scientific reports

Check for updates

OPEN

Global miniaturization of broadband antennas by prescreening and machine learning

Slawomir Koziel[1,2]✉, Anna Pietrenko-Dabrowska[2] & Ubaid Ullah[3]

The development of contemporary electronic components, particularly antennas, places significant emphasis on miniaturization. This trend is driven by the emergence of technologies such as mobile communications, the internet of things, radio-frequency identification, and implantable devices. The need for small size is accompanied by heightened demands on electrical and field properties, posing a considerable challenge for antenna design. Shrinking physical dimensions can compromise performance, making miniaturization-oriented parametric optimization a complex and heavily constrained task. Additionally, the task is multimodal due to typical parameter redundancy resulting from various topological modifications in compact antennas. Identifying truly minimum-size designs requires a global search approach, as the popular nature-inspired algorithms face challenges related to computational efficiency and the need for reliable full-wave electromagnetic (EM) simulation to evaluate device's characteristics. This study introduces an innovative machine learning procedure for cost-effective global optimization-based miniaturization of antennas. Our technique includes parameter space pre-screening and the iterative refinement of kriging surrogate models using the predicted merit function minimization as an infill criterion. Concurrently, the design task incorporates design constraints implicitly by means of penalty functions. The combination of these mechanisms demonstrates superiority over conventional techniques, including gradient search and electromagnetic-driven nature-inspired optimization. Numerical experiments conducted on four broadband antennas indicate that the proposed framework consistently yields competitive miniaturization rates across multiple algorithm runs at low costs, compared to the benchmark.

Keywords Antenna design, Miniaturization, Simulation-based design, Response features, Machine learning, Surrogate modeling

Present-day electronics, particularly in high-frequency design, heavily emphasizes miniaturization, a crucial aspect for an expanding array of application domains. Examples include mobile communications[1], the internet of things[2], medical imaging[3], wearable devices[4], and radio-frequency identification[5]. Developing compact antennas proves challenging because reducing size tends to compromise electrical parameters[6,7], while simultaneously facing (often stringent) performance requirements related to antenna characteristics such as impedance matching[8], gain[9], and sidelobe levels[10]. Moreover, modern antennas are expected to fulfil diverse functionalities such as multi-band operation[11], multiple-input-multiple-output (MIMO) operation[12], circular polarization[13], and polarization diversity[14]. Practical designs of miniaturized antennas necessitate working out size and performance trade-offs.

The initial phase in the development of compact antennas involves selecting their geometry, often based on parametric studies[15] and what is known as topology evolution[16]. This process frequently incorporates elements such as stubs[17], slots[18], defected ground structures (DGS)[19], stepped-impedance feeds[20], shorting pins[21], line meandering[22], and others. While these modifications contribute to size reduction, they also significantly intensify the design process. On one hand, compact antennas are parameterized using relatively large numbers of variables. On the other hand, their precise evaluation necessitates full-wave electromagnetic (EM) analysis. Given the former, determining the optimal antenna dimensions requires formal optimization, preferably employing global search procedures. The latter aspect makes the task computationally intensive, as any optimization involves extensive EM simulations of the structure at hand[23,24]. Additionally, explicit miniaturization is a constrained

[1]Engineering Optimization and Modeling Center, Reykjavik University, 102 Reykjavik, Iceland. [2]Faculty of Electronics, Telecommunications and Informatics, Gdansk University of Technology, 80-233 Gdansk, Poland. [3]Networks and Communication Engineering Department, Al Ain University, 112612 Abu Dhabi, United Arab Emirates. ✉email: koziel@ru.is

problem with costly conditions (e.g., acceptance thresholds for antenna reflection[25], gain[26], or axial ratio[27]), further complicating the issue[28]. Lastly, topological complexity results in counterintuitive relationships between certain antenna dimensions and electrical characteristics, potentially leading to misleading impressions that specific modifications are beneficial for size reduction when proper optimization may reveal their actual irrelevance[29]. It should also be mentioned that some recent works reported methods for unsupervised antenna design, which include automated generation of the antenna topology, either free-form topology optimization[30–32], or through adjustment of pre-defined structures (e.g., pixel antennas, etc.)[33–36]. In these cases, the development of antenna geometry is carried out simultaneously with parametric optimization.

Rigorous optimization methods have been gradually receiving attention in the antenna community, and are not only applied in solving a variety of design tasks (matching improvement[37], pattern synthesis[38], isolation enhancement in MIMO systems[39]), both in single-[40], and multi-criterial regime[41], but also used for global optimization[42–44], and uncertainty quantification[45–47]. An excellent review of the various optimization methods for antenna design, and exposition of real-world design scenarios can be found in[48]. As mentioned earlier, EM-driven size reduction is a particularly challenging task, not only because of being constrained, but also due to its multimodality. The first problem has been addressed through the development of both implicit[49–51], and explicit constraint handling approaches[42], all coupled with local (gradient-based) search routines[53]. Global optimization is nowadays mainly performed using nature-inspired population-based algorithms, a plethora of which have been developed over the years, e.g., genetic and evolutionary algorithms, differential evolution, particle swarm optimization (PSO), harmony search, ant systems, grey wolf optimization, firefly algorithm[54–61], and many others[62–66]. The global search capability of these methods is arguably linked to their intrinsic randomness, manifested in various forms such as stochastic selection procedures[67] and randomized relocation rules[68]. Additionally, these methods involve exchanging information within the ensemble of potential solutions handled by the procedure, employing operations like recombination[69] or biasing relocation towards the best solution found thus far[70]. However, when viewed from the standpoint of EM-driven design, nature-inspired methods suffer from a fundamental drawback – their considerable computational complexity. Conducting direct optimization at the EM simulation level is simply impractical given the typical number of merit function calls, which varies from hundreds to many thousands. An exception is when the individual EM simulation time is short (e.g., when using 2.5D solvers such as Sonnet ***em***) or the antenna structure at hand is simple and simulated without extra components (e.g., connectors). In such cases, direct EM-driven global optimization is still feasible.

The challenges associated with the excessive expenses of global search can be mitigated through the application of surrogate modelling techniques[71–74]. Surrogate-assisted nature-inspired methods, often referred to as machine learning procedures[75,76], typically operate in an iterative manner, refining the initial metamodel using electromagnetic (EM) data accumulated in the course of the optimization run[77]. The generation of infill points aims to enhance the model's predictive power (design space exploration[78]), identify the optimum (exploitation[79]), or achieve a balance between both objectives[80]. The infill points are the designs produced during the optimization process, e.g., by optimizing the underlying surrogate model, which are used to enhance the available dataset (for the purpose of improving the surrogate's reliability) and facilitate identification of the optimum. Various modelling methods applied for these purposes include kriging[81], Gaussian Process Regression (GPR)[82], and neural networks[83]. In the case of antennas, a limiting factor is high nonlinearity of frequency responses, rendering the modelling process challenging and necessitating significant investments in terms of training data acquisition[84]. Consequently, many of the reported methods are only demonstrated using low-dimensional examples[85–87]. Nonetheless, the recent literature provides a growing number of examples of utilizing machine learning schemes for antenna optimization, using diverse underlying surrogate modelling methods such as GPR or various types of artificial neural networks[88–92]. Possible ways of alleviating this difficulty are performance-driven modelling[93,94] (although their incorporation into global search frameworks is not trivial), variable-resolution modelling[95], but also a response feature technology[96]. The latter has been demonstrated successful in accelerating optimization processes[97], enabling quasi-global search capabilities[98], as well as reducing computational expenses of surrogate model construction[99].

This research introduces an innovative technique for globally reducing the size of broadband antennas through electromagnetic (EM) optimization. The proposed methodology is a machine learning procedure that leverages initial pre-screening of the parameter space. Furthermore, it employs implicit handling of design constraints, transforming the size reduction problems into an unconstrained task. The underlying surrogate modelling approach is kriging interpolation, whereas the infill criterion is based on minimizing the predicted merit function. The candidate designs (infill points) are rendered by globally optimizing the metamodel, using a particle swarm optimizer (PSO)[100]. The pre-screening strategy's utilization enables finding the search space region that contains the feasible region boundary, which reduces the size of the dataset needed to build a dependable metamodel. The proposed strategy has been verified using several microstrip antennas and compared to various benchmark techniques, including gradient search, nature-inspired algorithms, and machine learning procedures directly working with complete antenna responses (without pre-screening). The results obtained demonstrate competitive miniaturization rates achievable with the proposed algorithm, consistency across a set of independent runs, and low computational cost, averaging about two hundred EM antenna simulations.

The key original components of this article can be succinctly outlined as follows: (i) the introduction of an innovative machine learning procedure explicitly designed for reducing the size of broadband antennas, (ii) the integration of knowledge-based parameter space pre-screening, (iii) the implementation of implicit constraint handling to facilitate global search through nature-inspired algorithms, and (iv) the demonstration of practical advantages associated with the proposed method, including enhanced design reliability and computational efficiency. To the authors' best knowledge, no framework for size reduction has been reported to date that is comparable to the one presented in this work. This assessment encompasses not only the methodological components of the algorithm but also its overall effectiveness.

EM-driven size reduction. Explicit and implicit constraints

This section formulates optimization-based antenna size reduction task as a constrained nonlinear minimization problem. We also discuss explicit and implicit constraint handling, as well as provide a few specific examples. Finally, we briefly outline available solution approaches. The global optimization search suggested in this study is elucidated in Section "Global size reduction using machine learning and parameter space pre-screening".

Antenna miniaturization through EM-driven optimization

As previously highlighted, achieving a compact size is imperative for an expanding array of practical applications, encompassing mobile communications, wearable or implantable devices, and the internet of things. Electrically small antennas are typically crafted by introducing various topological modifications to basic structures (e.g., patches, monopoles), such as stubs[17], metamaterial components[101], or defected ground structures[19]. However, attaining the smallest possible size necessitates meticulous tuning of all antenna parameters. Traditional parametric studies (typically, based on sweeping one parameter at a time) are grossly unable to produce optimum designs, especially due to the necessity of handling multiple antenna characteristics. Utilization of formal numerical optimization is recommended instead.

Table 1 encompasses the nomenclature utilized in the article. The problem at hand can be defined as

$$\mathbf{x}^* = \arg\min_{\mathbf{x}\in X_f} \tag{1}$$

The feasible space X_f in (1) (cf. Table 1) is the region containing all points that satisfy the conditions imposed upon the antenna.

Design constraints. Explicit and implicit constraint handling

At this juncture, it is crucial to acknowledge that aside from strictly geometrical conditions (e.g., those associated with constraints imposed upon the ranges of antenna parameters), the constraints associated with electrical and field properties, making them costly to evaluate, necessitating electromagnetic (EM) analysis. Managing expensive constraints is typically challenging. A potential solution involves implicit constraint handling, wherein the original problem (1) is reformulated as follows:

$$\mathbf{x}^* = \arg\min_{\mathbf{x}} U_p(\mathbf{x}) \tag{2}$$

with the objective function U_p assuming the form of

$$U_p(\mathbf{x}) = U(\mathbf{x}) + \sum_{k=1}^{n_g+n_h} \beta_k c_k(\mathbf{x}) \tag{3}$$

The functions $c_k(\boldsymbol{x})$, acting as penalties, quantify the constraint violations, with β_k representing the penalty factors. It is important to note that although all constraints are treated uniformly here, most constraints in antenna design belong to the first category (inequality type). Some relevant examples are detailed in Table 2. The function $c(\boldsymbol{x})$ provided in the table gauges the relative constraint violation, with $c(\boldsymbol{x}) \neq 0$ only when the constraint is actually breached. Additionally, the use of the second power $[.]^2$ is to make sure that the penalty function is smooth at the feasible region boundary. The latter aids in the search across the feasible region, which is crucial since at least one of the constraints is active at the minimum-size design.

The implicit approach formally poses the miniaturization problem as an unconstrained task, yet the performance of the search process depends on the selection of the parameters β_k. Excessively low values may

Symbol	Meaning	Comment
$x = [x_1 \ldots x_n]^T$	Vector of antenna design parameters	Typically, the variables are antenna geometry parameters (dimensions in mm)
X	Parameter space	Typically, X is an interval $[\boldsymbol{l}\,\boldsymbol{u}]$, where $\boldsymbol{l} = [l_1 \ldots l_n]^T$ and $\boldsymbol{u} = [u_1 \ldots u_n]^T$ are lower and upper bounds on design parameters so that $l_k \le x_k \le u_k$, for $k = 1, \ldots, n$
$A(\boldsymbol{x})$	Antenna size	For planar antennas, $A(\boldsymbol{x})$ is typically a footprint area in mm^2
$\boldsymbol{R}_{EM}(\boldsymbol{x})$	Antenna responses at design $\boldsymbol{x}$	Aggregated antenna responses obtained through EM analysis
$S_{11}(\boldsymbol{x},f)$	Antenna reflection coefficient at the design $\boldsymbol{x}$ and frequency f	Reflection coefficient is a complex number; in the design process we handle its modulus $\lvert S_{11}\rvert$, expressed in decibels
$G(\boldsymbol{x},f)$	Antenna gain at the design $\boldsymbol{x}$ and frequency f	For example, G may stand for a realized gain in a broadside direction, expressed in decibels
$A_R(\boldsymbol{x},f)$	Antenna axial ratio at the design $\boldsymbol{x}$ and frequency f	Axial ratio in a specified (e.g., broadside) direction, expressed in decibels
$g_k(\boldsymbol{x}) \le 0$, $k = 1, \ldots, n_g$,	Inequality constraints	Typically, expressed using acceptance thresholds for selected antenna characteristics, e.g., $\lvert S_{11}(\boldsymbol{x},f)\rvert \le -10$ dB for f within the frequency range of interest
$h_k(\boldsymbol{x}) = 0$, $k = 1, \ldots, n_h$,	Equality constraints	Typically, expressed using target values for selected antenna characteristics
X_f	Feasible space	$X_f \subset X$ contains parameter vectors $\boldsymbol{x}$ for which all constraints are satisfied, i.e., $g_k(\boldsymbol{x}) \le 0$ for $k = 1, \ldots, n_g$, and $h_k(\boldsymbol{x}) = 0$ for $k = 1, \ldots, n_h$

Table 1. EM-driven antenna size reduction. Notation and terminology.

Constraint description	Penalty function
Given and operating bandwidth F, ensure that the antenna reflection $\|S_{11}(\boldsymbol{x},f)\|$ does not exceed −10 dB within F, i.e. $\|S_{11}(\boldsymbol{x},f)\| \leq -10$ dB for $f \in F$	$c(\mathbf{x}) = \left[\frac{\max\{S(\mathbf{x})+10,0\}}{10}\right]^2$ where $S(\boldsymbol{x}) = \max\{f\widehat{IF} : \|S_{11}(\boldsymbol{x},f)\|\}$
Ensure that the axial ratio $AR(\boldsymbol{x},f)$ of a CP antenna does not exceed 3 dB within the operating range F, i.e., $AR(\boldsymbol{x},f) \leq 3$ dB for $f \in F$	$c(\mathbf{x}) = \left[\frac{\max\{AR(\mathbf{x})-3,0\}}{3}\right]^2$ where $A_R(\boldsymbol{x}) = \max\{f \in F : AR(\boldsymbol{x},f)\}$
Ensure that variability of realized gain $G(\boldsymbol{x},f)$ is below 2 dB within the antenna operating range F, i.e., $\Delta G(\boldsymbol{x},f) \leq 2$ dB for $f \in F$, where $\Delta G(\boldsymbol{x},f) = \max\{f \in F : G(\boldsymbol{x},f)\} - \min\{f \in F : G(\boldsymbol{x},f)\}$ (note that $\Delta G(\boldsymbol{x},f)$ is the difference of the maximum gain over F, i.e., $\max\{f \in F : G(\boldsymbol{x},f)\}$ and the minimum gain over F, i.e., $\min\{f \in F : G(\boldsymbol{x},f)\}$)	$c(\mathbf{x}) = \left[\frac{\max\{G(\mathbf{x})-2,0\}}{2}\right]^2$ where $G(\boldsymbol{x}) = \max\{f \in F : \Delta G(\boldsymbol{x},f)\}$

Table 2. Examples of design constraints and penalty functions.

lead to severe violation of constraints. Values that are too high make the merit function extremely steep near the feasible region boundary, which makes it difficult to be explored. Appropriate selection of β_k is a non-trivial problem. Recently, some adaptation strategies have been proposed (e.g.,[25,27]). Also, the algorithms with explicit constraint handling have been introduced[52]. Both offer acceptable trade-offs between achievable miniaturization ratios and constraint control. Nevertheless, the mentioned techniques are all local methods, which does not address inherent multimodality of the size reduction task.

Global size reduction using machine learning and parameter space pre-screening

This section elucidates the details of the optimization-based antenna size reduction procedure presented in the work. We start by discussing the pre-screening procedure (Section "Parameter space pre-screening"), which aims at initial identification of the promising search space region based on possible prior knowledge about the antenna structure at hand, as well as general properties of antenna characteristics at the minimum-size design. Section "Surrogate model construction" briefly recalls kriging interpolation, which is employed in this work as the primary surrogate modelling method. At this point, it should be mentioned that essentially any other data-driven approach could be used in place of kriging (e.g., neural networks, polynomial chaos expansion, etc.) as the specific modelling approach is of secondary importance from the point of view of the ML process. Generation of the infill points is explained in Section "Generating Infill Points". The complete optimization procedure is summarized in Section "Complete optimization procedure". At this point it should be indicated that the major novelty of the proposed approach is in appropriate selection and combination of the ML and pre-screening concepts along with a suitable formulation of the design task (here, using implicit constraint handling). Clearly, none of these components by itself is new, yet specific implementation and their incorporation into the complete framework is original and efficient, as corroborated through extensive demonstration case studies provided in Section "Demonstration examples".

Parameter space pre-screening

The fundamental difficulty of global optimization of antenna systems is related to CPU-intensive evaluation of antenna responses, here, executed using EM simulation. This factor alone makes any sort of direct global search, e.g., using nature-inspired methods[62–66], prohibitive in most cases. At the same time, the employment of surrogate modelling methodologies is challenging due to large size of the search space, not only with respect to its dimensionality but also parameter ranges, although some recent works successfully aimed at addressing this issue, e.g.,[102–104].

Highly nonlinear characteristics of antennas only aggravate the problem. In the case of size reduction, an additional issue is the presence of expensive constraints (e.g., acceptance thresholds for reflection, gain, axial ratio, etc., cf. Section "EM-driven size reduction. Explicit and implicit constraints").

The aforementioned problems may be mitigated by appropriate pre-screening of the parameters space, which should be based on the known general properties of the system under design, but also any supplementary information that might be available (typically inferred from previous experience with the same system). In this work, we focus on broadband antennas, which will be used to illustrate the approach. Size reduction for this type of antennas is especially detrimental to impedance matching at lower frequencies, therefore the pre-screening process may be based on the location of the first antenna resonance, but also a knowledge of what antenna size is possible to be achieved.

Figure 1 shows the procedure for choosing random samples using the above criteria. The minimum and maximum antenna size, A_{min} and A_{max}, respectively, are obtained from a general engineering insight concerning this class of antennas (and, if available, information from prior design work on a particular structure). The first resonance acceptance range is determined by the target operating bandwidth. For example, for ultra-wideband antennas with $F = [3.1\ 10.6]$ GHz, one may set $F_0 = [2.5\ 4.5]$ GHz. The maximum in-band reflection level L_{max} is normally set in a relaxed manner (e.g., −4 dB), in order to only reject exceptionally bad designs. It should be noted that EM analysis is only executed for samples satisfying geometry conditions $A(\boldsymbol{x}) \geq A_{min}$ and $A(\boldsymbol{x}) \leq A_{max}$. Also, all samples for which EM analysis was carried out will be used for surrogate model construction, although the procedure is run until a prescribed number N_{pre} of observables that satisfy all pre-screening conditions have been identified. Figure 2 shows an exemplary compact wideband antenna and selected reflection responses obtained by sampling within the complete search space X, and observables generated using the pre-selection

1. Input parameters:
   - Target operating bandwidth $F = [f_1\ f_2]$;
   - Minimum antenna size A_{min};
   - Maximum antenna size A_{max};
   - First resonance location acceptance range $F_0 = [f_{0.1}\ f_{0.2}]$
   - Maximum in-band reflection level L_{max};
   - Number of pre-screened initial samples N_{pre};
2. Set $i = 0$ and $j = 0$;
3. Generate random sample $\boldsymbol{x}_{tmp} \in X$;
4. **if** $A(\boldsymbol{x}_{tmp}) \geq A_{min}$ AND $A(\boldsymbol{x}_{tmp}) \leq A_{max}$
   Perform EM simulation and acquire $\boldsymbol{R}_{EM}(\boldsymbol{x}_{tmp})$;
   $j = j + 1$;
   $\boldsymbol{x}_B^{(j)} = \boldsymbol{x}_{tmp}$;
   $\boldsymbol{R}_B^{(j)} = \boldsymbol{R}_{EM}(\boldsymbol{x}_{tmp})$;
   Extract f_0 (frequency location of the first resonance);
   Compute $L_0 = \max\{f \in F : |S_{11}(\boldsymbol{x}_{tmp},f)|\}$;
   if $f_0 \geq f_{0.1}$ and $f_0 \leq f_{0.2}$ and $L_0 \leq L_{max}$
   $i = i + 1$;
   end
   end
5. **if** $i < N_{pre}$
   Go to 3;
   else
   $N_{init} = j$;
   End;
   end

Fig. 1. Pre-screening procedure for size reduction of broadband antennas.

procedure of Fig. 1. The latter are confined to a relatively small subset of the space X, which facilitates a subsequent identification of the metamodel.

A similar pre-screening criteria may be readily developed for other types of antennas, e.g., multi-band structures, circular polarization devices, etc. In each case, appropriate acceptance conditions should be formulated depending on the constraints imposed upon electrical and field responses.

Surrogate model construction

In this research, we use kriging interpolation[71,105] as the modelling method of choice. Following the pre-screening procedure, an initial surrogate model $\boldsymbol{s}^{(0)}(\boldsymbol{x})$ is constructed based on the set of training designs $\left\{\boldsymbol{x}_B^{(j)}, R_B^{(j)}\right\}, j = 1, \ldots, N_{init},$ with $R_B^{(j)} = R_{EM}\left(\boldsymbol{x}_B^{(j)}\right)$ cf. Figure 1. Subsequently, the surrogate model is refined to $\boldsymbol{s}^{(j)}$, $j = 1, 2, \ldots$, by incorporating the extended set of samples $\{\boldsymbol{x}_B^{(j)}, \boldsymbol{R}_B^{(j)}\}$, $j = 1, \ldots, N_{init}, N_{init} + 1, \ldots$, obtained as described in Section "Generating Infill Points". For the reader's convenience, a concise overview of kriging is outlined in Fig. 3, considering scalar system responses. The extension to vector-valued functions is straightforward. Specifically, the components of the vector-valued metamodel $\boldsymbol{s}^{(j)}$ represent kriging interpolation surrogates created for pertinent antenna responses (in this case, the reflection characteristic) at all individual frequencies within the frequency sweep under consideration.

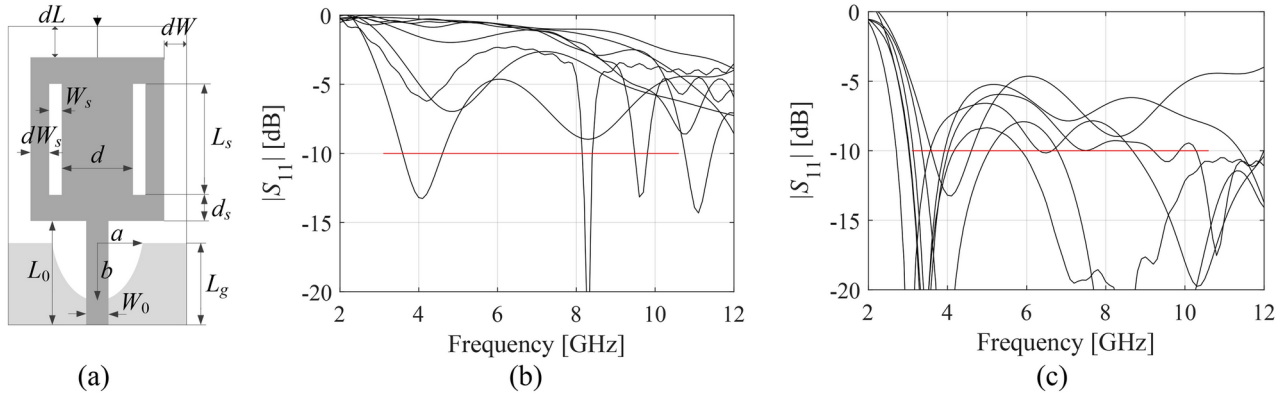


Fig. 2. A broadband antenna example: (**a**) antenna structure, (**b**) $|S_{11}|$ responses at random designs, (c) reflection responses at designs obtained using pre-screening; horizontal line denotes target operating frequency range (here, 3.1 GHz to 10.6 GHz) and level (here, −10 dB). Note that the responses obtained through pre-screening generally correspond to higher-quality designs, e.g., fulfilling the constraint $|S_{11}(\boldsymbol{x},f)| \leq -10$ dB for $f \in F$; also, due to constraints imposed on minimum/maximum size, they are confined to a relatively small subset of the search space, which facilitates a subsequent identification of the fast metamodel.

Generating Infill Points

The pre-screening stage (Section "Parameter space pre-screening") and a rendition of the initial metamodel $\boldsymbol{s}^{(0)}$ (Section "Surrogate model construction") is followed by a machine learning (ML) process, where a series of designs (infill points) is generated as $\boldsymbol{x}^{(i)}$, $i = N_{init}+1, N_{init}+2, \ldots$ ML also produces the refined surrogate models $\boldsymbol{s}^{(j)}$, $j = 1, 2, \ldots$.

The design $\boldsymbol{x}^{(i+1)}$ is found through optimization of the current metamodel, i.e., we have

$$\mathbf{x}^{(i+1)} = \arg\min_{\mathbf{x}\in X} U_P(\mathbf{s}^{(i)}(\mathbf{x})) \tag{4}$$

The merit function U_p has identical analytical form as the function described in Section "Design constraints. Explicit and implicit constraint handling"; however, it is evaluated using the surrogate model rather than EM simulation. This is indicated by explicitly including $\boldsymbol{s}^{(i)}$ as an argument of U_P. The problem (4) is solved in a global sense by means of the particle swarm optimization (PSO) algorithm[100]. As the metamodel is computationally cheap, a particular choice of the algorithm is of minor importance. Furthermore, it can be executed without much concern about computational budget as the cost of virtually unlimited number of merit function evaluations can be neglected when compared to a single EM analyzis of the antenna.

Upon solving (4), EM analysis is carried out at $\boldsymbol{x}^{(i+1)}$ to obtain $\boldsymbol{R}_{EM}(\boldsymbol{x}^{(i+1)})$. This data complements the training dataset. The refined surrogate model $\boldsymbol{s}^{(i+1)}$ is then constructed from the sample set $\{\boldsymbol{x}_B^{(j)}, \boldsymbol{R}_B^{(j)}\}$, $j = 1, \ldots, N_{init}$, $N_{init}+1, \ldots, N_{init}+i, N_{init}+i+1$, and employed as predictor for the subsequent iteration.

In the context of machine learning, generating the candidate design as per (4) is analogous to employing the infill criterion based on minimization of the predicted objective function[106]. This choice is motivated by two main factors. Firstly, the global accuracy of the metamodel is of secondary importance in comparison to the primary goal of identifying the constrained optimum in the process. Secondly, the pre-screening stage outlined in Section "Parameter space pre-screening" enables a rough allocation of the most promising region within the search space, along with generating several samples in that area. This ensures a satisfactory quality of the initial surrogate.

The termination criteria of the ML process are as follows:

- Convergence in argument, $||\boldsymbol{x}^{(i+1)} - \boldsymbol{x}^{(i)}|| < \varepsilon$,
- No improvement of the merit function over the last $N_{no\_improve}$ iterations.

The default values of the control parameter utilized in the numerical experiments of Section "Demonstration examples" are: $\varepsilon = 10^{-3}$, $N_{no\_improve} = 20$.

Complete optimization procedure

Here, we put together the proposed machine learning size reduction procedure. The pseudocode of the method is provided in Fig. 4. Meanwhile, Fig. 5 provides its flow diagram. The two main stages of the algorithm include pre-selection described in Section "Parameter space pre-screening", and the core search process (generating a sequence of infill points), elaborated on in Section "Generating Infill Points". As mentioned in Section "Surrogate model construction", the surrogate modelling technique of choice is kriging interpolation. Table 3 gathers the control parameters, which are only two, both related to the termination condition, and used to decide upon the resolution of identifying the optimum. For example, ε of 10^{-3} corresponds to 0.001 mm, which is more than sufficient for typical antenna components.

Function of interest:

$$f(\boldsymbol{x}) = \boldsymbol{g}(\boldsymbol{x})^T \boldsymbol{\beta} + Z(\boldsymbol{x}) \quad \text{(K1)}$$

where

- $\boldsymbol{g}(\boldsymbol{x}) = [g_1(\boldsymbol{x})\ g_2(\boldsymbol{x})\ \ldots\ g_N(\boldsymbol{x})]^T$ are known system responses,
- $\boldsymbol{\beta} = [\beta_1\ \beta_2\ \ldots\ \beta_N]^T$ are the unknown hyperparameters,
- $\boldsymbol{g}(\boldsymbol{x})^T\boldsymbol{\beta}$ - the regression component serving as a trend function for f,
- $Z(\boldsymbol{x})$ - a realization of a normally distributed Gaussian random process with zero mean and variance σ^2; $Z(\boldsymbol{x})$ manages localized variations from the trend

Model components:

- Covariance matrix of $Z(\boldsymbol{x})$

$$Cov[Z(\boldsymbol{x}^{(i)})Z(\boldsymbol{x}^{(j)})] = \sigma^2 \boldsymbol{R}([R(\boldsymbol{x}^{(i)}, \boldsymbol{x}^{(j)})]) \quad \text{(K2)}$$

  where $\boldsymbol{R}$ is a $p \times p$ correlation matrix with $R_{ij} = R(\boldsymbol{x}^{(i)},\boldsymbol{x}^{(j)})$, and $R(\boldsymbol{x}^{(i)}, \boldsymbol{x}^{(j)})$ is the correlation function between data samples $\boldsymbol{x}^{(i)}$ and $\boldsymbol{x}^{(j)}$,
- Correlation function: Gaussian function of the form

$$R(\boldsymbol{x}, \boldsymbol{y}) = \exp\left[-\sum_{k=1}^{n} \theta_k \mid x_k - y_k \mid^2\right] \quad \text{(K3)}$$

  where θ_k are the unknown correlation parameters, and x_k and y_k are the k^{th} elements of the vectors $\boldsymbol{x}$ and $\boldsymbol{y}$, respectively.

Kriging model:

$$\boldsymbol{s}(\boldsymbol{x}) = \boldsymbol{g}(\boldsymbol{x})^T \boldsymbol{\beta} + \boldsymbol{r}^T(\boldsymbol{x})\boldsymbol{R}^{-1}(\boldsymbol{h} - \boldsymbol{G}\boldsymbol{\beta}) \quad \text{(K4)}$$

where

$$\boldsymbol{r}(\boldsymbol{x}) = [R(\boldsymbol{x}, \boldsymbol{x}^{(1)}) \ldots R(\boldsymbol{x}, \boldsymbol{x}^{(p)})] \quad \text{(K5)}$$

$$\boldsymbol{f} = [f(\boldsymbol{x}^{(1)})\ f(\boldsymbol{x}^{(2)})\ \ldots\ f(\boldsymbol{x}^{(p)})]^T \quad \text{(K6)}$$

and $\boldsymbol{G}$ is a $p \times N$ matrix with $G_{ij} = P_j(\boldsymbol{x}^{(i)})$. The vector of model parameters $\boldsymbol{\beta}$ can be computed as

$$\boldsymbol{\beta} = (\boldsymbol{G}^T\boldsymbol{R}^{-1}\boldsymbol{G})^{-1}\boldsymbol{G}^T\boldsymbol{R}^{-1}\boldsymbol{f} \quad \text{(K7)}$$

Model identification: optimization of maximum likelihood to identify hyperparameters θ_k

$$-[p\ln(\sigma^2) + \ln|\boldsymbol{R}|]/2 \quad \text{(K8)}$$

In (8), both σ^2 and $\boldsymbol{R}$ are functions of θ_k.

Fig. 3. Kriging interpolation modeling [71].

Table 4 puts together the pre-screening-related parameters, most of which are problem dependent. As mentioned earlier, the minimum and maximum antenna size thresholds normally come from both general engineering insight and possibly antenna-specific information, for example, previous experience with the same or similar structures. The frequency and level acceptance threshold are more generic and typically established in a relaxed manner (e.g., the frequencies $f_{0.1}$ and $f_{0.2}$ may be set at about $0.8f_1$ and $1.5f_1$, respectively).

Demonstration examples

This part of the paper addresses verification of the machine learning algorithm for global size reduction of antennas, introduced in Section "Global size reduction using machine learning and parameter space pre-screening". We start by outlining the verification case studies, which include four broadband microstrip structures. An outline of the experimental setup is followed by presentation of the results, as well as optimization process visualization for selected algorithm executions. Our approach is juxtaposed against several benchmark techniques: (i) a multiple start gradient-based search, (ii) gradient-based search with adaptive penalty coefficients, (iii) nature-inspired optimization using a particle swarm optimization routine, and (iv) a machine-learning-based algorithm, which does not employ the pre-screening procedure. The numerical experiments are oriented towards investigating the performance indicators of the considered methods, i.e., the ability to identify the optimum designs and to control design constraints, as well as computational efficiency. It should also be emphasized that for illustration purposes, antenna size reduction is carried out using a single constraint, which is a condition $|S_{11}| \leq -10$ dB

1. Input parameters:
   - Design specifications, including target operating bandwidth $F = [f_1\ f_2]$;
   - Design space $X = [\boldsymbol{l}\ \boldsymbol{u}]$ (here, $\boldsymbol{l}$ and $\boldsymbol{u}$ are lower and upper design variable bounds);
   - Pre-selection parameters:
     - Antenna size bounds A_{min} and A_{max};
     - First resonance location acceptance range $F_0 = [f_{0.1}\ f_{0.2}]$;
     - Maximum in-band reflection level L_{max};
     - Number of pre-screened initial samples N_{pre};
   - Termination thresholds ε and $N_{no\_improve}$;
2. Generate the set of initial samples $\{\boldsymbol{x}_B^{(k)}, \boldsymbol{R}_B^{(k)}\}_{k=1,\ldots,Ninit}$, cf. Section 3.1;
3. Construct initial surrogate model $\boldsymbol{s}^{(0)}(\boldsymbol{x})$;
4. Set $i = 0$;
5. Obtain infill point $\boldsymbol{x}^{(i+1)}$ by solving (4) using the PSO algorithm:

$$\boldsymbol{x}^{(i+1)} = \arg\min_{\boldsymbol{x}\in X} U_P(\boldsymbol{s}^{(i)}(\boldsymbol{x}))$$

6. Update the dataset: $\{\boldsymbol{x}_B^{(k)}, \boldsymbol{R}_B^{(k)}\}_{k=1,\ldots,Ninit+i+1}$, with $\boldsymbol{x}_B^{(Ninit+j)} = \boldsymbol{x}^{(j)}$ for $j = 1, 2, \ldots$, and $\boldsymbol{R}_B^{(k)} = \boldsymbol{R}_{EM}(\boldsymbol{x}_B^{(k)})$;
7. Set $i = i + 1$;
8. Construct the surrogate model $\boldsymbol{s}^{(i)}(\boldsymbol{x})$ using the updated dataset;
9. **if** $||\boldsymbol{x}^{(i)} - \boldsymbol{x}^{(i-1)}|| < \varepsilon$ OR no objective function improvement for $N_{no\_improve}$ iterations
   Go to 11;
   end
10. Go to 5;
11. Return $\boldsymbol{x}^* = \boldsymbol{x}^{(i)}$;

Fig. 4. Pseudocode of the proposed machine-learning-based size reduction algorithm.

over the frequency range of interest. It can also be noted that fulfilling this constraint automatically ensures that the operating bandwidth of the antenna is achieved as well. Optimization under multiple constraints (e.g., associated with gain variability, etc.) will be addressed elsewhere.

Verification antennas

The validation of our algorithm is conducted using four broadband antenna structures depicted in Figs. 6(a), (b), (c), and (d). These antennas are denoted as Antenna I[107], Antenna II[108], Antenna III[109], and Antenna IV[110], respectively. Essential parameters of Antennas I through IV, such as substrate data, geometry variables, and search spaces, are provided in Fig. 6(e). It is worth noting that all four antennas are microstrip-fed monopoles, and the target operating bandwidth is from 3.1 to 10.6 GHz. Further details about the structures can be found in the referenced literature[107–110]. The electromagnetic analysis is carried out using the time-domain solver of CST Microwave Studio. The EM models incorporate subminiature version A (SMA) connectors[111]. The average simulation times of the EM models of Antennas I, II, III, and IV are 2.5, 7.1, 4.5, and 1.8 min, respectively.

The main design goal is to reduce the antenna size $A(\boldsymbol{x})$, which is understood here as the substrate area. Optimization is constrained with a condition imposed upon the impedance matching. More specifically, we have $|S_{11}(\boldsymbol{x},f)| \leq -10$ dB for the frequencies from 3.1 GHz to 10.6 GHz. As mentioned earlier, we employ penalty functions for handling constraints, as elaborated on in Section "Design constraints. Explicit and implicit constraint handling".

One should emphasize that the search spaces are extensive. The number of parameters is seven, eleven, eleven, and fourteen for Antenna I through IV, respectively. Also, the ranges of design variables are significant: the average ratio between upper and lower bound is 20, 5, 35, and 13 (parameters with zero lower bound are excluded), for Antenna I, II, III, and IV, respectively. Antenna IV is particularly challenging, as it will become evident based on the results shown in Section "Results".

It should be mentioned that for compact antennas, it is the electrical size that is often considered instead of the absolute area expressed in, e.g., mm^2. The electrical size is defined as $k{\cdot}a$, where k is the wave number and a is the radius of the smallest sphere enclosing the antenna. However, from practical perspective absolute area is normally more important because of the certain physical space allocated for an antenna in a particular system. From this perspective, the particular value of $k{\cdot}a$ is of secondary importance.

Experimental setup

The proposed algorithm has been used to optimize the antennas of Section "Verification antennas" for minimum size. The control parameters have been set to $\varepsilon = 10^{-3}$, and $N_{no\_improve} = 20$ (cf. Section "Global size reduction using machine learning and parameter space pre-screening"), except Antenna IV, for which $N_{no\_improve}$ has been set to 40. The reason is that Antenna IV is by far the most complex test case, and it is expected that the objective

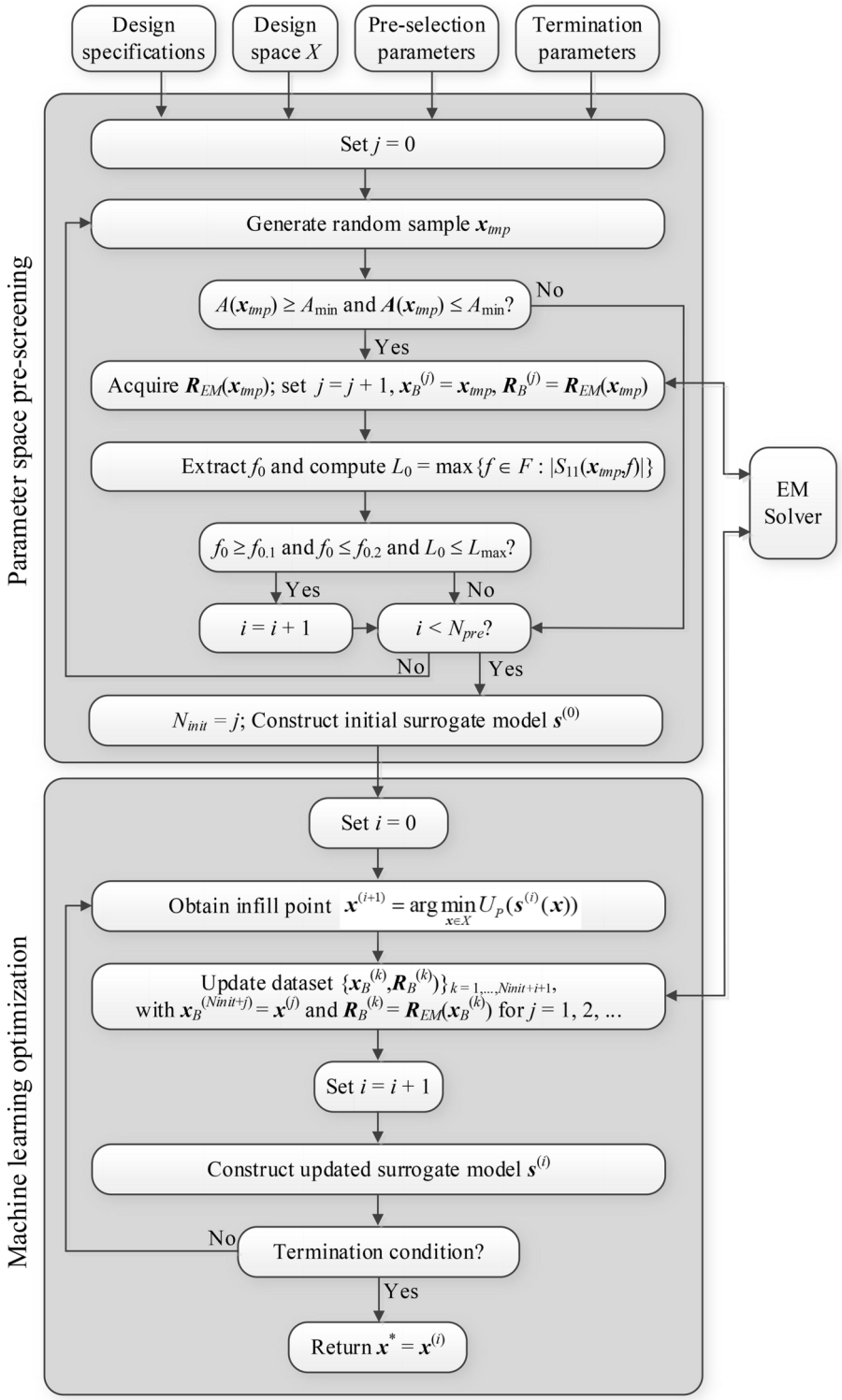


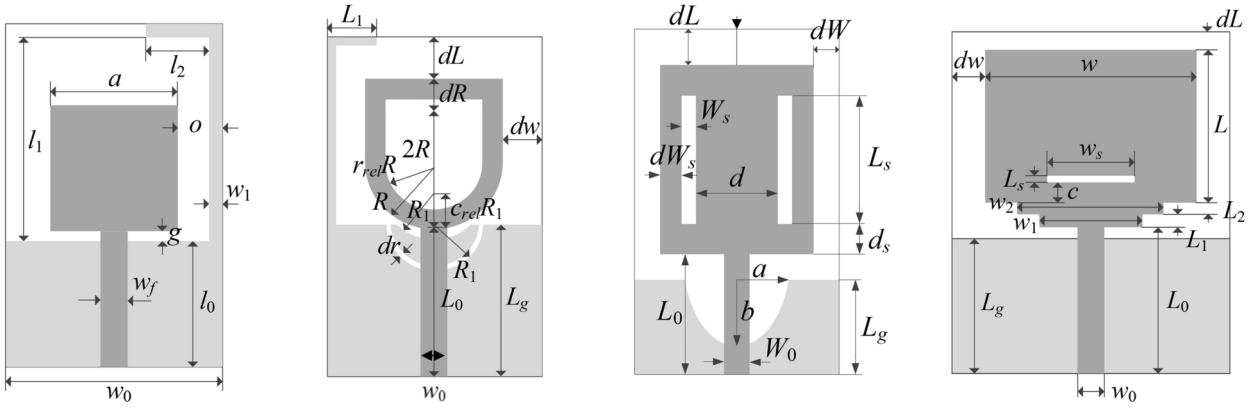
Fig. 5. Proposed machine-learning-based antenna size reduction algorithm: flow diagram.

Parameter	Meaning	Default value
ε	Termination threshold for convergence in argument, cf. Section "Surrogate model construction"	10^{-2}
$N_{no\_improve}$	Termination threshold for no objective function value improvement, cf. Section "Surrogate model construction"	20

Table 3. Suggested miniaturization procedure: control parameters.

Parameter	Meaning	Default value
A_{min}	Minimum antenna size	Problem dependent
A_{max}	Maximum antenna size	Problem dependent
$F_0 = [f_{0.1} f_{0.2}]$	Frequency acceptance bounds for first antenna resonance	Problem dependent
L_{max}	Acceptance level for for maximum in-band $\|S_{11}\|$	4 dB
N_{pre}	Number of pre-selected data points	80

Table 4. Proposed miniaturization procedure: pre-screening parameters.



Antenna	I [107]	II [108]	III [109]	IV [110]
Substrate	RF-35 ($\varepsilon_r = 3.5$, $h = 0.762$ mm)	RF-35 ($\varepsilon_r = 3.5$, $h = 0.762$ mm)	FR-4 ($\varepsilon_r = 4.4$, $h = 1.52$ mm)	RO4350 ($\varepsilon_r = 3.48$, $h = 0.762$ mm)
Designable Parameters [mm]	$\boldsymbol{x} = [l_0\ g\ a\ l_1\ l_2\ w_1\ o]^T$	$\boldsymbol{x} = [L_0\ dR\ R\ r_{rel}\ dL\ dw\ Lg\ L_1\ R_1\ dr\ c_{rel}]^T$	$\boldsymbol{x} = [L_0\ dR\ R\ r_{rel}\ dL\ dw\ Lg\ L_1\ R_1\ dr\ c_{rel}]^T$	$\boldsymbol{x} = [L_0\ L_1\ L_2\ L\ dL\ L_g\ w_1\ w_2\ w\ dw\ L_s\ w_s\ c]^T$
Other Parameters [mm]	$w_0 = 2o + a$, $w_f = 1.7$	$w_0 = 1.7$	$W_0 = 3.0$	$w_0 = 1.7$
Parameter space X	$\boldsymbol{l} = [10\ 8\ 4\ 5\ 1\ 0.1\ 0.2]^T$ $\boldsymbol{u} = [35\ 20\ 15\ 12\ 15\ 10\ 3]^T$	$\boldsymbol{l} = [4\ 0\ 3\ 0.1\ 0\ 0\ 4\ 0\ 2\ 0.2\ 0.2]^T$ $\boldsymbol{u} = [15\ 6\ 8\ 0.9\ 5\ 8\ 15\ 6\ 5\ 1\ 0.9]^T$	$\boldsymbol{l} = [2\ 2\ 2\ 0.2\ 2\ 0\ 0.1\ 0.1\ 0\ 0.01\ 0.01]^T$ $\boldsymbol{u} = [15\ 15\ 20\ 3\ 15\ 15\ 3\ 8\ 5\ 0.8\ 0.8]^T$	$\boldsymbol{l} = [5\ 0.1\ 0.1\ 5\ 0\ 5\ 0.1\ 0.1\ 5\ 0\ 0.01\ 0.1\ 0.01]^T$ $\boldsymbol{u} = [20\ 2\ 3\ 20\ 5\ 20\ 1\ 1\ 25\ 10\ 0.2\ 0.9\ 0.3]^T$

(e)

Fig. 6. Test antenna structures: (**a**) Antenna I, (**b**) Antenna II, (**c**) Antenna III, (**d**) Antenna IV (light gray shade used to indicate ground planes), (**e**) essential parameters.

function value would fluctuate more than in the case of Antennas I through III from one iteration to another. Table 5 shows the values of pre-screening parameters, which are identical in terms of F_0, L_{max}, and N_{pre}. Also, the maximum footprint area threshold A_{max} has been set identical for all antennas. The only differences are in the case of the minimum area threshold, which varies between 180 mm^2 for Antenna III and 250 mm^2 for Antenna II. These numbers come from prior work with the same structures. Nonetheless, the differences are minor, e.g., setting $A_{min} = 200$ mm^2 for all antennas would be just fine. The penalty coefficients β were set to 10^4 which provides a good trade-off between the miniaturization rate and the quality of constraint handling. In particular, given that the typical antenna size is a few hundred mm^2, the contribution of the penalty term (cf. (3)) is 100 for constraint violation of 0.1 dB, and quickly goes up if the violation increases. This setup is generally sufficient to ensure that the constraint violation at the optimized design does not exceed a fraction of dB, which is practically acceptable.

Our procedure has been juxtaposed against four state-of-the-art algorithms outlined in Table 6. The first benchmark method is a gradient-based optimizer[112], initialized from a random starting point. It uses penalty functions (cf. Section "Design constraints. Explicit and implicit constraint handling") with a fixed penalty coefficient β set to 10^3 (version I), 10^4 (version II), and 10^5 (version III). This particular setup allows us to

Parameter	Antenna			
	I	II	III	IV
A_{min}	200 mm^2	250 mm^2	180 mm^2	200 mm^2
A_{max}	350 mm^2	350 mm^2	350 mm^2	350 mm^2
$F_0=[f_{0.1}f_{0.2}]$	[2.5 4.5]	[2.5 4.5]	[2.5 4.5]	[2.5 4.5]
L_{max}	4 dB	4 dB	4 dB	4 dB
N_{pre}	80	80	80	80

Table 5. Pre-screening parameters for Antennas I through IV.

Algorithm	Algorithm type	Setup
I	Trust-region gradient based optimizer[112]	Algorithm setup: • Random initial design; • Response gradients estimated using finite differentiation[115]; • Termination criteria based on convergence in argument and reduction of the trust region size[112]; • Implicit constraint handling with fixed penalty coefficient: $\beta=10^3$ (version I), $\beta=10^4$ (version II), and $\beta=10^5$ (version III),
II	Trust-region gradient based optimizer with adaptive penalty coefficients[107]	Algorithm setup: • Random initial design; • Response gradients estimated using finite differentiation; • Termination criteria based on convergence in argument and the trust region size reduction; • Implicit constraint handling with adaptive penalty coefficients[107]
III	Particle swarm optimizer (PSO)	Algorithm setup: • Swarm size $N=10$, • Standard control parameters ($\chi=0.73$, $c_1=c_2=2.05$); • Number of iterations set to 100
IV	Machine learning procedure	Algorithm similar to that of Section "Global size reduction using machine learning and parameter space pre-screening": • Initial surrogate set up to ensure relative RMS error not higher than 10% with the maximum number of training samples equal to 400; • No pre-screening procedure applied; • Infill criterion: minimization of the projected objective function improvement[113]

Table 6. Benchmark size reduction procedures.

illustrate the design trade-offs when the penalty factors are too small or too large, as compared to the setup utilized by the proposed technique. The second method is a trust-region optimizer using an adaptive penalty function method[107], where the penalty coefficient is dynamically adjusted to improve the control over the design constraint[107]. Gradient-based methods are included to illustrate the necessity of global search for the specified test problems. The third method employed is a particle swarm optimizer (PSO)[113], chosen as a representative nature-inspired approach. It is important to highlight that the maximum number of objective function evaluation for algorithm is set to 1000, which is relatively low for this category of techniques but almost impractical for electromagnetic (EM)-driven optimization (typical running time of two to three days!). The fourth method is a machine learning procedure using the same infill criterion as described in Section "Global size reduction using machine learning and parameter space pre-screening" (predicted objective function improvement); however, it is executed without any pre-screening procedure. Due to this, the initial surrogate is constructed to achieve ten percent of relative RMS error (estimated using cross-validation[114]), with the maximum number of data samples set to 400 (whichever occurs first). This method is included to corroborate the relevance of the pre-screening stage.

Results

Tables 7, 8, 9, and 10 gather the numerical results. Note that the performance figures reported in the tables (antenna size, constraint violation, and CPU cost) are the average values and the standard deviations computed for ten independent runs of each method (benchmark algorithms and the proposed procedure). The computational cost of the optimization process is expressed in the following format: $N_{EM}\times\boldsymbol{R}$ [N_T h], where N_{EM} is the average number of EM simulations (averaged over the ten algorithm runs), whereas N_T is the total optimization time (the underlying unit is hours). Figures 7, 8, 9, and 10 show the antenna responses, as well as the changes of the antenna size and violation of the impedance matching condition over the algorithm iterations, for the two selected algorithm runs.

Discussion

The data compiled in Tables 7 through 10 enable us to draw several observations regarding the operation of the suggested machine learning technique and to compare it to the benchmark methods. The analysis encompasses various aspects of the search process, including design quality measured by the attained antenna size and violation of constraints, reliability assessed through solution repeatability, and computational efficiency.

- Design quality. Exquisite design quality is one of the essential advantages of the presented approach. For all four antenna structures, our technique yields the smallest size while ensuring excellent control over the reflection characteristics (average constraint violation is a small fraction of a decibel). The achieved footprint

Optimization algorithm		Performance figure				
		Antenna size A [mm^2][a]	Std(A) [mm^2][b]	Constraint violation C [dB][c]	Std(C) [dB][d]	CPU cost[e]
Algorithm I	$\beta = 10^3$	318.1	42.6	1.2	0.4	43.8 × $\boldsymbol{R}$ [1.8 h]
	$\beta = 10^4$	317.7	42.3	0.4	0.7	42.2 × $\boldsymbol{R}$ [1.8 h]
	$\beta = 10^5$	318.8	43.3	0.1	0.2	41.4 × $\boldsymbol{R}$ [1.7 h]
Algorithm II		314.1	42.3	0.3	0.2	50.0 × $\boldsymbol{R}$ [2.1 h]
Algorithm III		360.9	67.5	0.5	0.9	1,000 × $\boldsymbol{R}$ [42.0 h]
Algorithm IV		256.7	11.9	0.3	0.3	473.6 × $\boldsymbol{R}$ [20.0 h]
Machine learning with pre-screening stage (this work)		251.8	2.0	0.1	0.05	215.3 × $\boldsymbol{R}$ [9.0 h]

Table 7. Antenna I: optimization results. [a]Optimized footprint area of the circuit averaged over ten algorithm runs. [b]Standard deviation of the optimized footprint area averaged over ten algorithm runs. [c]Violation of the design constraint, defined as $C = \max\{f \in [3.1\ 10.6]\ \text{GHz}: |S_{11}(\boldsymbol{x},f)|\} + 10$ dB, averaged over ten algorithm runs. [d]Standard deviation of the constraint violation, averaged over ten algorithm runs. [e]Cost expressed in terms of equivalent number of EM analyzes (averaged over ten algorithm runs.)

Optimization algorithm		Performance figure				
		Antenna size A [mm^2][a]	Std(A) [mm^2][b]	Constraint violation C [dB][c]	Std(C) [dB][d]	CPU cost[e]
Algorithm I	$\beta = 10^3$	250.4	24.0	1.2	0.5	124.2 × $\boldsymbol{R}$ [14.6 h]
	$\beta = 10^4$	318.6	60.0	0.1	0.1	180.3 × $\boldsymbol{R}$ [21.2 h]
	$\beta = 10^5$	331.6	63.4	0.1	0.1	133.2 × $\boldsymbol{R}$ [15.7 h]
Algorithm II		281.6	37.1	0.2	0.2	181.7 × $\boldsymbol{R}$ [21.4 h]
Algorithm III		399.4	143.6	0.6	0.4	1,000 × $\boldsymbol{R}$ [118 h]
Algorithm IV		310.5	83.5	0.7	0.4	467.9 × $\boldsymbol{R}$ [55.2 h]
Machine learning with pre-screening stage (this work)		259.1	27.6	0.7	0.5	162.3 × $\boldsymbol{R}$ [19.2 h]

Table 8. Antenna II: optimization results. [a]Optimized footprint area of the circuit averaged over ten algorithm runs. [b]Standard deviation of the optimized footprint area averaged over ten algorithm runs. [c]Violation of the design constraint, defined as $C = \max\{f \in [3.1\ 10.6]\ \text{GHz}: |S_{11}(\boldsymbol{x},f)|\} + 10$ dB, averaged over ten algorithm runs. [d]Standard deviation of the constraint violation, averaged over ten algorithm runs. [e]Cost expressed in terms of equivalent number of EM analyzes (averaged over ten algorithm runs.)

Optimization algorithm		Performance figure				
		Antenna size A [mm^2][a]	Std(A) [mm^2][b]	Constraint violation C [dB][c]	Std(C) [dB][d]	CPU cost[e]
Algorithm I	$\beta = 10^3$	212.8	14.3	1.0	0.4	164.9 × $\boldsymbol{R}$ [12.1 h]
	$\beta = 10^4$	255.0	25.1	0.2	0.1	138.1 × $\boldsymbol{R}$ [10.2 h]
	$\beta = 10^5$	280.1	47.4	0.1	0.1	154.0 × $\boldsymbol{R}$ [11.3 h]
Algorithm II		215.6	3.6	0.3	0.1	189.9 × $\boldsymbol{R}$ [14.0 h]
Algorithm III		425.7	145.8	0.2	0.2	1,000 × $\boldsymbol{R}$ [74.2 h]
Algorithm IV		194.8	2.2	0.2	0.2	451.6 × $\boldsymbol{R}$ [33.5 h]
Machine learning with pre-screening stage (this work)		194.6	1.4	0.08	0.05	151.3 × $\boldsymbol{R}$ [11.2 h]

Table 9. Antenna III: optimization results. [a]Optimized footprint area of the circuit averaged over ten algorithm runs. [b]Standard deviation of the optimized footprint area averaged over ten algorithm runs. [c]Violation of the design constraint, defined as $C = \max\{f \in [3.1\ 10.6]\ \text{GHz}: |S_{11}(\boldsymbol{x},f)|\} + 10$ dB, averaged over ten algorithm runs. [d]Standard deviation of the constraint violation, averaged over ten algorithm runs. [e]Cost expressed in terms of equivalent number of EM analyzes (averaged over ten algorithm runs.)

Optimization algorithm		Performance figure				
		Antenna size A [mm^2][a]	Std(A) [mm^2][b]	Constraint violation C [dB][c]	Std(C) [dB][d]	CPU cost[e]
Algorithm I	$\beta=10^3$	727.9	236.0	1.7	1.5	180.3×$\boldsymbol{R}$ [5.3 h]
	$\beta=10^4$	829.5	206.4	1.0	1.9	211.2×$\boldsymbol{R}$ [6.2 h]
	$\beta=10^5$	842.8	130.2	0.4	0.9	248.0×$\boldsymbol{R}$ [7.2 h]
Algorithm II		753.9	243.0	0.9	0.8	230.3×$\boldsymbol{R}$ [6.7 h]
Algorithm III		457.8	59.1	0.7	0.4	1,000×$\boldsymbol{R}$ [29.0 h]
Algorithm IV		482.7	196.8	0.7	0.5	947.2×$\boldsymbol{R}$ [27.5 h]
Machine learning with pre-screening stage (this work)		202.2	12.5	0.2	0.06	472.5×$\boldsymbol{R}$ [13.7 h]

Table 10. Antenna IV: optimization results. [a]Optimized footprint area of the circuit averaged over ten algorithm runs. [b]Standard deviation of the optimized footprint area averaged over ten algorithm runs. [c]Violation of the design constraint, defined as $C=\max\{f\in[3.1\ 10.6]\ \text{GHz}:|S_{11}(\boldsymbol{x},f)|\}+10$ dB, averaged over ten algorithm runs. [d]Standard deviation of the constraint violation, averaged over ten algorithm runs. [e]Cost expressed in terms of equivalent number of EM analyzes (averaged over ten algorithm runs.)

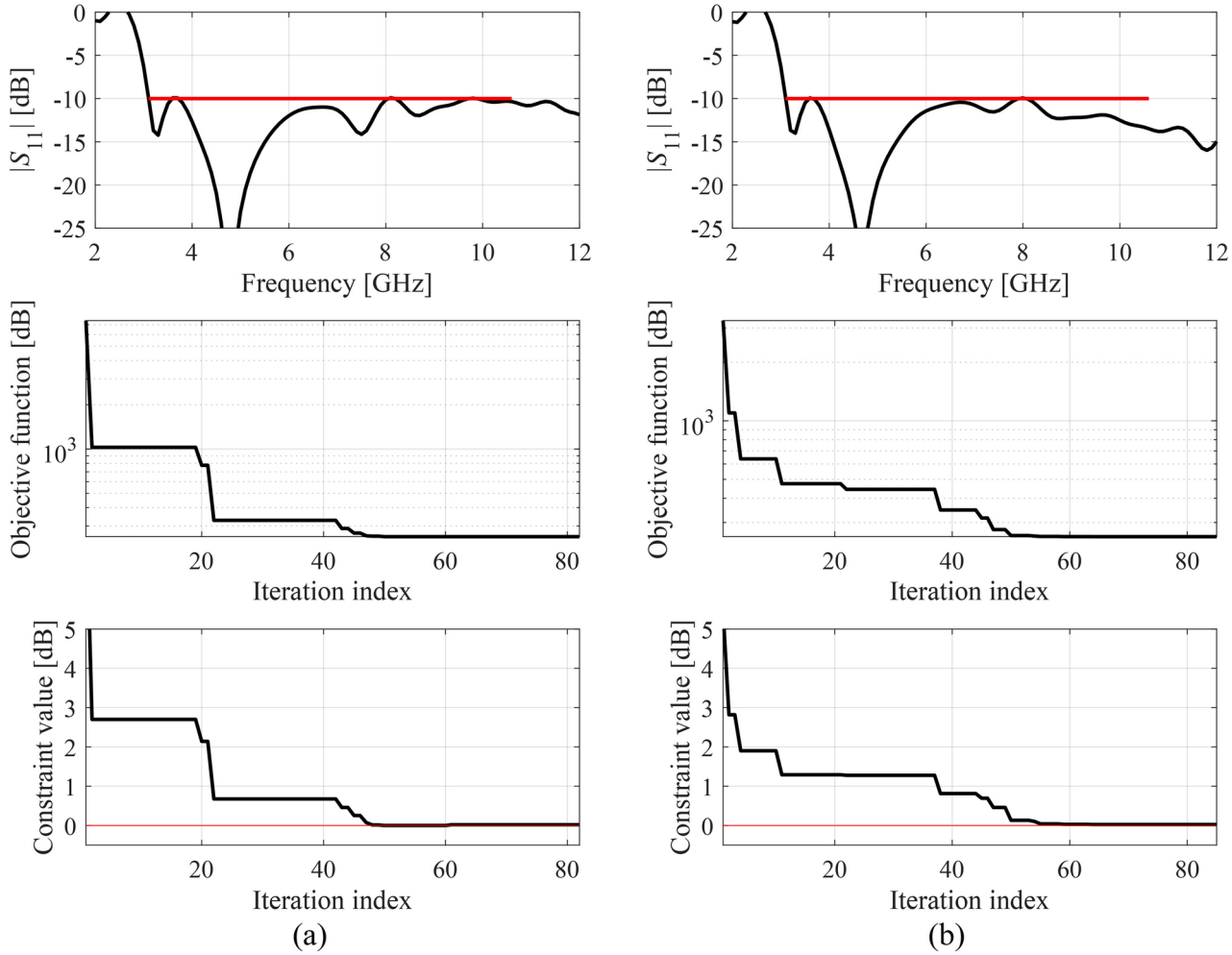


Fig. 7. $|S_{11}|$ of Antenna I at the minimum-size designs rendered by means of our algorithm (top), the antenna size versus iteration index (middle), and design constraint violation versus iteration index (bottom), shown for the representative search runs: (**a**) run 1, (**b**) run 2. The iteration counter starts after constructing the initial surrogate model. The horizontal lines mark the target operating frequency range (here, from 3.1 GHz to 10.6 GHz), and the acceptance level of –10 dB.

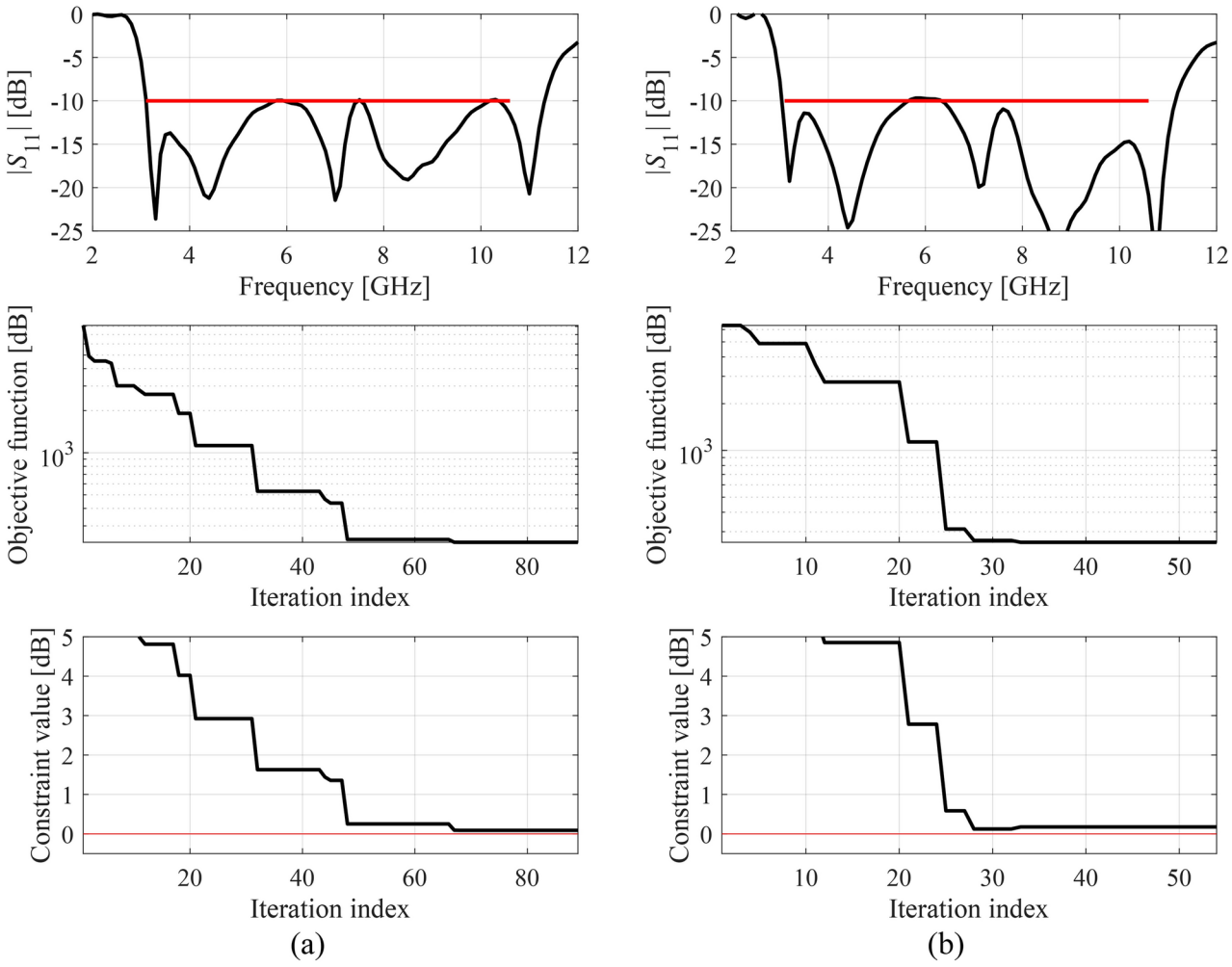


Fig. 8. $|S_{11}|$ of Antenna II at the minimum-size designs rendered by means of our algorithm (top), the antenna size versus iteration index (middle), and design constraint violation versus iteration index (bottom), shown for the representative search runs: (**a**) run 1, (**b**) run 2. The iteration counter starts after constructing the initial surrogate model. The horizontal lines mark the target operating frequency range (here, from 3.1 to 10.6 GHz), and the acceptance level of –10 dB.

areas are significantly smaller than for all benchmark methods, and the difference is particularly pronounced for Antenna IV, which is the most challenging case. Gradient based methods produce inferior results, which corroborates that the considered design tasks are multimodal. Furthermore, constraint control is a problem for Algorithm I: a clear trade-off between miniaturization rate and constraint violation is observed, which indicates the issues related to manual setup of the penalty coefficients. Algorithm II (adaptive penalty factor adjustment) effectively alleviates this difficulty, yet it is unable to overcome a local nature of gradient search. Algorithm III, although formally being a global optimization procedure, also produces inferior results, which is related to limited computational budget assigned to it (1,000 objective function evaluations). This budget is significant (and almost prohibitive) in terms of the CPU time, yet, seems insufficient from the point of view of rendering satisfactory outcome. Finally, the machine learning algorithm without pre-screening stage (Algorithm IV) produces results that are relatively close to those generated by our method, but its performance is inconsistent (e.g., the results are poor for Antennas II and IV).

- Reliability. Here, we are mainly interested in repeatability of results. The standard deviation of the antenna footprint area and of constraint violation is indicative of performance consistency. According to the obtained outcomes, the proposed algorithm demonstrates the highest level of result repeatability, excelling in both antenna size and constraint violation. For Antenna I and III, standard deviation of the size is less than one percent of the average value, and it is by over an order of magnitude lower than for other methods (on the average). Furthermore, the small standard deviation indicates that the solutions found by the algorithm are most likely close to global optimum. Regarding Antennas II and IV, the standard deviation is approximately ten percent and six percent of the average, respectively, representing a substantial reduction compared to the benchmark methods. This is pronounced for Antenna IV, where most of the benchmark methods exhibit large result variability, especially for Algorithms I and II (gradient-based routines) but also Algorithm IV.

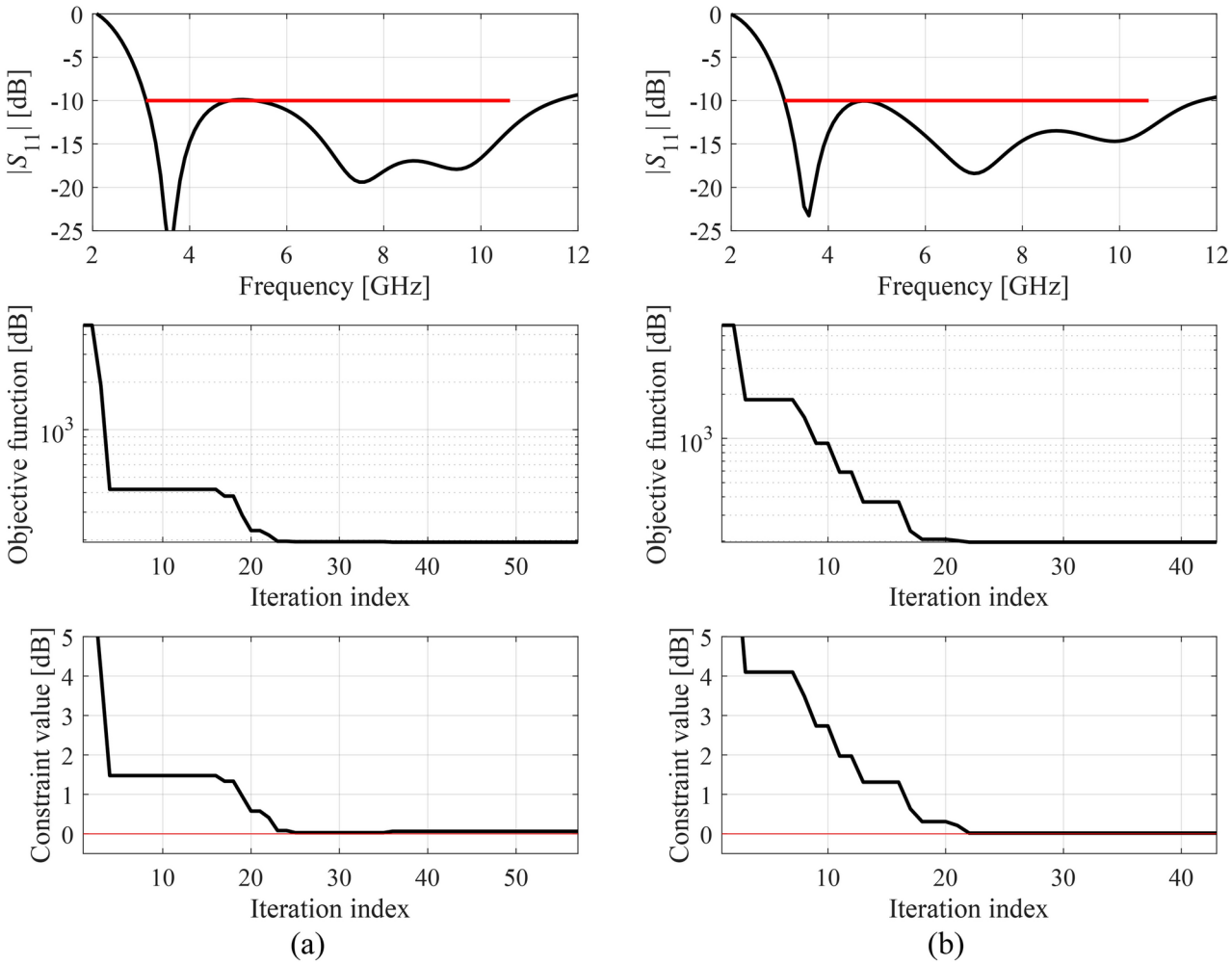


Fig. 9. $|S_{11}|$ of Antenna III at the minimum-size designs rendered by means of our algorithm (top), the antenna size versus iteration index (middle), and design constraint violation versus iteration index (bottom), shown for the representative search runs: (**a**) run 1, (**b**) run 2. The iteration counter starts after constructing the initial surrogate model. The horizontal lines mark the target operating frequency range (here, from 3.1 to 10.6 GHz), and the acceptance level of –10 dB.

- Computational efficiency. The expenses associated with the optimization process are notably higher for the proposed algorithm than for local methods (Algorithms I and II), with a more pronounced effect observed for Antennas I and IV. However, for Antennas II and III, the costs are comparable, highlighting the computational efficiency of the presented technique—achieving global search capability at costs similar to local algorithms. In comparison to the nature-inspired optimization method (PSO), our technique exhibits considerably better efficiency, even though, as mentioned earlier, PSO operates with a low budget compared to the typical setting for this class of algorithms. Additionally, the proposed framework is significantly faster than Algorithm IV, underscoring the importance of the pre-screening stage. The average reduction of the computational cost is as high as sixty percent, and it is accompanied by the added benefit of improved design quality. It should also be noted that the computational cost of the optimization process is not monotonically increasing with respect to the parameter space dimensionality (e.g., the costs are higher for Antenna I than for Antennas II and III). The likely reason is more complex relationship between antenna responses and geometry variables (e.g., higher nonlinearity), which normally translates to higher expenses due to problems with rendering accurate surrogate model. At this point, it should be reiterated that local optimizers (Algorithms I and II) are definitely faster than global search methods, and our purpose was not to beat them in terms of computational efficiency. The point was to illustrate that the results obtained by local optimizers are significantly worse in terms of design quality than those obtained using global methods (Algorithms III and IV, and the proposed technique), due to their inability to identify global optima. This has been clearly demonstrated in Tables 7 through 10.

For additional validation, selected optimum designs were simulated, and the results were presented in Figs. 11, 12, 13, and 14, for Antennas I through IV, respectively. As can be observed, despite considerable size reduction, all antennas maintain omnidirectional radiation as expected for UWB antennas, and excellent level of efficiency.

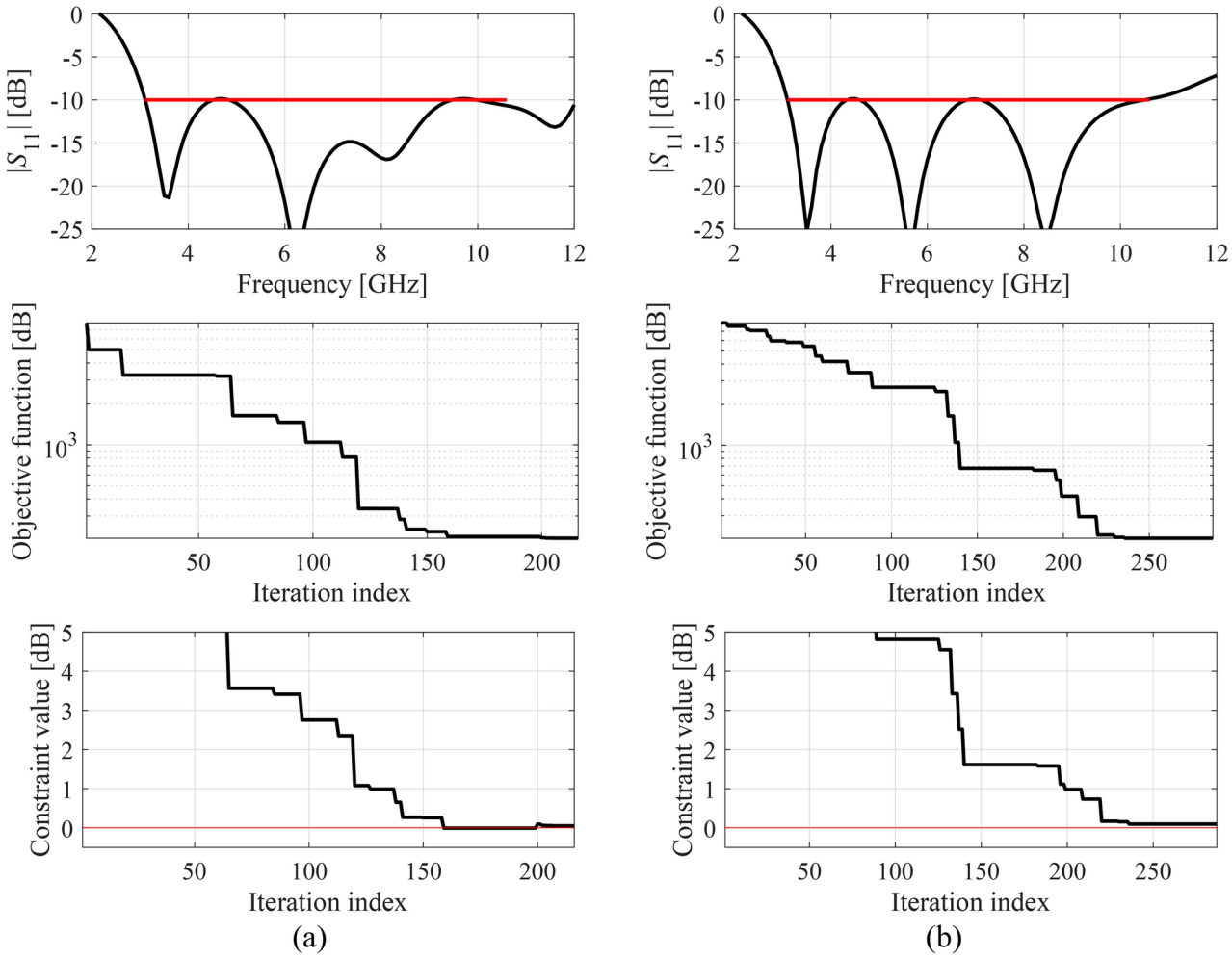


Fig. 10. $|S_{11}|$ of Antenna IV at the minimum-size designs rendered by means of our algorithm (top), the antenna size versus iteration index (middle), and design constraint violation versus iteration index (bottom), shown for the representative search runs: (**a**) run 1, (**b**) run 2. The iteration counter starts after constructing the initial surrogate model. The horizontal lines mark the target operating frequency range (here, from 3.1 to 10.6 GHz), and the acceptance level of –10 dB.

Also, the maximum in-band reflection is close to –10 dB, which means that the design constraint is reliably controlled.

The designs presented in Figs. 11, 12, 13, and 14, were fabricated and experimentally validated. Figure 15 shows the photographs of antenna prototypes, as well as a comparison of EM-simulated and measured reflection characteristics. As observed, the datasets are well-aligned with minor discrepancies resulting from measurement inaccuracies and fabrication/assembly imperfections.

For the sake of supplementary illustration, the results gathered in Tables 7 through 10 have been visualized in Fig. 16, which shows the average antenna size, constraint violation, and the optimization cost.

In summary, the presented machine learning approach significantly surpasses the entire range of the benchmark methods in terms of performance. Our technique does not only render designs of higher quality w.r.t. the achieved antenna size and constraint control, but it is also computationally cheaper than the global search methods included in the benchmark set. This level of performance corroborates utility of the algorithmic tools incorporated into the procedure, including implicit constraint handling, knowledge-based parameter space pre-screening, as well as the predicted objective function improvement as the infill criterion.

In a similar vein, one should note that the our technique is superior to Algorithm IV with respect to all performance indicators, which demonstrates the relevance of the parameter space pre-screening. The latter allows us to quickly identify the promising subsets of the search space, which enables construction—at lower computational cost—a surrogate model featuring improved predictive power. This, in turn, carries over to overall computational efficiency and dependability of the antenna miniaturization process.

Conclusion

This paper introduces an innovative machine learning approach designed for the explicit simulation-based miniaturization of antenna structures. The inherent challenges in this undertaking stem from the high costs

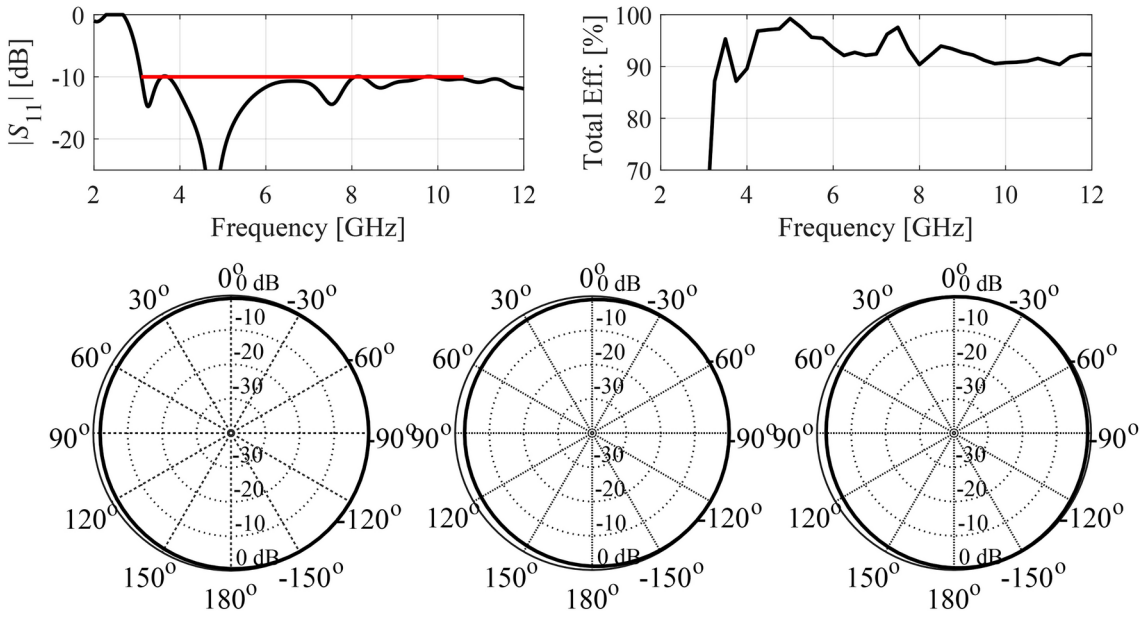


Fig. 11. Antenna I: simulation results for a selected design found using the proposed technique. Shown are reflection response (target operating bandwidth marked using the horizontal line), total efficiency, and H-plane radiation patterns at 4 GHz, 6 GHz, and 8 GHz (from left to right).

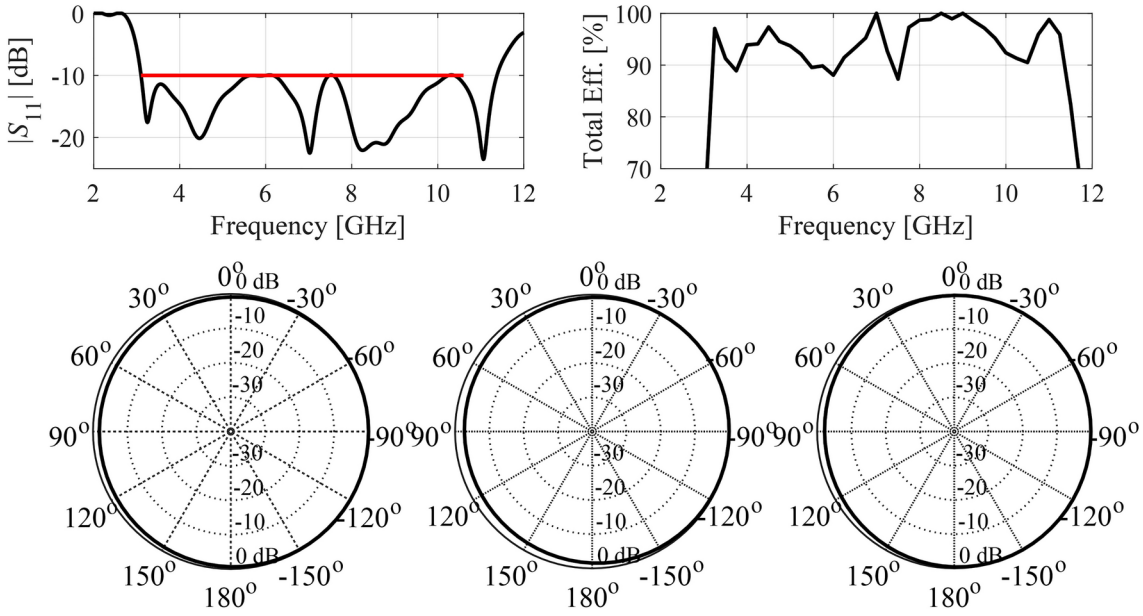


Fig. 12. Antenna II: simulation results for a selected design found using the proposed technique. Shown are reflection response (target operating bandwidth marked using the horizontal line), total efficiency, and H-plane radiation patterns at 4 GHz, 6 GHz, and 8 GHz (from left to right).

associated with evaluating antenna responses, necessitating full-wave EM simulation, the presence of expensive constraints, and the typically extensive search spaces, both with respect to dimensionality and parameter ranges. The presented methodology harnesses a pre-screening strategy to pinpoint promising regions within the parameter space, effectively mitigating the computational expenses. Furthermore, it incorporates an implicit constraint handling mechanism, effectively converting the design problem into a formally unconstrained task. The machine learning scheme, utilizing predicted merit function improvement as an infill criterion, expedites the algorithm's convergence.

The proposed algorithm has been verified based on four broadband antennas. It was also favourably compared to several benchmark methods, including both local and global optimization procedures. The achieved results

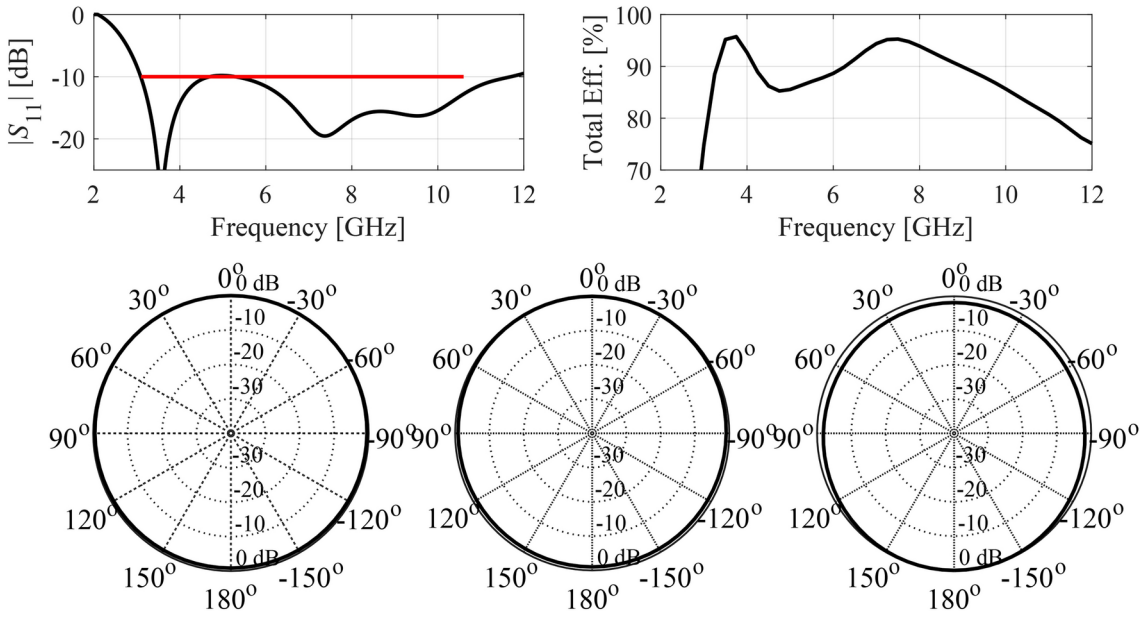


Fig. 13. Antenna III: simulation results for a selected design found using the proposed technique. Shown are reflection response (target operating bandwidth marked using the horizontal line), total efficiency, and H-plane radiation patterns at 4 GHz, 6 GHz, and 8 GHz (from left to right).

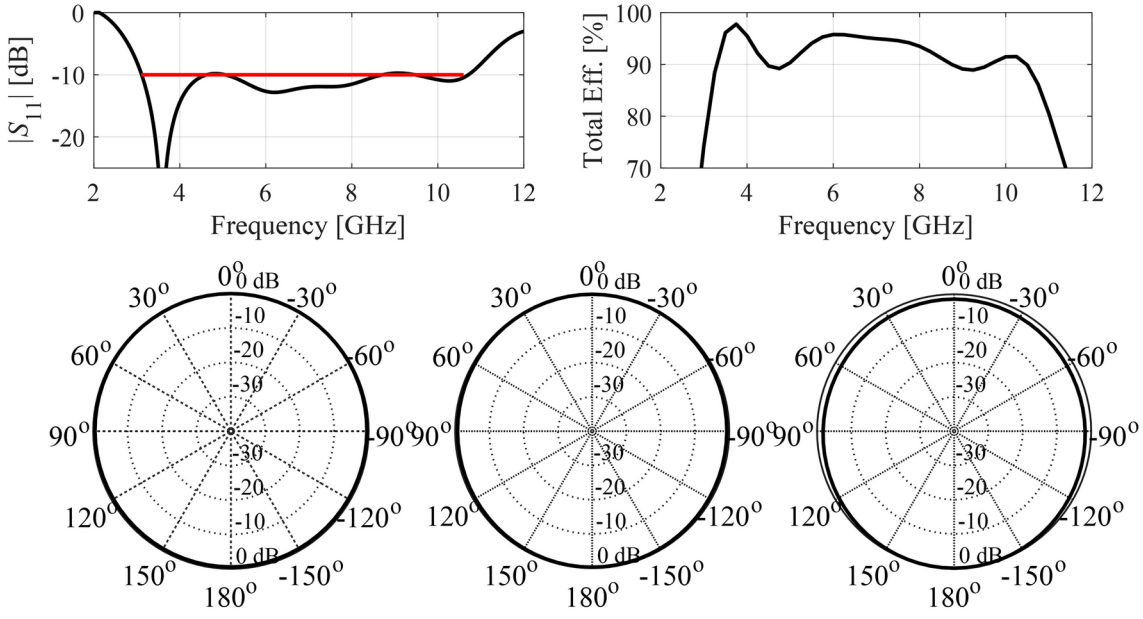


Fig. 14. Antenna IV: simulation results for a selected design found using the proposed technique. Shown are reflection response (target operating bandwidth marked using the horizontal line), total efficiency, and H-plane radiation patterns at 4 GHz, 6 GHz, and 8 GHz (from left to right).

corroborate remarkable performance of our method with respect to design quality, solution repeatability, but also computational efficiency. The standard deviation of the antenna size is only about four percent of the average (while being below one percent for two out of four verification structures), which is indicative of allocating the global optimum. Furthermore, comparison with machine learning approach that does not rely on the pre-screening stage corroborates the relevance of the latter in terms of improving all performance indicators, but also leading to sixty-percent computational savings.

The size reduction method outlined in this paper presents a promising alternative to existing state-of-the-art techniques, particularly when global search requirements are identified, and current methods, especially nature-inspired algorithms, prove excessively costly for direct handling of electromagnetic (EM) simulation antenna

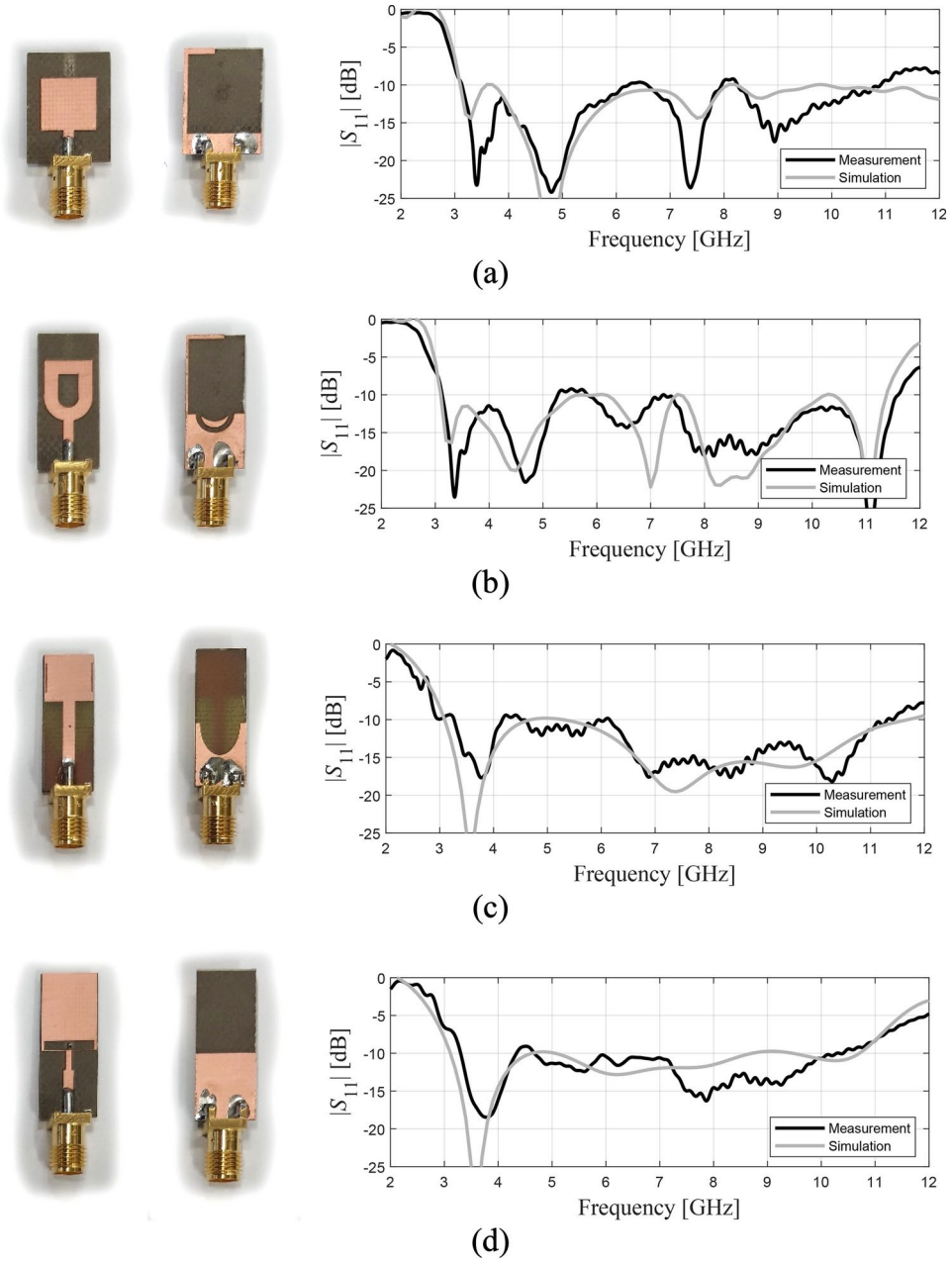


Fig. 15. Experimental validation of the optimized designs of Antennas I through IV presented in Figs. 11, 12, 13, and 14: (**a**) Antenna I, (**b**) Antenna II, (**c**) Antenna III, (**d**) Antenna IV. Shown are antenna prototype, front (left) and back (right), and the comparison between EM-simulated (grey) and measured (black) reflection responses.

models. Future efforts will concentrate on enhancing the pre-screening stage and incorporating methods to more efficiently address dimensionality-related challenges. Another topic for further research is to investigate alternative data-driven modelling techniques to construct the surrogate employed by the machine learning scheme (e.g., neural networks, polynomial chaos expansion, GPR). Furthermore, application of the proposed approach to size reduction of other types of antennas, such as MIMO structures, or directive antennas (e.g., quasi-Yagi) will be considered. While this is generally straightforward as our methodology does not leverage any specific properties of the antenna under optimization, additional constraints might have to be considered (e.g., concerning port isolation, gain, etc.).

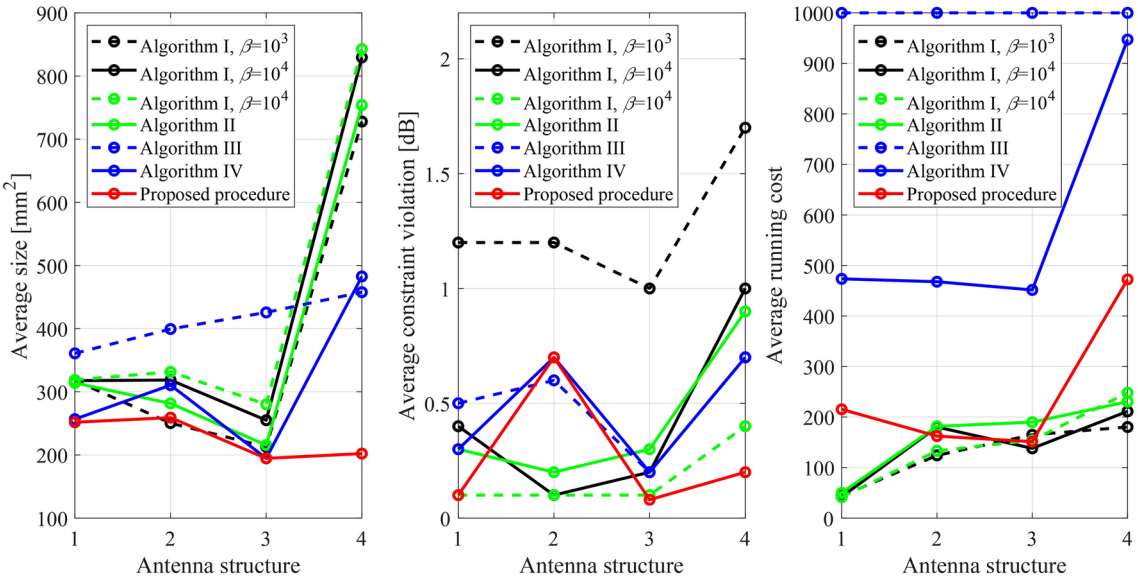


Fig. 16. Visual comparison of results obtained by the proposed and the benchmark methods. Shown are the average antenna size (left), average constraint violation (middle), and average running cost in terms of the number of EM simulations of the antenna at hand (right).

Data availability

The datasets used and/or analysed during the current study available from the corresponding author on reasonable request.

Received: 6 May 2024; Accepted: 15 November 2024
Published online: 27 November 2024

References

1. Sun, L., Li, Y. & Zhang, Z. Wideband decoupling of integrated slot antenna pairs for 5G smartphones. *IEEE Trans. Ant. Prop.* **69**(4), 2386–2391 (2021).
2. Oh, J.-I., Jo, H.-W., Kim, K.-S., Cho, H. & Yu, J.-W. A compact cavity-backed slot antenna using dual mode for IoT applications. *IEEE Ant. Wireless Propag. Lett.* **20**(3), 317–321 (2021).
3. Mahmud, M. Z., Islam, M. T., Misran, N., Kibria, S. & Samsuzzaman, M. Microwave imaging for breast tumor detection using uniplanar AMC based CPW-fed microstrip antenna. *IEEE Access* **6**, 44763–44775 (2018).
4. Xu, L., Xu, J., Chu, Z., Liu, S. & Zhu, X. Circularly polarized implantable antenna with improved impedance matching. *IEEE Ant. Wireless Propag. Lett.* **19**(5), 876–880 (2020).
5. Erman, F., Koziel, S., Hanafi, E., Soboh, R. & Szczepanski, S. Miniaturized metal-mountable U-shaped inductive-coupling-fed UHF RFID tag antenna with defected microstrip surface. *IEEE Access* **10**, 47301–47308 (2022).
6. Koziel, S. Objective relaxation algorithm for reliable simulation-driven size reduction of antenna structure. *IEEE Ant. Wireless Prop. Lett.* **16**(1), 1949–1952 (2017).
7. Liu, J., Esselle, K. P., Hay, S. G. & Zhong, S. Effects of printed UWB antenna miniaturization on pulse fidelity and pattern stability. *IEEE Trans. Ant. Prop.* **62**(8), 3903–3910 (2014).
8. Cheng, T., Jiang, W., Gong, S. & Yu, Y. Broadband SIW cavity-backed modified dumbbell-shaped slot antenna. *IEEE Ant. Wireless Propag. Lett.* **18**(5), 936–940 (2019).
9. Ali, M. Z. & Khan, Q. U. High gain backward scanning substrate integrated waveguide leaky wave antenna. *IEEE Trans. Ant. Prop.* **69**(1), 562–565 (2021).
10. Chopra, R. & Kumar, G. Series-fed binomial microstrip arrays for extremely low sidelobe level. *IEEE Trans. Ant. Propag.* **67**(6), 4275–4279 (2019).
11. Ameen, M., Thummaluru, S. R. & Chaudhary, R. K. A compact multilayer triple-band circularly polarized antenna using anisotropic polarization converter. *IEEE Ant. Wireless Propag. Lett.* **20**(2), 145–149 (2021).
12. Abdullah, M. & Koziel, S. A novel versatile decoupling structure and expedited inverse-model-based re-design procedure for compact single-and dual-band MIMO antennas. *IEEE Access* **9**, 37656–37667 (2021).
13. Ullah, U., Al-Hasan, M., Koziel, S. & Ben Mabrouk, I. Series-slot-fed circularly polarized multiple-input-multiple-output antenna array enabling circular polarization diversity for 5G 28-GHz indoor applications. *IEEE Trans. Ant. Prop.* **69**(9), 5607–5616 (2021).
14. Huang, H., Gao, S., Lin, S. & Ge, L. A wideband water patch antenna with polarization diversity. *IEEE Ant. Wireless Propag. Lett.* **19**(7), 1113–1117 (2020).
15. Zhu, Y. & Dong, Y. A novel compact wide-stopband filter with hybrid structure by combining SIW and microstrip technologies. *IEEE Microwave Wireless Comp. Lett.* **31**(7), 841–844 (2021).
16. Sharon Giftsy, A. L., Kommuri, U. K. & Dwivedi, R. P. Flexible and wearable antenna for biomedical application: progress and opportunity. *IEEE Access* **12**, 90016–90040 (2024).
17. Zhou, Y., Jiang, T., Li, H. & Chen, F. A 5G MIMO multiband low-profile antenna design for automotive shark-fin systems. *IEEE Ant. Wireless Propag. Lett.* **23**(5), 1588–1592 (2024).
18. Qin, X. & Li, Y. Compact dual-polarized cross-slot antenna with colocated feeding. *IEEE Trans. Ant. Propag.* **67**(11), 7139–7143 (2019).
19. He, X. L. & Zhang, Y. A Polarization convert reflective surface-based wideband antenna array decoupling structure. *IEEE Ant. Wireless Propag. Lett.* **23**(2), 883–887 (2024).

20. Haq, M. A. & Koziel, S. Feed line alterations for optimization-based design of compact super wideband MIMO antennas in parallel configuration. *IEEE Ant. Wireless Prop. Lett.* **18**(10), 1986–1990 (2019).
21. Wei, M. et al. Low-profile dual-band microstrip patch antenna with shorting pins. *IEEE Ant. Wireless Propag. Lett.* **23**(6), 1834–1838 (2024).
22. Ding, K., Gao, C., Qu, D. & Yin, Q. Compact broadband MIMO antenna with parasitic strip. *IEEE Ant. Wireless Propag. Lett.* **16**, 2349–2353 (2017).
23. Kovaleva, M., Bulger, D. & Esselle, K. P. Comparative study of optimization algorithms on the design of broadband antennas. *IEEE J. Multiscale Multiphys. Comp. Techn.* **5**, 89–98 (2020).
24. Koziel, S., Pietrenko-Dabrowska, A. & Al-Hasan, M. Frequency-based regularization for improved reliability optimization of antenna structures. *IEEE Trans. Ant. Prop.* **69**(7), 4246–4251 (2020).
25. Koziel, S. & Pietrenko-Dabrowska, A. Global EM-driven optimization of multi-band antennas using knowledge-based inverse response-feature surrogates. *Knowl. Based Systems* **227**, 107189 (2021).
26. Sang, L., Wu, S., Liu, G., Wang, J. & Huang, W. High-gain UWB Vivaldi antenna loaded with reconfigurable 3-D phase adjusting unit lens. *IEEE Ant. Wireless Propag. Lett.* **19**(2), 322–326 (2020).
27. Mahrokh, M. & Koziel, S. Explicit size-reduction of circularly polarized antennas through constrained optimization with penalty factor adaptation. *IEEE Access* **9**, 132390–132396 (2021).
28. Koziel, S., Cheng, Q. S. & Li, S. Optimization-driven antenna design framework with multiple performance constraints. *Int. J. RF Microwave CAE* **28**(4), 21208 (2018).
29. Haq, M. A. & Koziel, S. On topology modifications for wideband antenna miniaturization. *AEU - Int. J. Electr. Comm.* **94**, 215–220 (2018).
30. Zhu, S.-H., Yang, X.-S., Wang, J. & Wang, B.-Z. Design of MIMO antenna isolation structure based on a hybrid topology optimization method. *IEEE Trans. Ant. Propag.* **67**(10), 6298–6307 (2019).
31. Wang, J., Yang, X.-S., Ding, X. & Wang, B.-Z. Topology optimization of conical-beam antennas exploiting rotational symmetry. *IEEE Trans. Ant. Propag.* **66**(5), 2254–2261 (2018).
32. Wu, Y. et al. An optimized multiband antenna for UWB ad hoc networks based on topology optimization theory. *IEEE Trans. Ant. Propag.* **72**(5), 3896–3911 (2024).
33. Soltani, S., Lotfi, P. & Murch, R. D. Design and optimization of multiport pixel antennas. *IEEE Trans. Ant. Propag.* **66**(4), 2049–2054 (2018).
34. Lotfi, P., Soltani, S. & Murch, R. D. Printed endfire beam-steerable pixel antenna. *IEEE Trans. Ant. Propag.* **65**(8), 3913–3923 (2017).
35. Jiang, F. et al. Pixel antenna optimization based on perturbation sensitivity analysis. *IEEE Trans. Ant. Propag.* **70**(1), 472–486 (2022).
36. Koziel, S. et al. On unsupervised artificial intelligence-assisted design of antennas for high-performance planar devices. *Electronics* **12**(16), 3462 (2023).
37. Wolff, M. W. & Nanzer, J. A. Application of pseudoweights in antenna array optimization and design. *IEEE Ant. Wireless Propag. Lett.* **23**(5), 1478–1482 (2024).
38. Liu, Y., Li, M., Haupt, R. L. & Guo, Y. J. Synthesizing shaped power patterns for linear and planar antenna arrays including mutual coupling by refined joint rotation/phase optimization. *IEEE Trans. Ant. Propag.* **68**(6), 4648–4657 (2020).
39. Niu, Z., Zhang, H., Chen, Q. & Zhong, T. Isolation enhancement in closely coupled dual-band MIMO patch antennas. *IEEE Ant. Wireless Propag. Lett.* **18**(8), 1686–1690 (2019).
40. Zhang, Y.-X., Jiao, Y.-C. & Zhang, L. Antenna array directivity maximization with sidelobe level constraints using convex optimization. *IEEE Trans. Ant. Propag.* **69**(4), 2041–2052 (2021).
41. Koziel, S. & Pietrenko-Dabrowska, A. Fast multi-objective optimization of antenna structures by means of data-driven surrogates and dimensionality reduction. *IEEE Access* **8**, 183300–183311 (2020).
42. Li, X. & Luk, K. M. The grey wolf optimizer and its applications in electromagnetics. *IEEE Trans. Ant. Prop.* **68**(3), 2186–2197 (2020).
43. Papathanasopoulos, A., Apostolopoulos, P. A. & Rahmat-Samii, Y. Optimization assisted by neural network-based machine learning in electromagnetic applications. *IEEE Trans. Ant. Propag.* **72**(1), 160–173 (2024).
44. Huang, P.-Q., Zeng, S., Wu, X., Liu, H.-L. & Zhang, Q. A multiobjective evolutionary algorithm for network planning in in-building distributed antenna systems. *IEEE Trans. Network Sc. Eng.* **11**(3), 3002–3014 (2024).
45. Easum, J. A., Nagar, J., Werner, P. L. & Werner, D. H. Efficient multi-objective antenna optimization with tolerance analysis through the use of surrogate models. *IEEE Trans. Ant. Prop.* **66**(12), 6706–6715 (2018).
46. Ravindran, K. & Vinoy, K. J. A reduced-order PCE-based time-domain method for large uncertainties. *IEEE Trans. Ant. Propag.* **71**(6), 5158–5165 (2023).
47. Zhu, X., Di Rienzo, L., Ma, X. & Codecasa, L. Enhanced multilevel monte carlo method applied to FDTD for probability distribution estimation. *IEEE Trans. Ant. Propag.* **71**(10), 8390–8395 (2023).
48. Grout, V. et al. Software solutions for antenna design exploration: a comparison of packages, tools, techniques, and algorithms for various design challenges. *IEEE Ant. Propag. Mag.* **61**(3), 48–59 (2019).
49. Johanesson, D. O. & Koziel, S. "Feasible space boundary search for improved optimization-based miniaturization of antenna structures", *IET Microwaves. Ant. Prop.* **12**(8), 1273–1278 (2018).
50. Mahrokh, M. & Koziel, S. Improved-efficacy EM-based antenna miniaturization by multi-fidelity simulations and objective function adaptation. *Energies* **15**(2), 403 (2021).
51. Ullah, U., Al-Hasan, M., Koziel, S. & Ben Mabrouk, I. EM-driven size reduction and multi-criterial optimization of broadband circularly-polarized antennas using Pareto front traversing and design extrapolation. *Sci. Rep.* **12**, 9877 (2022).
52. Koziel, S. & Pietrenko-Dabrowska, A. On EM-driven size reduction of antenna structures with explicit constraint handling. *IEEE Access* **9**, 165766–165772 (2021).
53. Pietrenko-Dabrowska, A. & Koziel, S. Cost-efficient EM-driven size reduction of antenna structures by multi-fidelity simulation models. *Electronics* **10**, 1536 (2020).
54. Goldberg, D. E. & Holland, J. H. *Genetic algorithms and machine learning* (Springer, 1988).
55. Michalewicz, Z. *Genetic algorithms + data structures = evolution programs* (Springer, 1996).
56. Han, H., Zhou, H., Huang, Y. & Hou, Y. Robust multiobjective particle swarm optimization with feedback compensation strategy. *IEEE Trans. Cybernetics* **54**(2), 1062–1074 (2024).
57. Peng, F. & Chen, X. An efficient antenna optimization framework based on the NP to P problem. *IEEE Trans. Ant. Propag.* **72**(6), 4818–4828 (2024).
58. Lee, J., Choi, J. & Kang, J. Harmony search-based optimization for multi-RISs MU-MISO OFDMA systems. *IEEE Wireless Comm. Lett.* **12**(2), 257–261 (2023).
59. Zhu, D. Z., Werner, P. L. & Werner, D. H. Design and optimization of 3-D frequency-selective surfaces based on a multiobjective lazy ant colony optimization algorithm. *IEEE Trans. Ant. Propag.* **65**(12), 7137–7149 (2017).
60. Le, T. A. & Yang, X.-S. Generalized firefly algorithm for optimal transmit beamforming. *IEEE Trans. Wireless Comm.* **23**(6), 5863–5877 (2024).
61. Li, X. & Guo, Y.-X. Multiobjective optimization design of aperture illuminations for microwave power transmission via multiobjective grey wolf optimizer. *IEEE Trans. Ant. Propag.* **68**(8), 6265–6276 (2020).

62. Darvish, A. & Ebrahimzadeh, A. Improved fruit-fly optimization algorithm and its applications in antenna arrays synthesis. *IEEE Trans. Antennas Propag.* **66**(4), 1756–1766 (2018).
63. Xu, Z., Chu, B., Geng, H., Nian, X. & Zhang, C. Model-guided learning for wind farm power optimization. *IEEE Trans. Control Syst. Techn.* **32**(2), 428–439 (2024).
64. Al-Azza, A. A., Al-Jodah, A. A. & Harackiewicz, F. J. Spider monkey optimization: a novel technique for antenna optimization. *IEEE Antennas Wireless Propag. Lett.* **15**, 1016–1019 (2016).
65. Kovitz, J. M., Rahmat-Samii, Y. Ensuring robust antenna designs using multiple diverse optimization techniques. In *2013 IEEE Antennas and Propagation Society International Symposium (APSURSI)* , pp. 408–409, (2013).
66. Zhang, H. & Jia, R. Application of chaotic cat swarm optimization in cloud computing multi objective task scheduling. *IEEE Access* **11**, 95443–95454 (2023).
67. Bora, T. C., Lebensztajn, L. & Coelho, L. D. S. Non-dominated sorting genetic algorithm based on reinforcement learning to optimization of broad-band reflector antennas satellite. *IEEE Trans. Magn.* **48**(2), 767–770 (2012).
68. Chen, Z., Sun, X., Liu, J., Ren, B. & Wang, Z. Research on maximum power point tracking control in omnidirectional wireless power transfer system. *IEEE Trans. Ind. Electr.* **71**(7), 6612–6621 (2024).
69. John, M. & Ammann, M. J. Antenna optimization with a computationally efficient multiobjective evolutionary algorithm. *IEEE Trans. Ant. Propag.* **57**(1), 260–263 (2009).
70. Greda, L. A., Winterstein, A., Lemes, D. L. & Heckler, M. V. T. Beamsteering and beamshaping using a linear antenna array based on particle swarm optimization. *IEEE Access* **7**, 141562–141573 (2019).
71. Queipo, N. V. et al. Surrogate-based analysis and optimization. *Progr. Aerosp. Sci.* **41**(1), 1–28 (2005).
72. Liu, B. et al. An efficient method for antenna design optimization based on evolutionary computation and machine learning techniques. *IEEE Trans. Ant. Propag.* **62**(1), 7–18 (2014).
73. Koziel, S. & Ogurtsov, S. *Simulation-based optimization of antenna arrays* (World Scientific, 2019).
74. Hassan, A. K. S. O., Etman, A. S. & Soliman, E. A. Optimization of a novel nano antenna with two radiation modes using kriging surrogate models. *IEEE Photonic J.* **10**(4), 4800807 (2018).
75. Alzahed, A. M., Mikki, S. M. & Antar, Y. M. M. Nonlinear mutual coupling compensation operator design using a novel electromagnetic machine learning paradigm. *IEEE Ant. Wireless Prop. Lett.* **18**(5), 861–865 (2019).
76. Tak, J., Kantemur, A., Sharma, Y. & Xin, H. A 3-D-printed W-band slotted waveguide array antenna optimized using machine learning. *IEEE Ant. Wireless Prop. Lett.* **17**(11), 2008–2012 (2018).
77. Wu, Q., Wang, H. & Hong, W. Multistage collaborative machine learning and its application to antenna modeling and optimization. *IEEE Trans. Ant. Propag.* **68**(5), 3397–3409 (2020).
78. Forrester, A. I. J. & Keane, A. J. Recent advances in surrogate-based optimization. *Prog. Aerospace Sci.* **45**, 50–79 (2009).
79. Couckuyt, I., Declercq, F., Dhaene, T., Rogier, H. & Knockaert, L. Surrogate-based infill optimization applied to electromagnetic problems. *Int. J. RF Microw. Computt. Aided Eng.* **20**(5), 492–501 (2010).
80. Chen, C., Liu, J. & Xu, P. Comparison of infill sampling criteria based on Kriging surrogate model. *Sci. Rep.* **12**, 678 (2022).
81. Li, Y., Xiao, S., Rotaru, M. & Sykulski, J. K. A dual kriging approach with improved points selection algorithm for memory efficient surrogate optimization in electromagnetics. *IEEE Trans. Magn.* **52**(3), 1–4 (2016).
82. Yang, Y., Yu, W. M. & Cui, T. J. Efficient RCS modeling with an adaptive design-based gaussian process method. *IEEE Ant. Wireless Propag. Lett.* **23**(7), 2006–2010 (2024).
83. Ogut, M., Bosch-Lluis, X. & Reising, S. C. A deep learning approach for microwave and millimeter-wave radiometer calibration. *IEEE Trans. Geosci. Remote Sens.* **57**(8), 5344–5355 (2019).
84. Yu, X. et al. A method to select optimal deep neural network model for power amplifiers. *IEEE Microwave Wireless Comp. Lett.* **31**(2), 145–148 (2021).
85. Lim, D. K., Yi, K. P., Jung, S. Y., Jung, H. K. & Ro, J. S. Optimal design of an interior permanent magnet synchronous motor by using a new surrogate-assisted multi-objective optimization. *IEEE Trans. Magn.* **51**(11), 8207504 (2015).
86. Toktas, A., Ustun, D. & Tekbas, M. Multi-objective design of multi-layer radar absorber using surrogate-based optimization. *IEEE Trans. Microw. Theory Techn.* **67**(8), 3318–3329 (2019).
87. Lv, Z., Wang, L., Han, Z., Zhao, J. & Wang, W. Surrogate-assisted particle swarm optimization algorithm with Pareto active learning for expensive multi-objective optimization. *IEEE/CAA J. Autom. Sin.* **6**(3), 838–849 (2019).
88. Wu, Q., Chen, W., Yu, C., Wang, H. & Hong, W. Machine-learning-assisted optimization for antenna geometry design. *IEEE Trans. Ant. Propag.* **72**(3), 2083–2095 (2024).
89. Alquaydheb, I. N., Alfawaz, S. A., Ghadimi Avval, A., Ghayouraneh, S. & El-Ghazaly, S. M. Modeling, characterization, and machine learning algorithm for rectangular choke horn antennas. *IEEE Access* **12**, 61697–61707 (2024).
90. Cho, E. K. & Simsek, E. Enhancing the resolution of local near-field probing measurements with machine learning. *IEEE Trans. Microwave Theory Techn.* **72**(3), 1515–1519 (2024).
91. Yang, X. et al. Circularly polarized antenna array synthesis based on machine-learning-assisted surrogate modeling. *IEEE Trans. Ant. Propag.* **72**(2), 1469–1482 (2024).
92. Bilson, S., Hong Loh, T., Héliot, F. & Thompson, A. Physics-informed machine learning modelling of RF-EMF exposure in massive MIMO systems. *IEEE Access* **12**, 69410–69422 (2024).
93. Koziel, S. & Pietrenko-Dabrowska, A. *Performance-driven surrogate modeling of high-frequency structures* (Springer, 2020).
94. Koziel, S. Low-cost data-driven surrogate modeling of antenna structures by constrained sampling. *IEEE Antennas Wireless Prop. Lett.* **16**, 461–464 (2017).
95. Pietrenko-Dabrowska, A. & Koziel, S. Antenna modeling using variable-fidelity EM simulations and constrained co-kriging. *IEEE Access* **8**(1), 91048–91056 (2020).
96. Pietrenko-Dabrowska, A., Koziel, S. & Ullah, U. Reduced-cost two-level surrogate antenna modeling using domain confinement and response features. *Sci. Rep.* **12**, 4667 (2022).
97. Koziel, S. Fast simulation-driven antenna design using response-feature surrogates. *Int. J. RF Micr. CAE* **25**(5), 394–402 (2015).
98. Koziel, S. & Pietrenko-Dabrowska, A. Expedited feature-based quasi-global optimization of multi-band antennas with Jacobian variability tracking. *IEEE Access* **8**, 83907–83915 (2020).
99. Pietrenko-Dabrowska, A. & Koziel, S. Simulation-driven antenna modeling by means of response features and confined domains of reduced dimensionality. *IEEE Access* **8**, 228942–228954 (2020).
100. Vinod Chandra, S. S. & Anand, H. S. Nature inspired meta heuristic algorithms for optimization problems. *Computing* **104**, 251–269 (2022).
101. Blankrot, B. & Heitzinger, C. Efficient computational design and optimization of dielectric metamaterial structures. *IEEE J. Multiscale Multiphys. Comp. Techn.* **4**, 234–244 (2019).
102. Liu, Y. et al. An efficient method for antenna design based on a self-adaptive bayesian neural network-assisted global optimization technique. *IEEE Trans. Ant. Propag.* **70**(12), 11375–11388 (2022).
103. Dutta, K., Akinsolu, M. O., Kumar Mishra, P., Liu, B. & Guha, D. Application of machine learning-assisted global optimization for improvement in design and performance of open resonant cavity antenna. *IEEE Open J. Ant. Propag.* **5**(3), 693–704 (2024).
104. Mwang'amba, R., Mei, P., Akinsolu, M. O., Liu, B., Zhang, S. Gain bandwidth enhancement and sidelobe level stabilization of mm-wave lens antennas using AI-driven optimization. *IEEE Ant. Wireless Propag. Lett.* Early Access, (2024).

105. de Villiers, D.I., Couckuyt, I. and Dhaene, T. Multi-objective optimization of reflector antennas using kriging and probability of improvement. In *2017 IEEE International Symposium on Antennas and Propagation & USNC/URSI National Radio Science Meeting* (pp. 985-986). Ieee., San Diego, USA, (2017).
106. Liu, J., Han, Z., Song, W. Comparison of infill sampling criteria in kriging-based aerodynamic optimization. *28th Int. Congress of the Aeronautical Sciences*, pp. 1–10, Brisbane, Australia, 23–28. (2012).
107. Koziel, S. & Pietrenko-Dabrowska, A. Reliable EM-driven size reduction of antenna structures by means of adaptive penalty factors. *IEEE Trans. Ant. Propag.* **70**(2), 1389–1401 (2021).
108. Alsath, M. G. N. & Kanagasabai, M. Compact UWB monopole antenna for automotive communications. *IEEE Trans. Ant. Prop.* **63**(9), 4204–4208 (2015).
109. Unnsteinsson, S. D. & Koziel, S. Generalized Pareto ranking bisection for computationally feasible multi-objective antenna optimization. *Int. J. RF Microw. CAE* **28**(8), e21406 (2018).
110. Suryawanshi, D. R., Singh, B. A. A compact UWB rectangular slotted monopole antenna. *IEEE Int. Conf. Control, Instrumentation, Comm. Comp. Tech.* (ICCICCT), pp. 1130–1136, (2014).
111. SMA 32K101-400L5, Rosenberger Telematics GmbH, 4850 Timelkam, Austria (2022)
112. Conn, A.R., Gould, N.I.M., Toint, P.L., Trust Region Methods, MPS-SIAM Series on Optimization, (2000).
113. Kennedy, J. & Eberhart, R. C. *Swarm Intelligence* (Morgan Kaufmann, 2001).
114. Cawley, G. C. & Talbot, N. L. C. On over-fitting in model selection and subsequent selection bias in performance evaluation. *J. Machine Learn.* **11**, 2079–2107 (2010).
115. Levy, H. & Lessman, F. *Finite Difference Equations* (Dover Publications Inc., 1992).

Acknowledgements

The authors would like to thank Dassault Systemes, France, for making CST Microwave Studio available. This work was supported in part by the Icelandic Research Fund Grant 2410297, and by National Science Centre of Poland Grant 2022/47/B/ST7/00072.

Author contributions

Conceptualization, S.K. (Slawomir Koziel) and A.P.D. (Anna Pietrenko-Dabrowska); Data curation, S.K. and U.U. (Ubaid Ullah); Formal analysis. S.K.; Funding acquisition, S.K. and A.P.D.; Visualization, S.K., A.P.D., and U.U.; Writing—original draft, S.K.; Writing—review and editing, A.P.D. and U.U.; Sofware and Resources, S.K. and A.P.D.; Supervision, S.K.

Declarations

Competing interests

The authors declare no competing interests.

Additional information

Correspondence and requests for materials should be addressed to S.K.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.