. S. Cheng, H. Chen, and F. Jiang, "An efficient hybrid sampling method for neural network-based microwave component modeling and optimization," *IEEE Microw. Wireless Compon. Lett.*, vol. 30, no. 7, pp. 625–628, Jul. 2020.
- [25] A. Petrocchi, A. Kaintura, G. Avolio, D. Spina, T. Dhaene, A. Raffo, and D. M. M.-P. Schreurs, "Measurement uncertainty propagation in transistor model parameters via polynomial chaos expansion," *IEEE Microw. Wireless Compon. Lett.*, vol. 27, no. 6, pp. 572–574, Jun. 2017.
- [26] H. M. Torun and M. Swaminathan, "High-dimensional global optimization method for high-frequency electronic design," *IEEE Trans. Microw. Theory Techn.*, vol. 67, no. 6, pp. 2128–2142, Jun. 2019.
- [27] A. M. Alzahed, S. M. Mikki, and Y. M. M. Antar, "Nonlinear mutual coupling compensation operator design using a novel electromagnetic machine learning paradigm," *IEEE Antennas Wireless Propag. Lett.*, vol. 18, no. 5, pp. 861–865, May 2019.
- [28] D. Gorissen, K. Crombecq, I. Couckuyt, T. Dhaene, and P. Demeester, "A surrogate modeling and adaptive sampling toolbox for computer based design," *J. Mach. Learn. Res.*, vol. 11, pp. 2051–2055, Jul. 2010.
- [29] N. V. Queipo, R. T. Haftka, W. Shyy, T. Goel, R. Vaidynathan, and P. K. Tucker, "Surrogate based analysis and optimization," *Prog. Aerosp. Sci.*, vol. 41, no. 1, pp. 1–28, 2005.
- [30] S. Xiao, G. Q. Liu, K. L. Zhang, Y. Z. Jing, J. H. Duan, P. Di Barba, and J. K. Sykulski, "Multi-objective Pareto optimization of electromagnetic devices exploiting kriging with lipschitzian optimized expected improvement," *IEEE Trans. Magn.*, vol. 54, no. 3, Mar. 2018, Art. no. 7001704.
- [31] J. Zhang, F. Feng, and Q. J. Zhang, "Rapid yield estimation of microwave passive components using model-order reduction based neuro-transfer function models," *IEEE Microw. Wireless Comp. Lett.*, vol. 31, no. 4, pp. 333–336, Apr. 2021.

- [32] S. Koziel, A. Pietrenko-Dabrowska, and J. W. Bandler, "Computationally efficient performance-driven surrogate modeling of microwave components using principal component analysis," in *IEEE MTT-S Int. Microw. Symp. Dig.*, Los Angeles, CA, USA, Jul. 2020, pp. 68–71.
- [33] A. K. Prasad and S. Roy, "Accurate reduced dimensional polynomial chaos for efficient uncertainty quantification of microwave/RF networks," *IEEE Trans. Microw. Theory Techn.*, vol. 65, no. 10, pp. 3697–3708, Oct. 2017.
- [34] S. Koziel and L. Leifsson, "Surrogate-based aerodynamic shape optimization by variable-resolution models," *AIAA J.*, vol. 51, no. 1, pp. 94–106, Jan. 2013.
- [35] L. Leifsson and S. Slawomir, "Variable-resolution shape optimisation: Low-fidelity model selection and scalability," *Intern. J. Math. Mod. Num. Opt.*, vol. 6, no. 1, pp. 1–21, 2015.
- [36] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, pp. 436–444, May 2015.
- [37] J. Jin, F. Feng, J. Zhang, S. Yan, W. Na, and Q. Zhang, "A novel deep neural network topology for parametric modeling of passive microwave components," *IEEE Access*, vol. 8, pp. 82273–82285, 2020.
- [38] J. Jin, C. Zhang, F. Feng, W. Na, J. Ma, and Q.-J. Zhang, "Deep neural network technique for high-dimensional microwave modeling and applications to parameter extraction of passive microwave filters," *IEEE Trans. Microw. Theory Techn.*, vol. 67, no. 10, pp. 4140–4155, Oct. 2019.
- [39] S. Salman and X. Liu, "Overfitting mechanism and avoidance in deep neural networks," 2019, *arXiv:1901.06566*. [Online]. Available: <http://arxiv.org/abs/1901.06566>
- [40] X. Ying, "An overview of overfitting and its solutions," *J. Phys. Conf.*, vol. 1168, no. 2, 2019, Art. no. 022022.
- [41] L. Kouhalvandi, O. Ceylan, and S. Ozoguz, "Automated deep neural learning-based optimization for high performance high power amplifier designs," *IEEE Trans. Circuits Syst. I, Reg. Papers*, vol. 67, no. 12, pp. 4420–4433, Dec. 2020.
- [42] N. Calik, M. A. Belen, P. Mahouti, and S. Koziel, "Accurate modeling of frequency selective surfaces using fully-connected regression model with automated architecture determination and parameter selection based on Bayesian optimization," *IEEE Access*, vol. 9, pp. 38396–38410, 2021.
- [43] J. Mockus, "Application of Bayesian approach to numerical methods of global and stochastic optimization," *J. Global Optim.*, vol. 4, no. 4, pp. 347–365, Jun. 1994.
- [44] A. Kumar and J. Mayank, *Ensemble Learning for AI Developers*. New Delhi, India: Springer, 2020.
- [45] X. Dong, Z. Yu, W. Cao, Y. Shi, and Q. Ma, "A survey on ensemble learning," *Frontiers Comput. Sci.*, vol. 14, no. 2, pp. 241–258, 2020.
- [46] K. M. R. Alam, N. Siddique, and H. Adeli, "A dynamic ensemble learning algorithm for neural networks," *Neural Comput. Appl.*, vol. 32, no. 12, pp. 8675–8690, Jun. 2020.
- [47] S. Koziel and A. Pietrenko-Dabrowska, *Performance-Driven Surrogate Modeling of High-Frequency Structures*. New York, NY, USA: Springer, 2020.
- [48] S. Koziel and A. Pietrenko-Dabrowska, "Performance-based nested surrogate modeling of antenna input characteristics," *IEEE Trans. Antennas Propag.*, vol. 67, no. 5, pp. 2904–2912, May 2019.
- [49] S. Koziel and A. Pietrenko-Dabrowska, "Recent advances in high-frequency modeling by means of domain confinement and nested kriging," *IEEE Access*, vol. 8, pp. 189326–189342, 2020.
- [50] V. Sze, Y.-H. Chen, T.-J. Yang, and J. S. Emer, "Efficient processing of deep neural networks: A tutorial and survey," *Proc. IEEE*, vol. 105, no. 12, pp. 2295–2329, Dec. 2017.
- [51] T. Jiang, X. Yang, Y. Shi, and H. Wang, "Layer-wise deep neural network pruning via iteratively reweighted optimization," in *Proc. ICASSP*, Brighton, U.K., 2019, pp. 5606–5610.
- [52] B. Karlik and A. Vehbi, "Performance analysis of various activation functions in generalized MLP architectures of neural networks," *Int. J. Artif. Intell. Expert Syst.*, vol. 1, no. 4, pp. 111–122, 2015.
- [53] Z. Hu, J. Zhang, and Y. Ge, "Handling vanishing gradient problem using artificial derivative," *IEEE Access*, vol. 9, pp. 22371–22377, 2021.
- [54] H. Ide and T. Kurita, "Improvement of learning for CNN with ReLU activation by sparse regularization," in *Proc. IJCNN*, Anchorage, AK, USA, 2017, pp. 2684–2691.
- [55] R. Parhi and R. D. Nowak, "The role of neural network activation functions," *IEEE Signal Process. Lett.*, vol. 27, pp. 1779–1783, 2020.
- [56] Z. Chen, L. Deng, G. Li, J. Sun, X. Hu, L. Liang, Y. Ding, and Y. Xie, "Effective and efficient batch normalization using a few uncorrelated data for statistics estimation," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 32, no. 1, pp. 348–362, Jan. 2021.
- [57] S. Ioffe and C. Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift," in *Proc. Int. Conf. Mach. Learn.*, Lille, France, vol. 37, Jul. 2015, pp. 448–456.
- [58] S. Greenhill, S. Rana, S. Gupta, P. Vellanki, and S. Venkatesh, "Bayesian optimization for adaptive experimental design: A review," *IEEE Access*, vol. 8, pp. 13937–13948, 2020.
- [59] B. Shahriari, K. Swersky, Z. Wang, R. P. Adams, and N. de Freitas, "Taking the human out of the loop: A review of Bayesian optimization," *Proc. IEEE*, vol. 104, no. 1, pp. 148–175, Jan. 2016.
- [60] L. Yang and A. Shami, "On hyperparameter optimization of machine learning algorithms: Theory and practice," *Neurocomputing*, vol. 415, pp. 295–316, Nov. 2020.
- [61] S. Koziel and A. T. Sigurðsson, "Performance-driven modeling of compact couplers in restricted domains," *Int. J. RF Microw. Comput.-Aided Eng.*, vol. 28, no. 6, Aug. 2018, Art. no. e21296.
- [62] C.-H. Tseng and C.-L. Chang, "A rigorous design methodology for compact planar branch-line and rat-race couplers with asymmetrical T-structures," *IEEE Trans. Microw. Theory Techn.*, vol. 60, no. 7, pp. 2085–2092, Jul. 2012.
- [63] S.-F. Su, C.-C. Chuang, C. W. Tao, J.-T. Jeng, and C.-C. Hsiao, "Radial basis function networks with linear interval regression weights for symbolic interval data," *IEEE Trans. Syst., Man, Cybern. B, Cybern.*, vol. 42, no. 1, pp. 69–80, Feb. 2012.
- [64] A. Ameri, M. A. Akhaee, E. Scheme, and K. Englehart, "Regression convolutional neural network for improved simultaneous EMG control," *J. Neural Eng.*, vol. 16, no. 3, Jun. 2019, Art. no. 036015.
- [65] Y. Zhang and X. Xu, "Solubility predictions through LSBoost for supercritical carbon dioxide in ionic liquids," *New J. Chem.*, vol. 44, no. 47, pp. 20544–20567, 2020.



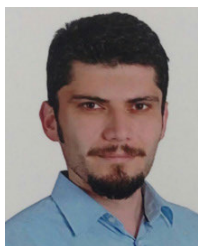
SLAWOMIR KOZIEL (Senior Member, IEEE) received the M.Sc. and Ph.D. degrees in electronic engineering from the Gdansk University of Technology, Poland, in 1995 and 2000, respectively, and the M.Sc. degree in theoretical physics and in mathematics, and the Ph.D. degree in mathematics from the University of Gdansk, Poland, in 2000, 2002, and 2003, respectively. He is currently a Professor with the Department of Engineering, Reykjavik University, Iceland. His research interests include CAD and modeling of microwave and antenna structures, simulation-driven design, surrogate-based optimization, space mapping, circuit theory, analog signal processing, evolutionary computation, and numerical analysis.



PEYMAN MAHOUTI received the M.Sc. and Ph.D. degrees in electronics and communication engineering from Yildiz Technical University, Turkey, in 2013 and 2016, respectively. He is currently an Associate Professor with the Department of Electronic and Communication, Istanbul University–Cerrahpasa, Turkey. The main research areas are analytical and numerical modeling of microwave devices, optimization techniques for microwave stages, and application of artificial intelligence-based algorithms. His research interests include analytical and numerical modeling of microwave and antenna structures, surrogate-based optimization, and application of artificial intelligence algorithms.



NURULLAH CALIK received the M.Sc. and Ph.D. degrees in electronics and communication engineering from Yildiz Technical University, Turkey, in 2013 and 2019, respectively. He was worked as a Postdoctoral Researcher with the Informatics Institute, Istanbul Technical University. He is currently an Assistant Professor with the Department of Biomedical Engineering, Istanbul Medeniyet University, Turkey. The main research areas are large-scale data analysis, signal and image processing, and AI applications in engineering. His research interests include especially deep learning regression, optimization, and surrogate modeling.



MEHMET ALI BELEN received the Ph.D. degree in electronics and communication engineering from Yildiz Technical University, in 2016. He is currently an Associate Professor with Iskenderun Technical University. His current activities include teaching and researching electromagnetics and microwaves along with developing additive manufacturing 3D printed microwave components for rapid prototyping. His current research interests include the areas of multivariable network theory, device modeling, computer aided microwave circuit design, monolithic microwave integrated circuits, and antenna arrays, and active/passive microwave components, especially in the field of metamaterial-based antennas and microwave filters.



STANISLAW SZCZEPANSKI received the M.Sc. and Ph.D. degrees in electronic engineering from the Gdańsk University of Technology, Poland, in 1975 and 1986, respectively. In 1986, he was a Visiting Research Associate with the Institute National Polytechnique de Toulouse (INPT), Toulouse, France. From 1990 to 1991, he was with the Department of Electrical Engineering, Portland State University, Portland, OR, USA, on a Kosciuszko Foundation Fellowship. From August 1998 to September 1998, he was a Visiting Professor with the Faculty of Engineering and Information Sciences, the University of Hertfordshire, Hatfield, U.K. He is currently a Professor with the Department of Microelectronic Systems, Faculty of Electronics, Telecommunications, and Informatics, Gdańsk University of Technology. He has published more than 160 papers and holds three patents. His teaching and research interests include circuit theory, fully integrated analog filters, high-frequency transconductance amplifiers, analog integrated circuit design, and analog signal processing.

...